UNIVERSIDAD DE VALLADOLID

E.T.S.I. TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO DE INGENIERÍA DE TECNOLOGÍAS ESPECÍFICAS DE TELECOMUNICACIÓN, MENCIÓN EN SISTEMAS ELECTRÓNICOS

Implementación de un monitor de actividad en la red social Facebook para uso en e-Learning

Autor:
Fátima Herrero Niño

Tutor:
Juan Pablo de Castro Fernández

Valladolid, septiembre de 2017

Universidad de Valladolid

Trabajo Fin de Grado

TÍTULO:	Implementación de un monitor de actividad en la red social Facebook para uso en e-Learning
AUTOR:	Fátima Herrero Niño
TUTOR:	J. Pablo de Castro Fernández
DEPARTAMENTO:	Teoría de la señal y Comunicaciones e Ingeniería Telemática

Tribunal

PRESIDENTE:	J. Pablo de Castro Fernández
VOCAL:	Mª Jesús Verdú Pérez
SECRETARIO:	Luisa Regueras Santos
SUPLENTE:	Francisco Merino Caminero
SUPLENTE 2:	Alonso Alonso Alonso

FECHA:
CALIFICACIÓN:

Resumen

En la actualidad, las redes sociales son una herramienta en la que podemos encontrar grandes posibilidades para desarrollar actividades educativas, pues facilitan la comunicación entre sus usuarios y el acceso a multitud de recursos, sin embargo, el desarrollo de este tipo de actividades no está muy extendido actualmente, quizás sea por la falta de herramientas que ayuden en su gestión y que posibiliten su evaluación.

En este contexto, mediante el desarrollo de este proyecto, se pretende diseñar una aplicación web que permita al docente la evaluación de sus alumnos mediante la participación en un grupo privado de Facebook. Este diseño pretende de forma visual y estadística dar a conocer medidas sobre la implicación de los alumnos en la actividad.

Palabras clave: Facebook, grafo, centralidad, Análisis de Redes Sociales, aplicación, D3, MySQL.

Abstract

At present, social networks are a tool in which we can find great posibilities to develop educational activities, as they facilitate communication between its users and acces to many resources, however, these types of activities are not widespread, maybe by the lack of tools that help in its management and enable its evaluation.

In this context, the development of this proyect aims to desing a web application that allows the teacher to evaluate their students through their participation in a private group of Facebook. This design intends visually and statistically to reveal measures on the involvement of students in the activity.

Keywords: Facebook, grahp, centrality, Social Analysis, aplication, D3, MySQL.

Índice general

ÍNDICE DE FIGURAS **7**

1. INTRODUCCIÓN **8**

1.1 DEFINICIÓN DEL TRABAJO 8
1.2 MOTIVACIÓN 8
1.3 OBJETIVOS 9

2. ESTADO DEL ARTE **9**

2.1 FACEBOOK 10
2.1.1 Posibilidades y limitaciones *12*
2.1.2 Crear la App *15*
2.1.3 API Graph *16*
2.2 ANÁLISIS DE REDES SOCIALES 18
2.2.1 Teoría de grafos *19*
2.2.2 Álgebra matricial *21*
2.2.3 Medidas de centralidad y poder *25*
2.2.4 "Shortest Path", Algoritmo de Dijkstra *29*
2.3 ALMACENAMIENTO DE DATOS 30
2.3.1 MySQL *31*
2.4 VISUALIZACIÓN DE DATOS 32
2.4.1 D3 ..... *33*

3. ACONDICIONAMIENTO DE LOS DATOS **34**

3.1 EL ENTORNO DE DESARROLLO 35
3.2 LA CONSULTA A LOS SERVIDORES DE FACEBOOK 35
3.3 ORGANIZACIÓN DE LOS DATOS 37
3.4 ALMACENAMIENTO DE LOS DATOS 38

4. MODELADO Y ANÁLISIS DE LOS DATOS **41**

4.1 GRAPHVIZ 41
4.2 LA MATRIZ DE ADYACENCIA Y LOS GRAFOS 41
4.3 MEDIDA DE CENTRALIDAD BASADA EN GRADOS 43
4.4 MEDIDA DE CENTRALIDAD BASADA EN CERCANÍA 45
4.5 MEDIDA DE CENTRALIDAD BASADA EN LA INTERMEDIACIÓN 49
4.6 CONCLUSIONES 51

5. VISUALIZACIÓN **51**

5.1 REGISTRO EN LA APLICACIÓN 52
5.1.1 Posibles casos de error en el registro *53*
5.2 ESTADÍSTICAS GENERALES 55
5.3 CENTRALIDAD DE GRADOS 59

5.4 CERCANÍA 60
5.5 GRAFOS DE DIJKSTRA 61
5.6 MATRIZ DE ENLACES 62
5.7 CONCLUSIONES 65

6. CONCLUSIONES 66

6.1 CONCLUSIONES GENERALES 66
6.2 CONCLUSIONES PERSONALES 66
6.3 LÍNEAS FUTURAS 67

7. BIBLIOGRAFÍA 67

Índice de figuras

Figura 2.1: *Número de descargas en Play Store en Junio de 2017* 11
Figura 2.2: *Número mensual de usuarios activos de Facebook en el mundo* 12
Figura 2.3: *Panel de la aplicación* 15
Figura 2.4: *Versiones API de Facebook disponibles actualmente* 17
Figura 2.5: *Los puentes de kônigsberg* 19
Figura 2.6: *Grafo no orientado* 20
Figura 2.7: *Grafo orientado y etiquetado* 21
Figura 2.8: *Grafo orientado y su matriz de adyacencia* 22
Figura 2.9: *Grafo orientado y sus listas de adyacencia* 23
Figura 2.10: *Ejemplo matriz de adyacencia* 24
Figura 2.11: *Ejemplo algoritmo BFS* 30
Figura 2.12: *Icono del sistema de gestión de bases de datos* 31
Figura 2.13: *D3, Data-Driven-Documents* 33
Figura 2.14: *Ejemplo de visualización con D3* 34

Figura 3.1: *Diagrama de la consulta a los servidores de Facebook* 37
Figura 3.2: *Estructura de la tabla* 39
Figura 3.3: *Ejemplo de reacción a un post* 40
Figura 3.4: *Ejemplo de reply* 40
Figura 3.5: *Ejemplo de Post* 40

Figura 4.1: *Grafo ponderado y dirigido de las interacciones* 43
Figura 4.2: *Centralidad de grados (entrada y salida)* 44
Figura 4.3: *Medidas descriptivas de la red en base a la centralidad de grado* 45
Figura 4.4: *Grafo dirigido representando los costes de los enlaces* 46
Figura 4.5: *Grafo de Dijkstra para el nodo "Ana Belén Maroto Herrero"* 47
Figura 4.6: *Centralidad de cercanía* 48
Figura 4.7: *Medidas descriptivas de la red en base a la centralidad de cercanía* 49
Figura 4.8: *Centralidad de intermediación* 50
Figura 4.9: *Medidas descriptivas de la red en base a la centralidad de intermediación* 51

Figura 5.1: *Página de inicio. Registro en la aplicación* 52
Figura 5.2: *Acceso al grupo de Facebook, obtención del identificador del grupo* 53
Figura 5.3: *Obtención del token de acceso* 53
Figura 5.4: *Error token de acceso expirado* 54
Figura 5.5: *Error token de acceso* 54
Figura 5.6: *Error el identificador introducido no existe* 54
Figura 5.7: *error clave secreta de la aplicación no valida* 54
Figura 5.8: *Nombre del grupo y barra de selección* 55
Figura 5.9: *Visualización estadísticas generales del grupo* 56
Figura 5.10: *Visualización de estadísticas en la publicación de reacciones* 57
Figura 5.11: *Visualización de estadísticas para un participante* 58
Figura 5.12: *Estadísticas generales con respecto al total de publicaciones* 58
Figura 5.13: *Visualización centralidad de grados* 59
Figura 5.14: *Visualización centralidad de grados para un participante* 60
Figura 5.15: *Visualización centralidad de cercanía para un nodo de la red* 61
Figura 5.16: *Visualización grafo de Dijkstra* 62
Figura 5.17: *Matriz de enlaces ordenada por cluster* 64
Figura 5.18: *Matriz de enlaces ordenada por frecuencia* 65

1. Introducción

En esta introducción se presentará el tema principal del trabajo fin de grado, los motivos del inicio del proyecto y los objetivos a cumplir.

1.1 Definición del trabajo

El trabajo expuesto mostrará el desarrollo de una aplicación como metodología de evaluación sistemática de la participación en una actividad docente basada en las redes sociales.

Partiremos de una actividad educativa que el profesor o docente plantea en la red social Facebook, principalmente a través un grupo privado de Facebook, aunque se pueden explorar también otras opciones como páginas de Facebook, los participantes serán el propio profesor y todos los alumnos que están matriculados en la asignatura, en dicha actividad el profesor planteará temas de discusión, en los que los alumnos podrán participar desde las diversas opciones que implementa Facebook.

1.2 Motivación

Las redes sociales han revolucionado en pocos años la forma de comunicarnos y compartir la información. Principalmente están basadas en una plataforma web y ofrecen multitud de servicios: mensajería instantánea, correo electrónico, blogs, foros, galerías para compartir fotos, vídeos y archivos, etc.

Dado el alto interés que las redes sociales suscitan entre los estudiantes, cada vez más instituciones educativas y universidades buscan la forma de utilizar estos sistemas.

En ese sentido, existen experiencias de distintos tipos. Hay universidades que han apostado por utilizar las redes sociales ya en funcionamiento, creando canales específicos para sus instituciones, como por ejemplo el IE University integrado en Facebook o MySpace. Otras universidades han apostado por crear sus propias redes sociales, como por ejemplo la Universidad Internacional Menéndez Pelayo, con la creación del portal UIMP 2.0, con más de 900 usuarios registrados [1]. Ejemplos todos ellos de una clara apuesta por las TIC 2.0, que no sólo ha concebido una nueva forma de comunicación sino que ha generado un nuevo escenario de colaboración, de internacionalización y de transferencia de conocimiento.

Como vemos [1], la utilización de las redes sociales en educación puede favorecer la interacción y la comunicación entre los diferentes agentes (profesores, estudiantes, egresados, empleadores, sociedad), fomentar una orientación práctica y profesional de los estudios y ampliar y reemplazar los espacios y tiempos de aprendizaje.

1.3 Objetivos

El objetivo principal es diseñar una aplicación que permita evaluar a los alumnos participantes en un grupo privado de Facebook mediante su participación en el mismo.

Las fases que seguiremos para su desarrollo son las siguientes:

1. Revisión del estado del arte para sentar las bases teóricas sobre las que se desarrollará el proyecto
2. Realizar un acondicionamiento de los datos
3. Modelado y análisis de los datos para extraer conocimiento de la información
4. Visualización de los datos

En la fase inicial del trabajo desarrollaremos los conceptos teóricos sobre los que nos apoyaremos para desarrollar el resto del trabajo, fundamentalmente basados en la plataforma Facebook y en el Análisis de Redes Sociales. También se presentan las herramientas que se utilizarán: herramientas específicas de la plataforma Facebook, el sistema de almacenamiento y la librería de grafos que se usará para la fase final de visualización.

En la segunda fase se presenta el entorno de desarrollo con el que trabajaremos y se exponen los pasos a seguir para obtener los datos con los que realizar el registro en la aplicación, se detalla la estructura de la consulta a los servidores de Facebook, se explica la organización de los datos recolectados y su almacenamiento.

La tercera fase expone las estructuras de datos utilizadas durante el desarrollo de la aplicación, así como el procedimiento llevado a cabo para examinar y analizar la información con el objetivo de generar conocimiento a través de la misma.

La última parte del trabajo será la visualización de la información. Es esencial en esta parte usar herramientas adecuadas y capaces de ofrecer al usuario final la información de una forma clara e interactiva. Por ello, la visualización se realizará con la librería D3.js de JavaScript, la cual cuenta con multitud de ejemplos que cumplen con los requisitos necesarios para trabajar con los datos que obtendremos.

2. Estado del arte

En estas páginas se presentará la información necesaria para conocer los conceptos teóricos que sustentan las bases de nuestro estudio y las herramientas que se utilizarán para su desarrollo.

Comenzaremos con la plataforma o red social Facebook, exponiendo su situación actual, las posibilidades que nos brinda en el contexto de nuestro trabajo y también sus limitaciones. Exploraremos los pasos necesarios para la obtención de la información que necesitamos, como crear una aplicación en Facebook, y el estudio del API Graph.

A continuación hablaremos sobre los conceptos teóricos en los que nos apoyaremos para realizar un análisis de la información extraída de Facebook. Tales como el Análisis de redes sociales, la Teoría de grafos, las medidas de centralidad y el algoritmo de Dijkstra.

En cuanto a las herramientas utilizadas expondremos el sistema seleccionado para almacenar los datos: MySQL, y por último la librería de grafos D3 como método de representación y visualización de la información.

2.1 Facebook

Facebook es hoy uno de los mejores ejemplos del concepto de red social 2.0, siendo una plataforma que facilita el intercambio de información, la interacción y la colaboración entre sus usuarios [2].

Dentro de las páginas web más visitadas en todo el mundo Facebook ocupa la tercera posición, solamente por detrás del buscador Google y de YouTube, por lo que en cuanto a redes sociales se refiere esta ocupa el primer lugar [3].

Facebook es actualmente la red social con mayor número de usuarios, las estadísticas del año 2016 revelan que Facebook generó 4 peta bytes de datos diariamente, medio millón de usuarios cada día, aproximadamente 6 nuevos perfiles al segundo. Los usuarios generaron ese año 4 millones de “likes” por minuto y de media pasaron unos 20 minutos al día en la plataforma. Además su app se consolida entre las más descargadas a nivel mundial, ya que el 47% de los usuarios únicamente acceden a la plataforma a través de sus smartphones [4].

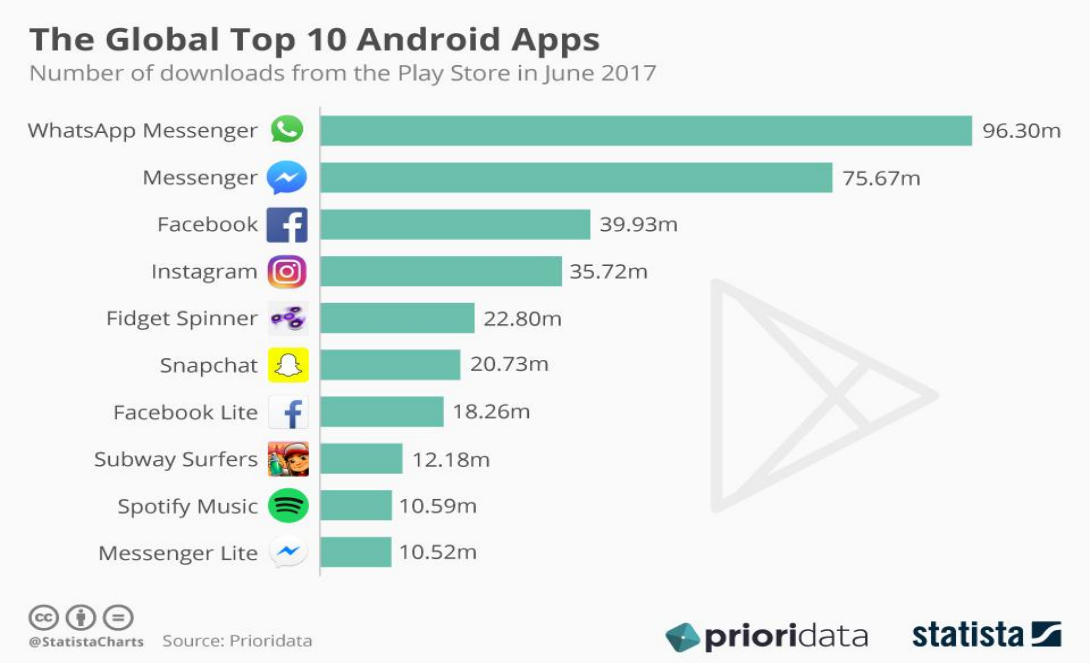


Figura 2.1: *Número de descargas en Play Store en Junio de 2017*

La red social más grande del mundo continúa creciendo. Como director ejecutivo y cofundador Mark Zuckerberg anunció recientemente en su página de Facebook que la red social ahora tiene más de 2 mil millones de usuarios activos mensuales. Para hacernos a la idea de lo que este número significa podemos decir que más de 1 de cada 4 personas en el mundo utilizan esta red social [5].

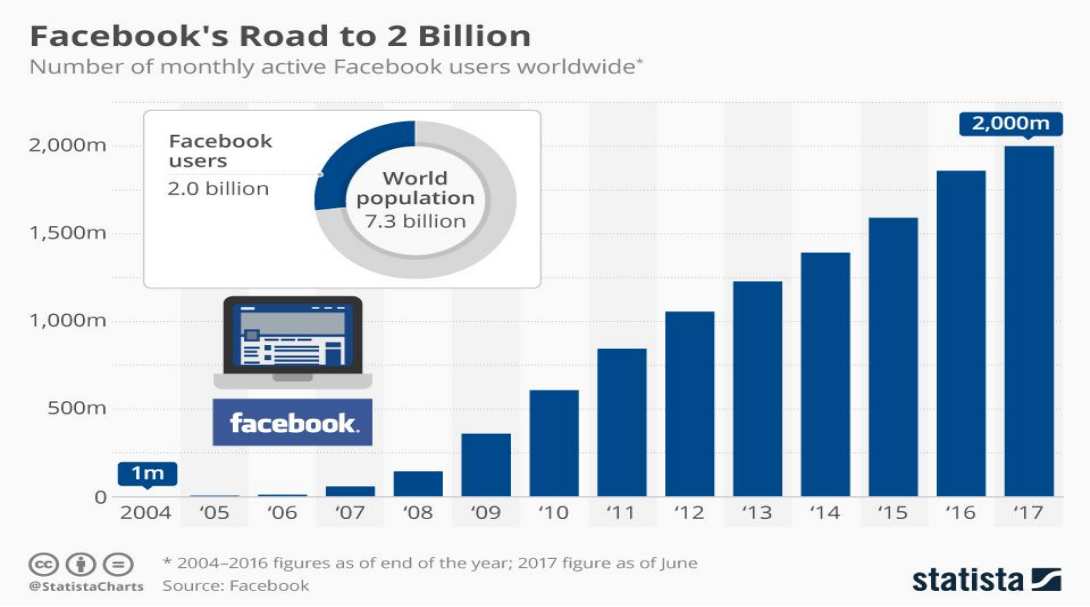


Figura 2.2: *Número mensual de usuarios activos de Facebook en el mundo*

Aunque su orientación inicial es la capacidad de conectar personas, no es la única característica de su éxito. Debemos destacar la potencia de la plataforma para compartir recursos, vincular contenidos presentes en internet a los perfiles de los usuarios y con ello poder estructurar procesos de aprendizaje colaborativo [2].

De lo anterior subyace la cuestión de modificar los flujos en la comunicación académica provocando un escenario en el que todos se relacionan con todos permanentemente (como si todo pasara en un foro) y de alterar el concepto de movilidad en el proceso de comunicación académica on-line: en vez de motivar al alumno a entrar en los entornos educativos, ¿por qué no llevar los contenidos académicos a los entornos en los que acostumbra a estar el alumno? ¿Por qué no usar las redes sociales como entornos virtuales docentes?

Gracias a su potencial de interactividad Facebook propicia la participación activa y el aprendizaje centrado en el estudiante. La colaboración entre los alumnos es una característica definitoria de las aulas constructivas, y Facebook presenta extensas posibilidades de interactividad social y de fomento de la colaboración y el aprendizaje colectivo [2].

2.1.1 Posibilidades y limitaciones

Es importante realizar un estudio previo de las posibilidades y, cómo no, también de sus limitaciones, esto es lo que se hizo para desarrollar “Web2Learn”, un proyecto de la Universidad Oberta de Cataluña cuyo principal objetivo fue generar conocimiento sobre las posibilidades de Facebook para el desarrollo de actividades de aprendizaje colaborativo a través de la creación de un grupo privado en la red Facebook [6].

Comenzando por los puntos fuertes de Facebook para el desarrollo del trabajo colaborativo y desde un punto de vista tecnológico enumeraron los siguientes:

- Sencillez y rapidez en la creación y administración de un grupo de trabajo. Un usuario individual crea, desde su perfil privado, el nuevo grupo de trabajo e invita al resto de los participantes.

- Simplicidad de uso de las herramientas nativas. Las funciones básicas de Facebook (muro, fórum, fotos…) resultan fáciles de utilizar, accesibles, intuitivas y visualmente bien estructuradas. El grupo puede comenzar a interactuar inmediatamente tras su creación.

- Chat, mensajería y etiquetado de imágenes. Estas funciones propias de los entornos 2.0 también se encuentran presentes en la red social. Sin embargo, en el caso del etiquetado (tagging), sólo está disponible para las imágenes, y su objetivo inicial es reconocer a personas en fotografías. Nada impide que utilicemos la función, por ejemplo, para documentar gráficos.

- Elevado nivel de conectividad externa. La extraordinaria expansión de Facebook, la presencia social en la red de marcas y eventos sociales y publicitarios, ha obligado a muchos servicios externos que proveen contenidos a implementar conectividad con Facebook.

- Capacidad de expansión interna. Gracias al desarrollo de aplicaciones de la propia red y de programadores independientes, los usuarios, así como los grupos, pueden ampliar las capacidades nativas mediante módulos adicionales.

- Potente soporte para el mobile learning. El soporte para dispositivos y sistemas operativos móviles (IPhone, Android, Maemo, etc.) ha constituido uno de los avances más espectaculares de la plataforma. Desde septiembre de 2007, momento en que se abrió el servicio, millones de usuarios utilizan conexiones móviles con Facebook, y se da la circunstancia de que estos son dos veces más activos en la plataforma que los usuarios que se conectan a través de ordenador. Esta flexibilidad es una indudable ventaja para desarrollar experiencias de mobile learning que resultarían poco o nada funcionales en otros entornos.

Las características de soporte que acabamos de describir revelan que estamos ante una plataforma tecnológicamente rica, adaptable y expandible, capaz de soportar experiencias de trabajo colaborativo en comunidades de aprendizaje. Sin embrago para emitir un diagnóstico más preciso quisieron atender también a las posibles carencias o limitaciones de la red social para los objetivos de su estudio.

Al realizar el estudio observaron que los elementos que obstaculizaban el desarrollo de su proyecto eran una combinación entre el escaso desarrollo de algunas herramientas nativas de Facebook, y la inexistencia de las mismas. Enumeraron las siguientes:

- Presencia de «ruido» y elementos distractores. Cuando se utiliza Facebook como red orientada al trabajo en grupo, ciertos elementos que resultan justificados en el uso lúdico-individual de la red pueden aportar factores de dispersión. Elementos como publicidad, avisos, sugerencias y solicitudes que resultan superfluos en entornos de trabajo en equipo.

- El sistema de comentarios desplegables en los muros tiende a dificultar la visualización de la información. Así como la diferencia entre los elementos «leídos» y «no leídos» no es fácilmente perceptible.

- Facebook carece de un auténtico sistema de etiquetado, filtrado, búsqueda y organización de la información. La velocidad en la producción de noticias, y la misma velocidad en la «desaparición» de lo «no nuevo» no son compatibles a priori con un «trabajo de almacén», en el que el conocimiento generado deba ser clasificado, guardado y recuperado cómoda y rápidamente.

- Los foros de los grupos de Facebook son excesivamente primarios. Por una parte, carecen de un editor de texto que preste funcionalidad a una exposición mínimamente compleja de un tema. Por otra, no permiten ningún tipo de anidado de las respuestas, con lo que se convierten, a pesar de su nombre, en simples listas de réplica a un tópico principal.

- Facebook carece de funcionalidades nativas propias de los entornos orientados específicamente a equipos de trabajo. No es posible subir archivos (excepción hecha de fotos y vídeos), editarlos entre varios usuarios, realizar listas de tareas, asignarlas a miembros o hacer su seguimiento.

- Los miembros individuales de un grupo no pueden crear eventos. Esto es así aunque posean el rango de administradores del grupo. Esta función resulta, en su fase de desarrollo actual, muy poco operativa, pues si se decide que varios usuarios sean administradores es porque se supone que deben tener acceso a la gestión de los eventos y al calendario del grupo.

- Facebook no proporciona, de manera nativa, sincronía bidireccional de audio y vídeo. Aunque los usuarios pueden dejar mensajes y grabaciones de vídeo, no es posible utilizar este canal para establecer conversaciones bidireccionales en tiempo real.

En cuanto a las limitaciones expuestas cabe destacar que algunas de estas carencias pueden ser insignificantes para las necesidades de muchos equipos de trabajo, que

aún con ellas encuentran en la red un magnifico soporte para sus interacciones, especialmente si se basan en el intercambio y discusión de recursos que se obtienen vía navegación e integración de contenidos externos.

2.1.2 Crear la App

Antes de la obtención de toda la información que se genera en un grupo de Facebook y con este mismo fin debemos crear una aplicación en la plataforma, indispensable para conectar con su interfaz.

El requisito inicial para ello es ser usuario de la propia red social, nuestra cuenta debe ser una cuenta de desarrollador, por lo que si ya contábamos con una cuenta personal debemos actualizarla para convertirla en cuenta de este tipo.

El siguiente paso sería seleccionar en el menú el apartado de "aplicaciones" y escoger la opción de "agregar una nueva aplicación".

Para continuar con la configuración de la aplicación tenemos que agregar la plataforma, dar un nombre a nuestra aplicación y seleccionar "crear nuevo identificador de la aplicación de Facebook" así Facebook nos dará un identificador único que representa la aplicación.

Por otro lado Facebook genera automáticamente una clave secreta de la aplicación y la asocia con ella, esta clave secreta sirve para autenticar las solicitudes que la aplicación hace a los servidores de Facebook y se debe tratar como una contraseña.

Figura 2.3: *Panel de la aplicación*

El identificador de la aplicación y la clave secreta serán necesarios para lanzar la consulta a los servidores de Facebook, y rescatar así los datos que necesitamos.

Podremos seguir configurando características de la aplicación como la categoría a la que pertenece (educación, finanzas, compras etc.), los participantes y sus roles

(desarrollador, administrador, evaluador etc.), características de seguridad, icono de la aplicación y correo electrónico de contacto, versión de la API que utiliza etc.

2.1.3 API Graph

La API Graph de Facebook es el modo principal para extraer e ingresar datos en la plataforma de Facebook. Se trata de una API basada en HTTP que podemos utilizar de manera programática para consultar datos, publicar nuevas historias, administrar anuncios, subir fotos y llevar a cabo varias tareas más propias de una aplicación. Este paso es vital para entender como está estructurada la información que extraemos de la plataforma, y por tanto conocer cómo debemos manipularla.

Como ya anunciamos Facebook no es solo una red social, si bien una de las características que la distingue del resto de las redes sociales es el uso de la misma para el desarrollo y construcción de aplicaciones con software de código abierto.

Facebook cuenta con varias interfaces de programación de aplicaciones (API por sus iniciales Application Programming Interfaces), a nivel general tenemos Atlas API para crear y gestionar campañas de publicidad, API de marketing que permite crear y gestionar anuncios en la plataforma y API Graph, el modo principal para que las aplicaciones lean y escriban en el grafo social de Facebook, además todas ellas cuentan con otras API's generalmente para obtener estadísticas, informes o mediciones acerca del uso, rendimiento o visualización de las aplicaciones.

Tiene varias versiones disponibles a las que se puede acceder en cualquier momento. Cada versión contiene un conjunto de campos básicos y operaciones perimetrales. Facebook garantiza que esas API básicas estarán disponibles y sin modificar en esa versión durante al menos dos años a partir de su lanzamiento.

Actualmente la versión más reciente de la API es la 2.10, que se presentó el 18 de julio de 2017. La lista de versiones disponibles actualmente se muestra en la siguiente imagen.

Versión	Ruta	Fecha de presentación	Disponible hasta
v2.10	/v2.10/{object}	18 de julio del 2017	
v2.9	/v2.9/{object}	18 de abril del 2017	18 de julio del 2019
v2.8	/v2.8/{object}	5 de octubre del 2016	18 de abril del 2019
v2.7	/v2.7/{object}	13 de julio del 2016	5 de octubre del 2018
v2.6	/v2.6/{object}	12 de abril del 2016	13 de julio del 2018
v2.5	/v2.5/{object}	7 de octubre del 2015	12 de abril del 2018
v2.4	/v2.4/{object}	8 de julio del 2015	7 de octubre del 2017

Figura 2.4: *Versiones API de Facebook disponibles actualmente*

Existe un registro de cambios de la plataforma de Facebook en el que se puede encontrar las actualizaciones o deprecaciones que incluyen las nuevas versiones.

Revisado el registro de cambios, nosotros trabajaremos con la última versión a fin de mantener la aplicación lo más actualizada posible, aunque podremos trabajar con cualquiera de las versiones disponibles, pues ninguna presenta características que impidan el correcto funcionamiento de la aplicación.

La API Graph recibe este nombre por la idea de una "gráfica social", la representación de la información en Facebook consta de:

- **Nodes:** básicamente elementos como un usuario, una foto, una página, un comentario, estos tienen un identificador único mediante el cual se accede a ellos.
- **Edges**: las conexiones entre esos elementos, como fotos de páginas o comentarios de fotos.
- **Fields**: información sobre esos elementos, como el cumpleaños de una persona o el nombre de una página.

Todos los nodos y perímetros se pueden leer simplemente con una solicitud GET HTTP al end-point relevante, o bien escribir mediante solicitudes POST HTTP.

Tenemos opción de realizar consultas anidadas; es decir, podremos anidar de forma eficaz varias consultas en una sola llamada. Esta funcionalidad se denomina "expansión de campo", y nos permite obtener uno o varios campos o perímetros del nodo principal, que pasarían a ser nodos de primer nivel, y uno o más perímetros o campos del nodo de primer nivel, que serían nodos de segundo nivel, así sucesivamente [7].

No hay limitación en cuanto a la cantidad de anidación de niveles, además existe un argumento para restringir la cantidad de objetos que quieres obtener en cada campo o perímetro.

Facebook ofrece también una guía completa de nodos raíz a los cuales se podemos acceder directamente y nodos no raíz, accesibles mediante los primeros o por enlaces a estos, [7].

Centrándonos en nuestro estudio, solamente necesitaremos leer información del nodo raíz que para nosotros será el grupo de Facebook. Realizaremos una consulta anidada para obtener en una única llamada todos los datos que precisamos, y reducir los tiempos de espera.

Para realizar la consulta necesitaremos un token de acceso, que podemos obtener desde el explorador de la API Graph iniciando sesión en Facebook y seleccionando la aplicación. Dependiendo de la consulta que realicemos necesitaremos o no permisos, estos se deben seleccionar a la hora de generar el token de acceso. El explorador de la API Graph es un entorno con el cual podemos practicar realizando solicitudes o consultas y observar los datos obtenidos.

Podemos encontrarlo en: https://developers.facebook.com/tools/explorer .

2.2 Análisis de Redes Sociales

Fundada por Jacob Levy Moreno en la década de 1930, la sociometría busca indagar la evolución y organización de grupos de personas y la posición de los distintos individuos en ellos, para ello se vale de algunas técnicas cuantitativas, como las sociomatrices y los sociogramas, para representar las relaciones sociales. El principal supuesto teórico de la sociometría es que el comportamiento de los actores podría ser explicado por el efecto de las relaciones sociales en ellos; y la metodología a la que recurre implica vincular sistemáticamente grafos matemáticos con problemas sociales con el objeto de graficar y medir las relaciones sociales, siendo así un antecedente principal del Análisis de Redes Sociales, [8].

Su fundador, Jacob Levy Moreno, la describió como:

“La sociometría tiene por objeto el estudio matemático de las propiedades psicológicas de las poblaciones; con este fin utiliza una técnica experimental fundada sobre los métodos cuantitativos y expone los resultados obtenidos por la aplicación de estos métodos. Persigue así una encuesta metódica sobre la evolución y la organización de los grupos y sobre la posición de los individuos en los grupos” [9].

El análisis de redes sociales (ARS) o social network analysis (SNA) en inglés, se focaliza en la estructura reticular de las relaciones sociales, buscando dar cuenta del efecto de las relaciones en el comportamiento de los individuos, los distintos grupos sociales y la sociedad en su conjunto.[8]

Obviamente una idea fundamental que sustenta el ARS es que las interacciones entre individuos y organizaciones en la red social, además de reflejar los flujos de conocimiento y comunicación, podrían tener un impacto relevante en el

comportamiento de los actores, así como resultados en las estructuras de poder identificables y en los procesos de aprendizaje [10].

No es solo análisis de las redes sociales entendidas como análisis de contenido de Social Media. Es un estudio numérico, algebraico, de una representación de conocimiento en formato de grafo. Un campo que mezcla la sociología y las matemáticas (el álgebra de grafos) en el que hay actores o entidades que interactúan, pudiendo representar estas acciones a partir de un grafo.

El interés por estudiar los patrones y estructura que esconden esta representación de nodos y aristas ha crecido en los últimos años a medida que ha aumentado la relación entre agentes. Es decir, a medida que han crecido las redes sociales (¿cómo se relacionan mis clientes en Facebook?), ha crecido la influencia de una persona en otra para comprar (los millenials confían más en la reputación de sus amigos que en la publicidad de las marcas), las redes de proveedores y clientes han aumentado sustancialmente (por la globalización de la economía y la interconexión internacional), etc., crece el interés por estudiar qué patrones pueden descubrirse para incrementar la inteligencia del negocio [9].

El ARS se ha consolidado como técnica de análisis de las relaciones gracias a la utilización del álgebra matricial y de los grafos. Las dos formas de representación nos dicen lo mismo de la estructura de relaciones de un conjunto y nos permiten determinar, tanto las características de la estructura como las propiedades de la posición de cada nodo en la red [10].

2.2.1 Teoría de grafos

La teoría de grafos surge cuando Euler, en el año 1736, explica matemáticamente la imposibilidad de resolver el acertijo de “los puentes de Kônigsberg” (hoy ciudad de Kalingrado) que consistía, básicamente, en intentar trazar un camino a través del cual cruzar los 7 puentes de Kônigsberg sin pasar dos veces por el mismo puente, [8].

Figura 2.5: *Los puentes de kônigsberg*

Euler determinó, en el contexto del problema, que los puntos intermedios de un recorrido posible necesariamente han de estar conectados a un número par de líneas. En efecto, si llegamos a un punto desde alguna línea, entonces el único modo de salir

de ese punto es por una línea diferente. Esto significa que tanto el punto inicial como el final serían los únicos que podrían estar conectados con un número impar de líneas. Sin embargo, el requisito adicional del problema dice que el punto inicial debe ser igual al final, por lo que no podría existir ningún punto conectado con un número impar de líneas [11].

Un grafo es un conjunto, no vacío, de objetos llamados vértices (o nodos) y una selección de pares de vértices, llamados aristas (edges en inglés) que pueden ser orientados o no. Típicamente, un grafo se representa mediante una serie de puntos (los vértices) conectados por líneas (las aristas).

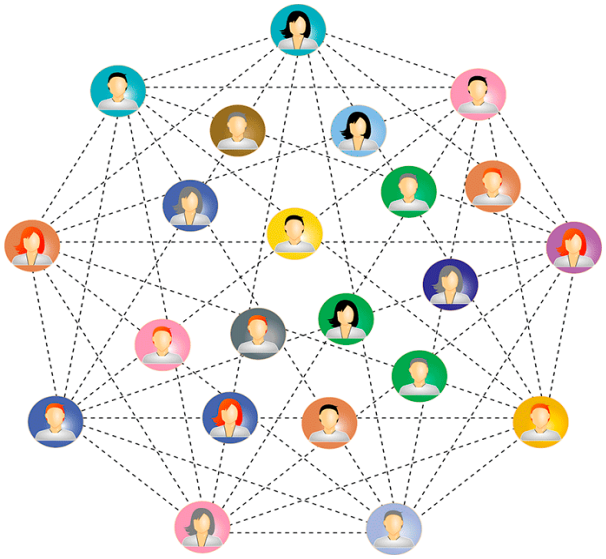


Figura 2.6: *Grafo no orientado*

En algunos casos, interesa asignar un sentido a las aristas del grafo, si queremos ver quien se relaciona con quien necesitamos utilizar grafos orientados. Además si queremos saber por ejemplo la frecuencia con la que dos nodos se comunican deberemos incluir en el grafo algún tipo de etiqueta en las aristas, que nos dé cuenta de alguna característica de la conexión.

Pasamos entonces a hablar de grafos orientados y etiquetados, propiedades que enriquecen sin duda alguna el grafo, pues ahora no solo sabemos si dos nodos están conectados o no, sino que sabemos quién establece dicha conexión, y en base a qué característica.

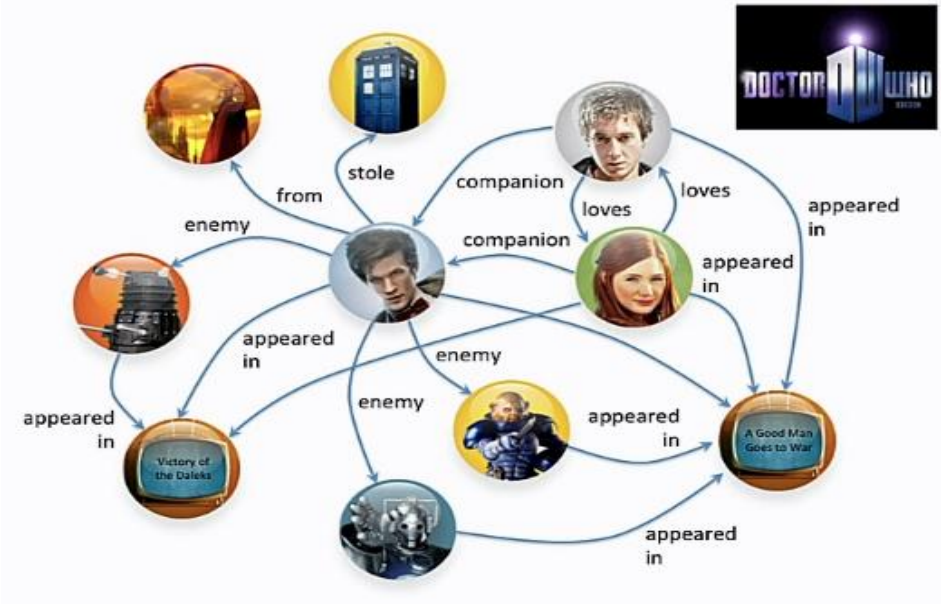


Figura 2.7: *Grafo orientado y etiquetado*

La teoría de grafos ha sido muy útil para el ARS porque tiene un vocabulario que puede ser utilizado para analizar muchas propiedades de las estructuras sociales, nos ofrece las operaciones matemáticas por las cuales esas propiedades pueden analizarse y medirse y nos permite probar teoremas sobre los grafos y, por tanto, deducir y someter a test determinados enunciados, [10].

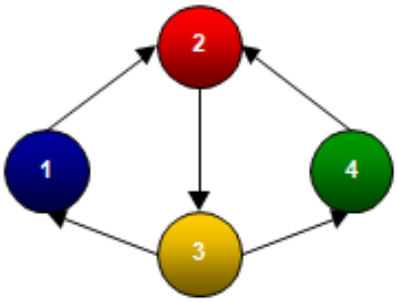
2.2.2 Álgebra matricial

Los grafos constituyen una manera muy útil de representar información sobre redes sociales. Sin embargo, cuando existen muchos actores y/o muchas clases de relaciones, éstos pueden hacerse tan visualmente complicados que se hace muy difícil identificar estructuras. En su lugar, también es posible representar la información sobre redes sociales en forma de matrices o listas. De esta forma, la representación de la información permite la utilización de herramientas matemáticas y de computación para identificar estructuras [12].

Existen diferentes formas de almacenar grafos en una computadora. La estructura de datos usada depende de las características del grafo y el algoritmo usado para manipularlo. Entre las estructuras más sencillas y usadas se encuentran las listas y las matrices, aunque frecuentemente se usa una combinación de ambas. Las listas son preferidas en grafos dispersos porque tienen un eficiente uso de la memoria. Por otro lado, las matrices proveen acceso rápido, pero pueden consumir grandes cantidades de memoria.

Matriz de adyacencia

Dado el grafo de la imagen, se representa su matriz de adyacencia, donde un "1" equivale a la existencia de un enlace y "0" representa el caso opuesto.



$$a = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Figura 2.8: *Grafo orientado y su matriz de adyacencia*

- ☺ Se puede determinar en un tiempo fijo y constante si un enlace (arco) pertenece o no al grafo.
- ☺ Es fácil determinar si existe o no un arco o enlace, solo se debe posicionar en la matriz.
- ☺ Es fácil determinar si existe un ciclo en el grafo, basta multiplicar la matriz por ella misma n veces hasta obtener la matriz nula (no hay ciclos) o bien una sucesión periódica de matrices (hay ciclo).

- ☹ Se requiere un almacenamiento igual al cuadrado de los nodos.
- ☹ El tiempo necesario para leer o examinar la matriz es proporcional al cuadrado de los nodos.

Listas de adyacencia

La lista de adyacencia para un vértice es una lista enlazada de todos los vértices adyacentes a él. Un grafo puede ser entonces representado por tantas listas de adyacencia como vértices tenga.

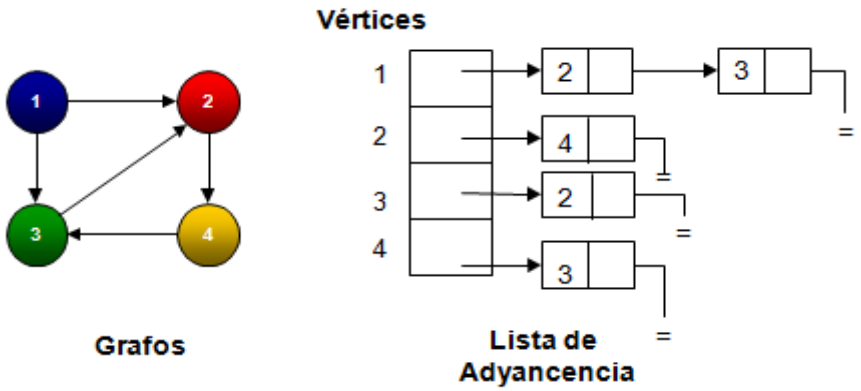


Figura 2.9: *Grafo orientado y sus listas de adyacencia*

- ☺ La lista de adyacencia requiere un espacio proporcional a la suma del número de vértices más el número de enlaces (arcos). Hace buen uso de la memoria.
- ☺ Se utiliza bastante cuando el número de enlaces es mucho menor que el número de nodos al cuadrado.
- ☺ En la representación con lista de adyacencia el tiempo necesario para examinar si existe un enlace entre dos nodos es proporcional al número de nodos, ya que en la lista de adyacencia del vértice examinado pueden estar todos los nodos.

Pasaremos a desarrollar ahora la matriz de adyacencia, que usaremos en nuestro estudio junto a los grafos como almacenamiento de la existencia de relaciones entre los actores de nuestro grupo de Facebook. Tomaremos la siguiente matriz de adyacencia como referencia para explicar algunos conceptos.

	Bob	Carol	Ted	Alice
Bob	-	1	0	0
Carol	1	-	1	0
Ted	1	1	-	1
Alice	0	0	1	-

Figura 2.10: *Ejemplo matriz de adyacencia*

La forma más común de matriz en el análisis de redes sociales es una matriz simple compuesta por tantas filas y columnas como actores existan en el conjunto de datos y dónde los elementos representan los vínculos entre los actores. La más simple y común de las matrices es la matriz binaria. Es decir, si existe un vínculo, se coloca un 1 en la celda, si no lo hay se escribe un cero. Este tipo de matriz es el punto de partida de casi todos los análisis de redes y se llama "matriz de adyacencia" porque representa quién está cerca de quién, o adyacente a quién en el "espacio social" mostrado por las relaciones que hemos medido. Por convención, en un grafo dirigido el origen de un vínculo es la fila y el objeto es la columna.

Es necesario tener presente que las filas representan el origen de los vínculos dirigidos y las columnas los destinos, éste es un ejemplo de una matriz "asimétrica" que representa vínculos dirigidos (vínculos que van de un origen a un receptor). Es decir, el elemento i,j no necesariamente es igual que el elemento j,i.

En la representación de datos de redes sociales como matrices, la pregunta que siempre surge es: ¿qué hago con los elementos de la matriz donde i=j? Es decir, por ejemplo, ¿Bob se ha considerado como amigo cercano de él mismo? A esta parte de la matriz se le conoce como la diagonal principal. A veces, el valor de la diagonal principal carece de significado y se ignora (se deja en blanco).

A menudo es conveniente hacer referencia a ciertas partes de una matriz utilizando una terminología resumida. Si se toman todos los elementos de una fila (por ejemplo, a quiénes escoge a Bob como amigos: 1,1,1,0), se examina el "vector fila" para Bob. Si sólo se observa quién escoge a Bob como amigo (la primera columna o 1,0,1,0), se examina el "vector columna" para Bob. A veces es útil hacer ciertas operaciones con los vectores fila o los vectores columna. Por ejemplo, si se suman los elementos de los vectores columnas en este ejemplo, puede obtenerse la medida de cuán "popular" es cada nodo (en términos de con cuánta frecuencia fueron objeto de un vínculo amistad),[12].

2.2.3 Medidas de centralidad y poder

Debido a la utilización de las redes sociales como metodología de aprendizaje aparece la necesidad del estudio de las reacciones, por ello, es necesario aplicar métodos que permitan analizar los roles participativos de los estudiantes, con el propósito de identificar el valor de cada individuo y su importancia dentro de la red, formada a partir de una actividad colaborativa, identificando los factores que benefician la interacción y la participación a partir de las relaciones dirigidas entre los participantes [6].

La perspectiva de redes ha contribuido en gran parte a la comprensión del poder social. Quizá aún más importante, la aproximación a las redes sociales enfatiza que el poder es inherentemente relacional. Un individuo no tiene poder en abstracto, se tiene poder porque se puede dominar a otros. Debido a que el poder es una consecuencia de los patrones de relación, la cantidad de poder en las estructuras sociales puede variar. Si un sistema está muy débilmente acoplado (baja densidad) poco poder puede ser ejercido; en sistemas de alta densidad existe el potencial para mayor poder. El poder es una propiedad sistémica (macro) y relacional (micro).

El poder en las redes sociales podría ser visto tanto como una propiedad micro (i.e. éste describe las relaciones entre actores) o como una propiedad macro (i.e. éste describe la población entera). Por tanto, como pasa en otros conceptos sociológicos claves, lo macro y lo micro están íntimamente conectados en la perspectiva de redes sociales.

Los analistas de redes a menudo describen la manera en la que un actor está insertado en una red relacional en la medida que se le imponen restricciones y se le ofrecen oportunidades. Los actores que se enfrentan a menos restricciones y tienen más oportunidades que otros, están en posiciones estructurales favorables. Tener una posición favorable significa que un actor puede extraer mejores ofertas en los intercambios y que será un foco para la deferencia y atención por parte de aquéllos en posiciones menos favorables.

Pero ¿a qué nos referimos con 'tener una posición favorable' y 'tener más oportunidades' y 'menos restricciones'? No existe una sola respuesta correcta y definitiva para estas difíciles preguntas. Sin embargo, el análisis de redes ha hecho importantes contribuciones en proveer definiciones precisas y mediciones concretas de las diferentes aproximaciones a la noción de poder que se atribuye a las posiciones en estructuras de relaciones sociales.

Existen en el análisis de redes tres enfoques o medidas muy extendidas que describen la ubicación de los individuos en términos de cuán cerca están del centro de la red, o visto de otra manera cuanto de central es con respecto al resto de la red.

Se definen así las medidas de centralidad siguientes: centralidad de grado, cercanía e intermediación, que pasaremos a describir a continuación, [14].

Centralidad de grado

Esta medida está basada en el número de conexiones directas que tiene un actor con el resto de actores de la red. Es intuitivo pensar que a mayor número de conexiones su posición en la red será más destacada. Es por tanto una medida muy simple pero eficaz que nos permite estimar la centralidad o poder de un individuo, aunque sin tener en cuenta la estructura global de la red.

Si utilizamos datos de relaciones orientadas es importante hacer una distinción entre el grado de salida, número de conexiones directas que inicia un determinado actor y que finalizan en diversos actores de la red, y grado de entrada o número de conexiones directas originadas por otros actores de la red y con destino un único actor. En redes dirigidas tenemos que calcular la centralidad de grado de entrada y también la de salida para cada actor.

La formulación matemática que se presenta a continuación, sirve para el cálculo de la centralidad de grado tanto de entrada como de salida de un determinado miembro de la red.

La notación es la siguiente: C es la centralidad, que es función de un nodo específico n_i donde i = 1,2,…, N, siendo N el número de nodos de la red, $\mathrm{g(n_i)}$ un enlace entre n_i y cualquier otro nodo, el índice G señala el tipo de C, en este caso se refiere a "grado". $C_G(n_i)$ está definida en el intervalo [0, N-1].

$$C_G(n_i) = \sum_{i=1}^{N} g(n_i)$$

En base a la idea de grados de entrada y salida, podremos decir que un actor es prominente o de prestigio si tiene un alto grado de entrada, pues los demás actores buscan relacionarse con él, lo que puede ser visto como un acto de deferencia o reconocimiento hacia su posición, indicando que esta es valiosa. Por otro lado si el grado de salida es alto, implicaría un deseo por parte del individuo de establecer conexiones que le permitirían hacer llegar sus inquietudes, razonamientos o conocimientos al resto del conjunto, pudiendo estos ser influenciados por sus puntos de vista, este tipo de actores son denominados actores influyentes.

De manera general, que un actor tenga altos grados de salida y entrada indicaría que es un actor central dentro de la red, que podría obtener más recursos del conjunto que otros actores con menos conexiones o grados, pudiendo además servir de conector o puente entre distintos pares de actores, posibilitando o no la conexión entre estos y la dispersión de los recursos.

Mediante la medida de centralidad de grados podemos definir la centralización de la red, este concepto recoge el significado de centralidad cuando se aplica al conjunto de la red. Por lo tanto en este punto hablaremos de la centralización de grado de la red, definida por la siguiente ecuación, que sigue la misma notación que la ecuación anterior a diferencia de que aquí C_G es la centralización de grado de la red y $C_G(n^*)$ es el valor máximo de la medida de centralidad de grado para los N actores de la red. C_G se define en el intervalo [0,1], aunque se suele expresar en tanto por ciento.

$$C_G = \frac{\sum_{i=1}^{N}[C_G(n^*) - C_G(n_i)]}{(N-1)(N-2)}$$

Esta medida tomará el valor máximo de 1 cuando un solo actor está conectado al resto que no tienen conexiones entre sí (grafo en forma de estrella) y el valor mínimo de 0 corresponderá a un grafo en el que todos los nodos tienen el mismo grado de centralidad; las mismas conexiones. Valores cercanos a la unidad, o en porcentaje al 100% de esta medida de centralización evidenciarán la existencia de actores muy influyentes o prestigiosos dentro de la red.

Centralidad de cercanía

La medida anterior basada en grados o conexiones directas, podría no servir por si sola sí lo que nos importa para dar una medida de centralidad de un actor son sus conexiones ya sean directas o indirectas.

En este sentido un actor puede estar conectado indirectamente con otros actores, aunque estos no estén muy bien conectados con el conjunto, siendo a lo mejor un actor central dentro de un subconjunto de la red. Esta medida de centralidad de cercanía está construida fundamentalmente en base al concepto de distancia geodésica, siendo esta la distancia de un actor a otro, hallada por el camino más corto entre ambos.

Para hallar el camino más corto de un actor a otro, necesitaremos un grafo cuyos enlaces representen el coste de la comunicación, los pesos o costes de los enlaces reflejan el flujo de la comunicación entre ambos. Así un enlace con un coste superior a otro implicaría que la comunicación entre esos actores es menos fluida o de peor calidad que la que pueda existir entre otros pares de actores.

La suma de todas las distancias geodésicas de un actor al resto de actores de la red nos darán un valor que llamaremos “lejanía” del individuo con respecto a la red.

Ahora podremos convertir esta medida en una medida de cercanía o centralidad de cercanía al tomar la inversa de la lejanía, y dividirla entre el número de nodos de la red menos uno.

Su formulación matemática sigue la notación anterior, aunque ahora el tipo de C es C que indica "cercanía" y $d(n_i, n_j)$ representa la distancia geodésica entre dos nodos.

$$C_C(n_i) = \frac{\left[\sum_{i=1}^{N} d(n_i, n_j)\right]^{-1}}{N-1}$$

En redes pequeñas y con alta densidad, las medidas de centralidad basadas en grados o en cercanía no difieren demasiado, debido al número de conexiones directas.

Para medir la centralización de la red basada en la medida de centralidad de cercanía Freeman propone la siguiente fórmula:

$$C_C = \frac{\sum_{i=1}^{N} [C_c(n^*) - C_c(n_i)]}{[(N-1)(N-2)]/(2N-3)}$$

Sin embargo, está formula carece de sentido en redes donde los enlaces están ponderados, ya que el denominador representa el máximo teórico del numerador para redes donde todos los enlaces tienen el mismo coste.

Como nosotros trabajaremos con grafos ponderados, la fórmula anterior no nos servirá, pues nuestros enlaces tendrán costes distintos en función de la comunicación entre los actores, por lo que no podremos definir un máximo teórico para el denominador de esta fórmula. Alternativamente para medir la centralización de la red en cuanto a esta medida de centralidad podremos utilizar la varianza de los datos.

Centralidad de intermediación

La centralidad de intermediación como su nombre indica mide el poder o centralidad de un individuo basándose en sus apariciones en la ruta o camino geodésico entre pares de actores. De esta forma cuanto más frecuentemente aparezca en las rutas geodésicas entre pares de actores mayor importancia tendrá dentro de la red.

Se entiende entonces que si un actor está en medio de la comunicación de otros, este hace de intermediario, pudiendo conectarles y permitiendo o no el intercambio de información entre ellos, los actores dependen de este para comunicarse y el intermediario tiene la posibilidad de aprovechar los recursos que otros comparten sin ser estrictamente el destinatario, simplemente aprovechándose de su ventajosa posición.

Sin embargo, se puede dar la posibilidad de que exista más de un camino geodésico entre dos actores, si es así los intermediarios de una ruta no tienen por qué estar en otras rutas, lo que haría que estos perdieran poder.

Podríamos pensar también que los actores no tienen por qué comunicarse únicamente mediante los caminos geodésicos, si no que buscarán todas las rutas posibles para conectar aunque estas no sean las más rápidas o eficientes. Tenemos por tanto otra medida de centralidad denominada centralidad de flujo, que mide la frecuencia de aparición de un actor entre pares de actores que están simplemente conectados.

Esta mueva medida suma cuán involucrado está un actor en el flujo entre todos los pares de actores, se intuye por tanto que la cantidad de cálculos solamente para medir la intermediación de un nodo de la red es enorme, y que tanto la magnitud de los cálculos como la magnitud de la medida se incrementará con el tamaño y densidad de la red.

Nosotros nos centraremos en la centralidad de intermediación, para la cual se define la siguiente ecuación, en la que $N_{jk}(n_i)$ es el camino geodésico entre dos nodos en el que se encuentra el nodo n_i, y N_{jk} representa el número de caminos geodésicos entre dos nodos, que normalmente será 1.

$$C_I(n_i) = \frac{\sum_{j<k} N_{jk}\,(n_i)}{N_{jk}}$$

Con la fórmula de centralización de la red con respecto a la medida de intermediación nos ocurre lo mismo que con la medida de centralización de cercanía, por lo que continuaremos utilizando también en este caso la varianza como medida representativa de la variabilidad de la centralidad de intermediación en el conjunto de la red.

2.2.4 "Shortest Path", Algoritmo de Dijkstra

Como ya hemos comentado anteriormente, para obtener las medidas de centralidad de cercanía e intermediación necesitamos conocer los caminos más cortos entre pares de actores de la red. Surge la necesidad de utilizar algoritmos que nos ayuden a determinarlos.

El algoritmo de Dijkstra determina la ruta más corta desde un nodo origen hacia los demás nodos, para ello es requerido como entrada un grafo cuyas aristas posean pesos.

Algunas consideraciones:

Si los pesos de las aristas son de valor 1, bastará con usar el algoritmo Breadth First Search (BFS) o algoritmo de búsqueda por anchura en un grafo.

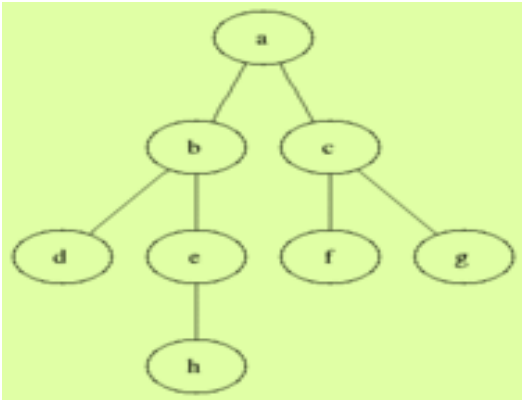


Figura 2.11: *Ejemplo algoritmo BFS*

BFS va formando un árbol a medida que va recorriendo un grafo, todo parte de un nodo inicial que será la raíz del árbol que se forma, luego ve los adyacentes a ese nodo y los agrega en una cola, la prioridad de la cola es FIFO (primero en entrar es el primero en salir), así que los siguientes nodos a evaluar serán los adyacentes previamente insertados.

Si los pesos de las aristas son negativos no se puede usar el algoritmo de Dijsktra, para pesos negativos existe otro algoritmo llamado Algoritmo de Bellmand-Ford.

El algoritmo de Dijkstra funciona de forma parecida al algoritmo de búsqueda por anchura, la diferencia es que ahora las aristas aunque con valores positivos pueden o no tener el mismo peso o coste, por lo que la "cola" que se forma será una cola con prioridad, en este caso prevalecerá el nodo con el menor coste.

El algoritmo parte de un vértice origen, a partir de ese vértice evaluaremos sus adyacentes, utilizando la técnica "greedy": utiliza el principio de que para que un camino sea óptimo, todos los caminos que contiene también deben ser óptimos, entre todos los adyacentes buscamos el que está más cerca del origen y evaluamos si podemos llegar más rápido desde este vértice a los demás. Después escogemos al siguiente más cercano (con las distancias ya actualizadas) y repetimos el proceso. Esto lo hacemos hasta que el vértice no utilizado más cercano sea nuestro destino. Al proceso de actualizar las distancias tomando como punto intermedio al nuevo vértice se le conoce como relajación, [15].

2.3 Almacenamiento de datos

La consulta lanzada a la plataforma Facebook nos arroja un conjunto de datos que analizaremos en repetidas ocasiones, para evitar lanzar la consulta cada vez que necesitemos disponer de los datos, necesitaremos almacenarlos.

No solo con el fin de disponer de los datos, sí no también de tenerles categorizados y estructurados necesitaremos disponer de una base de datos.

La base de datos nos proporciona el acceso a estos de forma más rápida y eficaz, su ordenación, y una visualización sencilla de los mismos. Utilizaremos para ello el sistema de gestión de bases de datos MySQL.

2.3.1 MySQL

MySQL es el sistema de gestión de bases de datos “open source” SQL más popular, desarrollado, distribuido y soportado por la compañía Oracle. Es un sistema de gestión de bases de datos relacionales.

Figura 2.12: *Icono del sistema de gestión de bases de datos*

Una base de datos relacional almacena los datos en tablas separadas en lugar de poner todos los datos en un gran almacén. Las estructuras de la base de datos se organizan en archivos físicos optimizados para la velocidad. El modelo lógico, con objetos como bases de datos, tablas, vistas, filas y columnas, ofrece un entorno de programación flexible.

La parte SQL de "MySQL" significa "Structured Query Language". SQL es el lenguaje estandarizado más común utilizado para acceder a las bases de datos. SQL está definido por el estándar ANSI / ISO SQL. El estándar SQL ha estado evolucionando desde 1986 y existen varias versiones.

Open Source significa que es posible que cualquiera use y modifique el software. Cualquiera puede descargar el software MySQL desde Internet y usarlo sin pagar nada. Se puede estudiar el código fuente y cambiarlo para adaptarlo a tus necesidades. El software MySQL utiliza la licencia GPL (General Public License), para definir lo que puede o no hacer con el software en diferentes situaciones. Existen también versiones con licencia comercial.

MySQL Server fue desarrollado originalmente para manejar grandes bases de datos mucho más rápido que las soluciones existentes y ha sido utilizado con éxito en entornos de producción altamente exigentes durante varios años. Aunque en constante desarrollo, MySQL Server ofrece hoy un rico y útil conjunto de funciones. Su conectividad, velocidad y seguridad hacen que MySQL Server sea muy adecuado para acceder a bases de datos en Internet. [16].

Algunas de sus características, [17]:

- Escrito en una combinación de C y C++
- Testado con una amplia gama de compiladores
- Proporciona soporte para diferentes sistemas operativos: Windows, Linux, Ubuntu, Debian, IBM, Mac etc.
- Utiliza diseño de servidor multicapa con módulos independientes
- Aprovecha la potencia de sistemas multiprocesador, gracias a su implementación multihilo
- Proporciona motores de almacenamiento transaccionales y no transaccionales
- Utiliza un sistema de asignación de memoria basado en hilos muy rápido
- Implementa tablas hash de memoria, que se utilizan como tablas temporales
- Implementa funciones SQL utilizando una biblioteca de clases altamente optimizada que debe ser lo más rápida posible.
- Proporciona el servidor como un programa independiente para su uso en un entorno de red cliente / servidor y como una biblioteca que se puede incrustar (enlazar) en aplicaciones independientes
- Soporta gran cantidad de tipos de datos para las columnas
- Dispone de API's en gran cantidad de lenguajes (C, C++, Java, PHP, etc.)

2.4 Visualización de datos

En nuestro caso la cantidad de datos que se pueden generar en un grupo de Facebook es la mayoría de las veces demasiado extensa como para extraer conclusiones relevantes mediante análisis manual. Los datos en bruto solo se vuelven útiles cuando se aplican métodos que permitan obtener una visión de los mismos.

Pocos de nosotros podemos extraer o identificar patrones entre filas de números, texto o en general de datos, pero si somos capaces de interpretar por ejemplo gráficos de barras o diagramas de flujo, extrayendo de las representaciones visuales una visión general, o particular de los datos que se manejan.

Por esta razón la visualización de los datos es la forma más rápida y directa de comunicar y compartir la información con otros.

La visualización de los datos trabaja con reglas que interpretan los datos y expresan sus valores como propiedades visuales, es el proceso de mapear información textual hacia información visual.

Mapear la información requiere el uso de la programación para acelerar la tarea, permitiéndonos manipular grandes conjuntos de datos, y obtener rápidamente una retroalimentación que nos indique si el diseño de la visualización es el adecuado.

El sistema se encargará de ejecutar la visualización y nuestro papel es conceptualizar y escribir las reglas que este debe tomar para que con los datos que le aportamos genere la visualización.

Normalmente necesitaremos varias visualizaciones para lograr una visión útil y representativa de la información. En muchas ocasiones las vistas estáticas pueden no ser demasiado útiles, por ello las visualizaciones interactivas, donde el usuario puede seleccionar lo que desea ver e interactuar con la información enriquecen las conclusiones que podamos extraer, [18].

2.4.1 D3

D3 o D3.js es una librería de JavaScript que nos permite crear visualizaciones de datos. La abreviatura D3 hace referencia al nombre completo de la herramienta Data-Driven-Documents.

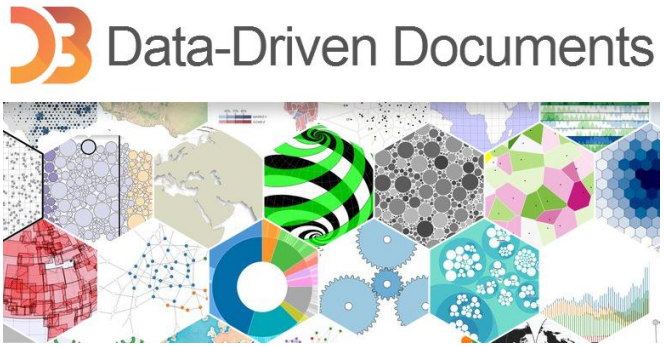


Figura 2.13: *D3, Data-Driven-Documents*

El autor de la librería es Mike Bostock aunque cuenta con aportaciones de otros desarrolladores. El proyecto es totalmente de código abierto y está disponible gratuitamente en GitHub.

D3 se lanzó bajo la licencia BSD (Berkeley Software Distribution), licencia de software libre, por lo que se permite utilizar y/o modificar código con fines comerciales y no comerciales.

Para generar visualizaciones con D3 necesitamos los datos o información que tienen que ser proporcionados por nosotros, normalmente en documentos de texto con el formato apropiado, herramientas basadas en tecnologías web tales como HTML, SGV y CSS, por último D3 se encarga de interconectar ambas partes y generar la visualización.

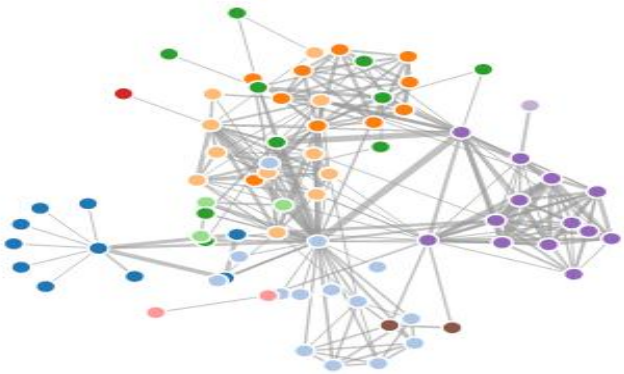


Figura 2.14: *Ejemplo de visualización con D3*

Fundamentalmente D3 es un software que facilita la generación y manipulación de documentos web con datos. Lo realiza cargando los datos en la memoria del navegador, enlazando los datos a elementos de página web y creando los nuevos elementos que necesite. Además permite hacer transiciones de elementos entre estados en respuesta a las interacciones del usuario.

La API de D3 está estructurada en una colección de módulos, estos módulos están diseñados para trabajar juntos. Aunque se pueden utilizar de forma independiente para realizar construcciones determinadas. Cada módulo permite un conjunto de operaciones y define una serie de métodos para cada operación. Por ejemplo: el módulo Arrays tiene definidas operaciones de búsqueda, manipulación, ordenación entre otras, en estas podemos encontrar una extensa lista de métodos.

D3 cuenta con multitud de ejemplos visuales, será nuestro papel valorar cuales se ajustan mejor al conjunto de datos con el que estemos trabajando.

A la hora de usar D3 debemos tener en cuenta que no soporta navegadores antiguos, su filosofía persigue la evolución de la web, y para ello incentiva a sus usuarios a disponer de las versiones más actuales de los navegadores.

Finalmente es importante saber que D3 no oculta la información. EL código de D3 se ejecuta en el lado del cliente de modo que para generar la visualización es necesario enviar los datos. Por lo tanto si no queremos mostrar los datos D3 no es el software más adecuado, [18].

3. Acondicionamiento de los datos

A continuación, se presentará la parte práctica correspondiente al trabajo de fin de grado.

En primer lugar se presenta el entorno de desarrollo desde el que se implementará la aplicación, conocido este, se explicará la estructura de la consulta a los servidores de

Facebook, una vez extraídos los datos se hablará de su organización y características, este paso es necesario antes de analizar los datos y determinar las posibles opciones de almacenamiento.

Por último se definen los pasos necesarios para volcar correctamente toda la información en el sistema de almacenamiento.

3.1 El entorno de desarrollo

Nosotros utilizaremos el IDE (Entorno de Desarrollo Integrado) Netbeans, una herramienta que nos permite compilar, depurar y ejecutar programas. Está escrito en Java, pero ofrece soporte a un amplio rango de tecnologías de desarrollo, entre otras: Java, PHP, C/C++, etc., y sistemas operativos: Windows, Linux, Mac OS etc. Además es un entorno de desarrollo gratuito y de código abierto.

En cuanto a la programación utilizaremos el lenguaje PHP, como producto PHP es un lenguaje de código abierto, y multiplataforma. Se puede utilizar para generar cualquier tipo de programa, aunque realmente es en el campo de la generación de páginas web donde más se utiliza.

Ahora que disponemos del entorno de desarrollo y del lenguaje de programación necesitamos un servidor de páginas web y un sistema de gestión de bases de datos. Esto nos lo ofrece el instalador XAMPP, un software libre y multiplataforma que nos proporcionará el servidor de páginas web Apache, PHP y MySQL.

3.2 La consulta a los servidores de Facebook

Para definir la consulta que queremos lanzar, primero debemos realizar un estudio de las posibilidades que ofrece Facebook. Ya mencionamos anteriormente que disponemos de una lista completa de nodos raíz a los que se puede acceder directamente. La documentación se puede consultar en el siguiente enlace:

https://developers.facebook.com/docs/graph-api/reference

Concretamente nuestro nodo raíz será el grupo de Facebook, accesible a través de su identificador. Nuestro grupo será de carácter privado, por lo que al obtener el token de acceso debemos marcar el permiso “user_managed_groups”, el cual permitirá leer el contenido del grupo de Facebook en el que el usuario sea administrador.

Continuamos explorando los campos accesibles desde el nodo raíz, en la descripción de los campos encontramos: el identificador del grupo, el nombre, su descripción, información sobre su foto e icono, dirección de email, el perfil del propietario del grupo, el número de solicitudes de acceso pendientes de confirmación, la configuración de privacidad etc., en definitiva parámetros que quedarán excluidos a la hora de realizar la consulta pues no aportan información relevante en nuestro estudio.

Examinando los edges a los que tenemos acceso desde el nodo “group” encontramos una extensa lista, de todos ellos nos quedaremos únicamente con el edge “feed”, que contiene la lista de todos los posts publicados en el grupo.

Un post es una entrada individual en el tablón o muro del grupo realizada por uno de sus miembros, para leerlos necesitamos marcar el permiso “user_posts” al generar el token de acceso. De cada post necesitaremos acceder a los siguientes campos:

- created_time: la fecha en la que se publicó el post
- message: el mensaje que se escribió en el post
- from: información sobre el perfil que publicó el mensaje

Y a los siguientes edges:

- reactions: contiene la lista de perfiles que reaccionaron al post y el tipo de reacción
- comments: contiene la lista de comentarios que ha recibido el post

Del edge “post/comments” necesitaremos los mismos campos que para el Post:

- created_time: la fecha en la que se publicó el comentario
- message: el mensaje que se escribió en el comentario
- from: información sobre el perfil que publicó el comentario

Y los mismos edges:

- reactions: contiene la lista de perfiles que reaccionaron al comentario y el tipo de reacción
- comments: contiene la lista de comentarios que ha recibido el comentario

Por último del edge “post/comments/comments” necesitaremos los campos:

- created_time: la fecha en la que se publicó el comentario
- message: el mensaje que se escribió en el comentario
- from: información sobre el perfil que publicó el comentario

Y el siguiente edge:

- reactions: contiene la lista de perfiles que reaccionaron al comentario y el tipo de reacción

La consulta tendrá como hemos especificado tres niveles de anidamiento, pues a partir de los comentarios de segundo nivel; es decir comentarios como respuesta a un comentario de un post, no existen más niveles de anidamiento, los comentarios se muestran secuencialmente. El diagrama de la consulta se muestra en la siguiente figura.

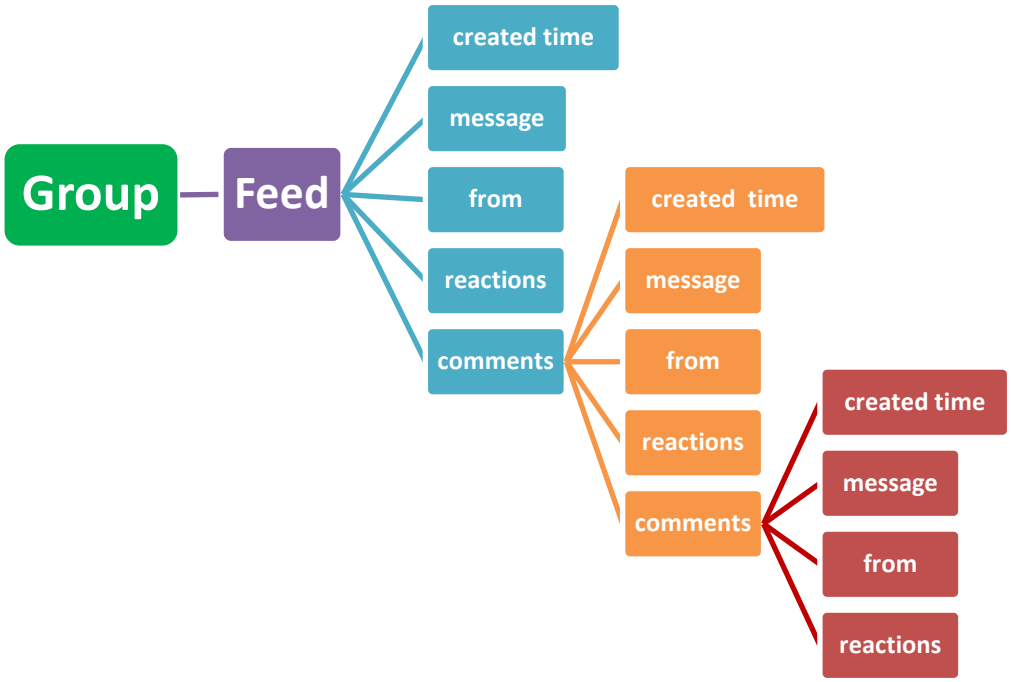


Figura 3.1: *Diagrama de la consulta a los servidores de Facebook*

La consulta se realizara mediante una solicitud GET al end-point relevante, en nuestro caso al grupo de Facebook, anidando todos los parámetros a los que necesitamos acceder para realizar solamente una llamada y así obtener los datos de forma eficiente.

3.3 Organización de los datos

Una vez realizada la consulta, el siguiente paso será el análisis de los datos para su posterior almacenamiento.

En esta fase diferenciaremos la información en tres niveles:

En un primer nivel tendremos los posts, estos como ya mencionamos son publicados por un miembro del grupo en el tablón o muro, por lo que no tienen un destinatario concreto pero serán visibles y notificados por Facebook al resto del grupo. Además representan la única forma de iniciar la comunicación entre los miembros del grupo, por lo que serán de vital y suma importancia.

En un segundo nivel tenemos los comentarios o replies que se realizan como respuesta a un post, al contrario que los posts estos si tienen destinatario concreto, este será la persona que escribió el post al que hace referencia el comentario. Lo mismo sucede con los rereplies, o comentarios como respuesta a un comentario, en este caso el destinatario será la persona que escribió el comentario o reply. En estos dos casos Facebook notifica su publicación a todos aquellos miembros que hubieran publicado en ese post anteriormente.

Por último un tercer nivel recoge las reacciones que cualquier información de los otros dos niveles pudiera recibir. Aquí el destinatario será obviamente la persona que escribió el post o comentario (reply o rereply) que genera la reacción. En este caso Facebook únicamente notifica la reacción al destinatario de la misma.

Como puede entenderse toda la información extraída no tiene la misma relevancia ni la misma difusión, por eso es importante realizar esta distinción, pues si el objetivo es evaluar la participación e interacción entre los miembros del grupo debemos pensar en las distintas formas de comunicación que se pueden establecer para asignarlas mayor o menor importancia en base a criterios como: la visibilidad y difusión de la publicación, el aporte de información y el inicio de la comunicación.

En nuestro estudio decidimos ignorar los replies y rereplies cortos, refiriéndonos con cortos a los comentarios con menos de dos palabras, para evitar contabilizar comentarios que expresaran conformidad o no con una publicación sin añadir motivos, ya que no aportan información relevante en el desarrollo de la actividad.

En un principio se pensó también en ignorar los comentarios o reacciones que una persona se pudiera hacer a sí misma, pero sí observamos la disposición de los rereplies (se colocan secuencialmente sin anidación), podríamos estar obviando información útil, pues puede que una persona escriba un rereply en un reply suyo como respuesta a otro miembro que este participando del tema.

Del mismo modo si pensamos en ignorar los comentarios que un miembro realice como respuesta a un post publicado por el mismo podríamos perder información relevante, pues puede que simplemente quiera añadir contenido a su post o puede que esté abriendo otra línea de discusión referente al tema de su publicación como post.

Tampoco se ignoraron finalmente las reacciones que una persona pudiera generar en sus propias publicaciones, pues las reacciones no tendrán el mismo peso que los posts o los comentarios, así que no distorsionarán demasiado los resultados y sí podrán formar parte de medidas de participación generales.

3.4 Almacenamiento de los datos

Una vez realizada la consulta, analizados y organizados los datos pasamos a su almacenamiento en una base de datos mediante MySQL.

La base de datos nos aportará rapidez y eficacia en cuanto a la disponibilidad de los datos, evitándonos realizar consultas a Facebook cada vez que queramos modificar o utilizar la información. Nos garantiza el acceso a los datos aunque no tengamos disponible conexión a red. Posibilita la ordenación y estructuración de la información en base a nuestros requerimientos y nos permite realizar búsquedas de forma ágil y eficiente.

Para ello, crearemos una tabla en la que almacenaremos toda la información que hemos catalogado como útil en el apartado anterior. La información se dispondrá en filas y columnas.

La elección de las columnas que integrarán la tabla muestra la necesidad de almacenar no solo la información estructurada y sus características, sino también quienes están relacionados con ella.

La siguiente imagen muestra la estructura de la tabla:

#	Nombre	Tipo	Cotejamiento	Atributos	Nulo	Predeterminado	Extra
1	**ID**	int(11)			No	*Ninguna*	AUTO_INCREMENT
2	**Groupname**	varchar(255)			No	*Ninguna*	
3	**FromName**	varchar(255)			No	*Ninguna*	
4	**ToName**	varchar(255)			Sí	*NULL*	
5	**TypeInteraction**	varchar(255)			Sí	*NULL*	
6	**WrittenMessage**	longtext			Sí	*NULL*	
7	**InteractionDate**	timestamp		on update CURRENT_TIMESTAMP	No	CURRENT_TIMESTAMP	ON UPDATE CURRENT_TIMESTAMP

Figura 3.2: *Estructura de la tabla*

Esta dispone de siete columnas, la primera columna "ID" está definida como un entero y es un valor que no puede ser nulo, representa un identificador único para cada fila de la tabla definido por SQL como PRIMARY KEY o clave primaria, cada tabla solo puede tener una clave primaria.

La segunda columna "Groupname" representa el nombre del grupo, este parámetro es bastante importante, pues se utiliza para conocer la información de la base de datos que pertenece a él. Esta columna será el modo de discernir que filas de la tabla pertenecen a un grupo determinado.

Otorga mayor flexibilidad a nuestra aplicación, dándonos la posibilidad de introducir en una misma tabla información sobre diversos grupos, y permitiéndonos la selección de los datos de un grupo concreto. Con esta variable, conseguimos que nuestra aplicación sea multigrupo.

Las columnas 3 y 4 contienen los nombres de quien realiza una la publicación "FromName" y hacia quien va dirigida "ToName", representando la existencia de comunicación entre pares de miembros del grupo, y el sentido de la misma.

El tipo de interacción "TypeInteraction" o quinta columna da a conocer el tipo o nivel de comunicación, distinguiremos seis tipos:

- Post : publicación de un post
- Reply : publicación de un comentario a un post
- Rereply : publicación de un comentario a un comentario de un post
- PostReaction : reacción ante un post
- ReplyReaction : reacción ante un reply o comentario
- RereplyReaction : reacción ante un rereply

Las columnas “Groupname”, “FromName”, “ToName” y “TypeInteraction” están definidas con el tipo VARCHAR (size), utilizado para cadenas de longitud variable (puede contener letras, números y caracteres especiales). El tamaño máximo se especifica entre paréntesis. Puede almacenar hasta 255 caracteres.

La sexta columna o “WrittenMessage” almacenará el texto que se escribe en cada publicación, si el tipo de interacción es una reacción, esta columna representará el tipo de reacción, será una cadena predefinida por Facebook, actualmente existen: NONE, LIKE, LOVE, WOW, HAHA, SAD, ANGRY y THANKFUL. Esta columna está definida con el tipo LONGTEXT, que soporta cadenas con una longitud máxima de 4.294.967.295 caracteres.

La última columna, “InteractionDate”, almacena la fecha en la que se realizó la publicación mediante el tipo TIMESTAMP cuyo formato es YYYY-MM-DD HH:MI:SS. En el caso especial de las reacciones Facebook no guarda la fecha, por lo que en nuestra tabla aparecerán con un valor nulo representado por ceros, 0000-00-00 00:00:00.

Las siguientes figuras muestran ejemplos de la información que se recoge en la tabla:

ID	Groupname	FromName	ToName	TypeInteraction	WrittenMessage	InteractionDate
19	FatimaGraph	Andrea Manzanares	Helena Llorente Franco	PostReaction	LIKE	0000-00-00 00:00:00

Figura 3.3: *Ejemplo de reacción a un post*

ID	Groupname	FromName	ToName	TypeInteraction	WrittenMessage	InteractionDate
25	FatimaGraph	Gloria Gonzalez	Helena Llorente Franco	Reply	En cuanto tenga un ratin lo veo	2017-01-16 06:01:15

Figura 3.4: *Ejemplo de reply*

ID	Groupname	FromName	ToName	TypeInteraction	WrittenMessage	InteractionDate
163	FatimaGraph	Gloria Gonzalez		Post	Hola guapitas!! Que tal la vuelta a la rutina??	2017-01-09 06:01:37

Figura 3.5: *Ejemplo de Post*

Cabe destacar que el flujo de comunicación con la base de datos está implementado de forma que evita la redundancia de información. Esto se consigue actualizando al completo la información almacenada en la base de datos cada vez que iniciamos sesión en la aplicación con un determinado grupo. Está actualización se lleva a cabo eliminando el contenido existente del grupo en la tabla, realizando la consulta a Facebook y analizando los datos para finalmente almacenarlos de nuevo.

Quizás este sistema de actualización podría llegar a ser ineficiente, pero no hay otra opción, pues la gran velocidad a la que los contenidos se generan, modifican o eliminan en las redes sociales hace que en cuestión de minutos o pocas horas nuestra base de datos, que para nosotros es la principal fuente de información, este desactualizada, lo que nos llevaría a extraer conclusiones desacertadas la mayoría de las veces.

4. Modelado y análisis de los datos

Con la información ya disponible para trabajar con ella en la base de datos, podemos comenzar a tratarla generando nuevas propiedades y estableciendo relaciones. Esto nos ayudará a modelar el conjunto de datos y posibilitará el análisis y creación de medidas de participación y centralidad con el fin de evaluar a los miembros del grupo de Facebook.

4.1 GraphViz

Durante la implementación de la aplicación, y concretamente en la fase que nos encontramos era necesario el uso de alguna herramienta que nos permitiese observar si nuestra implementación iba por el camino correcto, ofreciéndonos una retroalimentación necesaria para continuar el desarrollo.

En este contexto utilizamos GraphViz, una aplicación de visualización de gráficos de código abierto. Incluyendo su librería en el proyecto y mediante su interfaz Web fuimos observando la generación de los grafos que necesitábamos y comprobando si estos se estructuraban conforme a nuestras necesidades.

En lo sucesivo, y para enriquecer las explicaciones incluimos figuras que aluden a la utilización de esta herramienta.

4.2 La matriz de adyacencia y los grafos

Para tratar con la información que tenemos almacenada en la base de datos, utilizaremos la matriz de adyacencia y la generación de grafos.

Ambas estructuras recogerán las interacciones o comunicación entre los miembros del grupo, en cuanto a visualización la matriz de adyacencia representa la información de manera más limpia y clara que los grafos, más aún cuando el conjunto de miembros es grande o la red está muy acoplada (existe gran número de conexiones), aunque hablando de eficiencia los grafos permiten el acceso y tratamiento de la información de forma mucho más ágil que la matriz.

La matriz y el modelo de grafo tendrán las siguientes características:

1. Tendrá tantas filas (nodos para el grafo) como miembros participantes en el grupo de Facebook, cabe destacar la posibilidad de que existan miembros en el

grupo que no participen por lo que estos no aparecerán en la matriz (o grafo), cada fila (nodo del grafo) tendrá tantas columnas (enlaces provenientes de otros nodos) como miembros se hayan comunicado con el miembro que representa la fila. Será una matriz (o grafo) de tamaño variable, en función de las comunicaciones entre los miembros.

2. Los valores que almacenarán reflejarán no solo la existencia o no de comunicación entre pares de miembros sino también la calidad de la comunicación, en base al número de veces que se han comunicado y al tipo de comunicación.

3. Las filas no puede tener una columna que contenga el mismo nodo que representa la fila, reflejando que un usuario no puede generar relación al comunicarse consigo mismo. En el caso del grafo un nodo no puede tener un enlace hacia el mismo.

Teniendo clara su estructura, es momento de definir la información que contendrá. Necesitaremos toda la información de la base de datos a excepción de:

- Aquellas filas que representan comunicación de un miembro consigo mismo, es decir, aquellas filas donde las columnas FromName y ToName contengan la misma cadena.

- Las filas que representen la publicación de posts, aquellas en las que la columna TypeInteraction contenga la cadena Post, pues un post ya dijimos que no tiene destinatario concreto.

Ahora debemos asignar los valores numéricos a las interacciones dependiendo del tipo de interacción, para representar la importancia o nivel de la comunicación. A los comentarios (replies o rereplies) les asignaremos un valor numérico de 10, mientras a cualquiera de los tres tipos de interacción que representan reacciones les asignaremos la quinta parte del valor de los comentarios, un valor de 2.

Vamos a representar el grafo de las interacciones mediante GraphViz. Este será un grafo orientado y ponderado, en este caso los valores de los enlaces representan el flujo de comunicación entre los miembros del grupo.

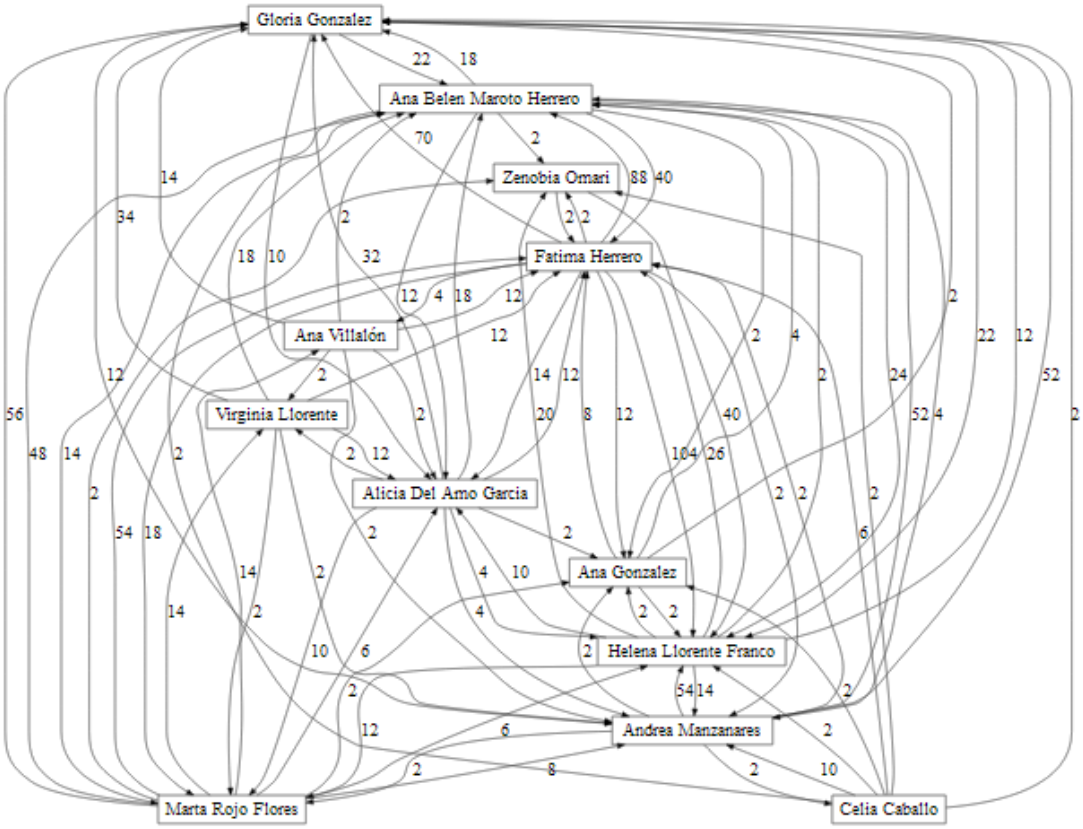


Figura 4.1: *Grafo ponderado y dirigido de las interacciones*

Aunque es cierto que visualmente el grafo anterior no representa la información de forma clara, si permite un acceso a los datos de forma rápida, además de posibilitar el uso de algoritmos para examinar la información, característica que no encontramos en la matriz de adyacencia. Por esta razón haremos uso de ambos para el cálculo de las medidas de centralidad que se desarrollan a continuación.

4.3 Medida de centralidad basada en grados

Tanto trabajando desde la matriz de adyacencia, como si lo hacemos desde la información representada en forma de grafo, estamos en disposición de calcular la centralidad basada en grados (salida y entrada), en este cálculo prescindiremos de los valores de los enlaces, únicamente nos interesa examinar la existencia de comunicación, sin importar su valor.

Desde la matriz de adyacencia obtenemos el grado de entrada “número de nodos que se comunican con un nodo dado” simplemente haciendo un recuento del número de columnas que tiene su fila. Para obtener el grado de salida “numero de nodos con los que un nodo dado se comunica” tendríamos que examinar todas las columnas de todas las filas a excepción de la fila del nodo y contar en cuantas de estas filas aparece como columna el nodo que se está examinando.

Desde el grafo deberíamos analizar por cada nodo el número de enlaces que este inicia, “grado de salida” y el número de enlaces que concluyen en él, “grado de entrada”.

Recogemos los valores obtenidos en nuestro estudio en la siguiente figura.

	Nombre	Grado Entrada	Grado Salida	Porcentaje Grado de Entrada	Porcentaje Grado de Salida
1	Gloria Gonzalez	10	4	90.91%	36.36%
2	Marta Rojo Flores	6	10	54.55%	90.91%
3	Fatima Herrero	10	9	90.91%	81.82%
4	Helena Llorente Franco	9	8	81.82%	72.73%
5	Andrea Manzanares	9	7	81.82%	63.64%
6	Ana Belen Maroto Herrero	10	8	90.91%	72.73%
7	Alicia Del Amo Garcia	7	8	63.64%	72.73%
8	Zenobia Omari	5	2	45.45%	18.18%
9	Celia Caballo	1	7	9.09%	63.64%
10	Virginia Llorente	3	7	27.27%	63.64%
11	Ana VillalÃ³n	3	6	27.27%	54.55%
12	Ana Gonzalez	7	4	63.64%	36.36%

Figura 4.2: *Centralidad de grados (entrada y salida)*

La figura anterior recoge la tabla que representa los cálculos de la medida de centralidad basada en grados para el conjunto del grupo. De ella podemos extraer que los actores definidos por los números 1,3 y 6 están a la cabeza en cuanto a grado de entrada, expresando el deseo de otros participantes de comunicarse con ellos, lo que implica que estos tres actores sean los más prestigiosos, sus posiciones son valiosas pues podrán hacer el papel de intermediarios posibilitando la comunicación y la dispersión de los recursos dentro de la red. En cuanto a grados de salida son los actores 2 y 3 los que se encuentran a la cabeza, estos actores se pueden considerar los más influyentes pues expresarán y compartirán sus opiniones con un gran número de miembros. El actor 3 es el que mayor grado de salida y entrada tiene indicando por tanto que además de ser un miembro de prestigio será también influyente, su posición será la más importante y valiosa.

Las dos últimas columnas de la tabla recogen la medida estandarizada con respecto al máximo de grados de entrada y salida que en cualquier caso será el número de nodos de la red menos uno, pues podríamos querer comparar la centralidad basada en grados de dos personas que se encuentren en redes diferentes que tengan distinto tamaño o densidad.

También será importante establecer medidas estadísticas que reflejen la distribución de la medida de centralidad basada en grados entre los miembros del grupo. Estas medidas resultan interesantes ya que indican si los miembros del grupo son

homogéneos o heterogéneos en cuanto a la posición estructural que ocupan dentro de la red.

Se recogen en la siguiente figura algunas medidas de interés:

	Grado Entrada	Grado Salida	Porcentaje Grado de Entrada	Porcentaje Grado de Salida
Promedio	6.67	6.67	60.61	60.61
Desviacion Estandar	2.98	2.21	27.1	20.1
Varianza	8.89	4.89	734.62	404.04
Maximo	10	10	90.91	90.91
Minimo	1	2	9.09	18.18
Coeficiente de Variacion	44.72%	33.17%		

Figura 4.3: *Medidas descriptivas de la red en base a la centralidad de grado*

Vemos que de media los actores tienen un valor de 6.67 tanto en grados de entrada como de salida, buen valor dado que representa el 60.61% del valor máximo que podrían tener. En cuanto a los rangos de entrada y de salida, no existe mucha diferencia, (1-10) y (2-10). La desviación estándar y la varianza, indican que existe mayor variabilidad en los grados de entrada que en los de salida.

Claramente los nodos del grupo son más homogéneos en cuanto a grados de salida que a grados de entrada, ya que su rango es menor y su variabilidad más baja o lo que es lo mismo, son más homogéneos en cuando a influencia que a prominencia.

El coeficiente de variación nos ayuda a comparar redes de distinto tamaño, indicando la dispersión o variabilidad de la red en cuanto a esta medida. Volvemos a ver como existe menor dispersión en cuanto al grado de salida.

Por último damos las medidas de centralización de grados de la red, cuya fórmula fue definida en el apartado 2.2.3. Obtenemos los siguientes valores:

- Centralización de grado de entrada: 36.36 %
- Centralización de grado de salida: 36.36 %

Ambos valores revelan que la red está lejos de comportarse como la red en estrella, donde solo es uno es actor más influyente y prestigioso, aunque revela la existencia de actores más prestigiosos o más influyentes que otros, como ya hemos argumentado antes.

4.4 Medida de centralidad basada en cercanía

La medida de centralidad de cercanía evidencia que un actor puede no estar vinculado directamente con algunos actores, pero si puede llegar a comunicarse con ellos a

través de otros. Trata de medir lo cercano o lejano que se encuentra un actor con respecto al conjunto de la red.

Para esta medida necesitaremos disponer de un grafo que contenga como característica de los enlaces un valor que represente el coste entre nodos. Por esto inicialmente modelaremos el grafo de relaciones anteriormente expuesto, en el que los valores de los enlaces indican el flujo o nivel de comunicación entre nodos, para generar otro en el que se indique el coste de cada enlace.

Para llevar a cabo la transformación de los valores de los enlaces del grafo anterior, buscaremos el enlace con mayor valor (que representa el mayor flujo de comunicación), y le sumaremos uno. Ahora restaremos a este número el valor de cada enlace, obteniendo así una correspondencia inversa entre el flujo o nivel de comunicación entre dos nodos y el coste del enlace. El enlace que en el grafo anterior presenta mayor flujo de comunicación ahora pasará a tener coste 1, del resto podemos decir que a menor flujo de comunicación, mayor será su coste. Volvemos en este caso a utilizar Grahpviz para asegurarnos que el nuevo grafo se haya generado correctamente, manteniendo su estructura pero representando costes en vez de flujo de la comunicación.

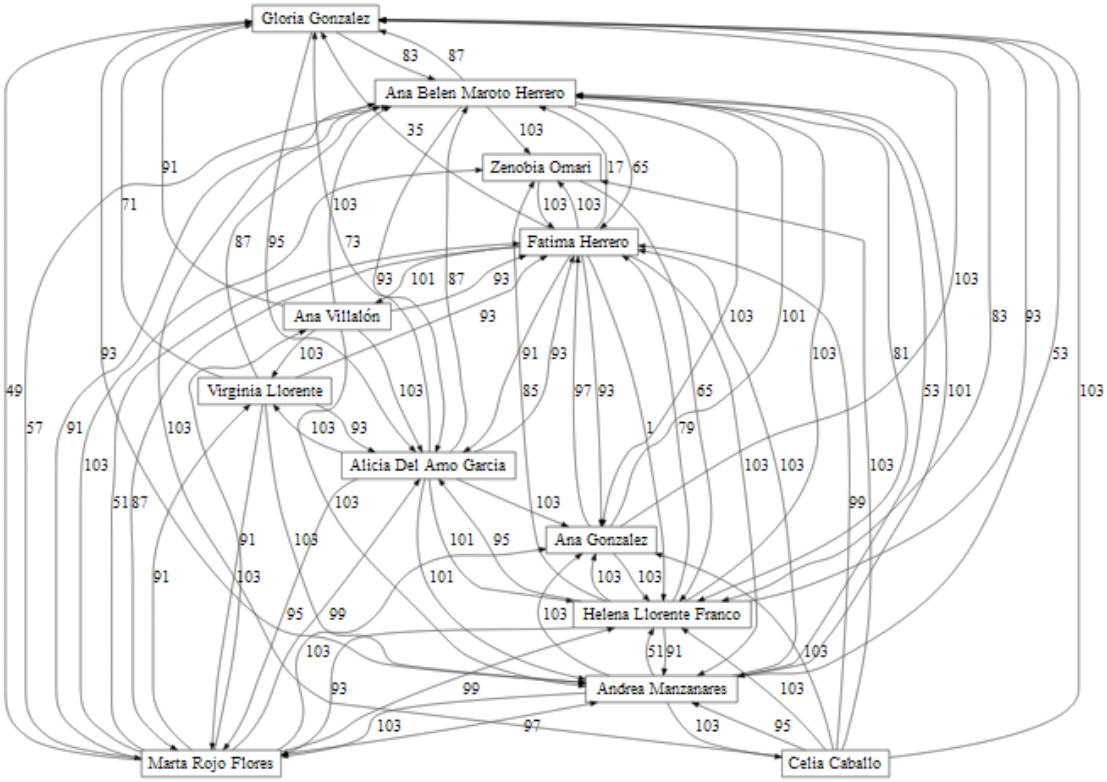


Figura 4.4: *Grafo dirigido representando los costes de los enlaces*

La generación de este grafo es necesaria para utilizar el algoritmo de Dijkstra y obtener así un grafo para cada nodo que represente los caminos con menor coste o caminos

geodésicos del nodo al resto de la red, la técnica usada para hallar los costes de los enlaces se ha utilizado con el fin de no generar enlaces con costes negativos, ya que el algoritmo de Dijkstra solo trabaja con números positivos.

Volvemos a utilizar Grahpviz para asegurarnos de que el algoritmo funciona según lo esperado. Representamos a modo de ejemplo el grafo de Dijkstra para uno de los miembros del grupo.

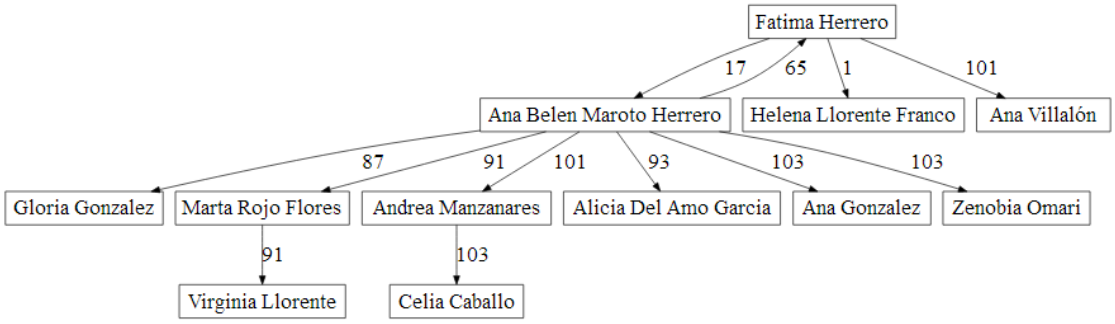


Figura 4.5: *Grafo de Dijkstra para el nodo "Ana Belén Maroto Herrero"*

Como se puede observar el algoritmo de Dijkstra nos dará también el camino geodésico hacia uno mismo en caso de que exista. Este deberá de ser excluido a la hora de establecer las medidas de cercanía e intermediación.

Para el cálculo del índice de cercanía necesitamos primero calcular las distancias geodésicas (suma de los costes de los enlaces de un camino geodésico) de un nodo a cada nodo que está conectado.

Teniendo en cuenta que un nodo puede no estar conectado a todos los nodos de la red deberemos asignar una distancia geodésica al camino inexistente. Nosotros la hemos modelado como el número de nodos de la red menos uno multiplicado por 100. Ahora estamos en disposición de realizar el cálculo de la lejanía haciendo la suma de todas las distancias geodésicas.

Tomando la inversa de la lejanía y dividiéndola entre el número de nodos menos uno estaremos calculando el índice de cercanía para cada miembro del grupo. Los resultados obtenidos se muestran en la tabla que recoge la siguiente figura, indicar que los valores de cercanía expuestos han sido trasladados a un rango de valores más altos para facilitar comparaciones.

	NOMBRE	Lejania	Cercania
1	Ana Belen Maroto Herrero	1261	7.21
2	Fatima Herrero	976	9.31
3	Andrea Manzanares	1239	7.34
4	Marta Rojo Flores	993	9.15
5	Gloria Gonzalez	1673	5.43
6	Helena Llorente Franco	1293	7.03
7	Alicia Del Amo Garcia	1318	6.9
8	Celia Caballo	1559	5.83
9	Ana Gonzalez	1908	4.76
10	Zenobia Omari	1780	5.11
11	Virginia Llorente	1318	6.9
12	Ana VillalÃ³n	1441	6.31

Figura 4.6: *Centralidad de cercanía*

Podemos observar como los actores 2 y 4 son los más cercanos al conjunto de la red, pues su lejanía con respecto a esta es la menor, sus índices de cercanía se convierten en los más altos del conjunto. Se corrobora mediante estas observaciones que estos dos actores podrán extraer más del conjunto de la red, son los actores que mejores posiciones ocupan. En cuanto al actor 9 que representa todo lo contrario a los actores 2 y 4 se podría decir que es sin duda el más lejano al conjunto, por lo que tendrá mayores dificultades para obtener recursos y para comunicarse con el resto.

Si observamos la correspondencia de los actores 2 y 4 de esta medida, en la tabla de centralidad de grados, vemos que en aquella son los que mayores conexiones directas tienen, por lo que no sorprende que los resultados de esta medida se asemejen a los de centralidad de grados, debido a que un gran número de caminos geodésicos son conexiones directas. En otras redes se podrían observar mayores diferencias entre las dos medidas de centralidad.

La siguiente figura recoge las medidas estadísticas para el conjunto de la red a cerca de la medida de centralidad de cercanía.

	Cercania
Promedio	6.77
Suma	81.28
Desviacion Estandar	1.37
Varianza	1.87
Maximo	9.31
Minimo	4.76
Coeficiente de Variacion	20.18%

Figura 4.7: *Medidas descriptivas de la red en base a la centralidad de cercanía*

Aquí se manifiesta lo que veníamos argumentando, si observamos el valor de la varianza o de la desviación estándar con respecto al promedio vemos que la variabilidad de la medida es baja indicando que no existen actores muy superiores en cercanía al resto, lo mismo representa el rango de valores (mínimo y máximo).

El coeficiente de variación representa que la red está poco centralizada en cuanto a la cercanía como ya comentábamos, además nos aportaría la posibilidad de realizar comparaciones con respecto a la centralización en esta medida con otras redes.

4.5 Medida de centralidad basada en la intermediación

Para el cálculo de la centralidad de intermediación seguiremos necesitando los grafos de Dijkstra para cada miembro del grupo, así será sencillo contabilizar cuantas veces aparece cada nodo en el conjunto de todos los caminos geodésicos de la red.

	NOMBRE	Intermediacion	Intermediacion Normalizada
1	Ana Belen Maroto Herrero	4	3.64%
2	Fatima Herrero	31	28.18%
3	Andrea Manzanares	10	9.09%
4	Marta Rojo Flores	9	8.18%
5	Gloria Gonzalez	0	0%
6	Helena Llorente Franco	18	16.36%
7	Alicia Del Amo Garcia	1	0.91%
8	Celia Caballo	0	0%
9	Ana Gonzalez	0	0%
10	Zenobia Omari	0	0%
11	Virginia Llorente	0	0%
12	Ana VillalÃ³n	0	0%

Figura 4.8: *Centralidad de intermediación*

La columna de intermediación simplemente representa un recuento por persona de las veces que aparece en el medio de algún camino geodésico. Para tratar de dar una medida más representativa de los datos se ha añadido una segunda columna de intermediación pero ya normaliza. La normalización se ha llevado a cabo haciendo un recuento de todos los caminos geodésicos en los que un actor puede participar como intermediario, para ello del total de los caminos geodésicos que se forman en la red, de forma particular para cada persona se han descontado aquellos caminos en los que participa como nodo de inicio o nodo final.

Da la casualidad de que con nuestra red, todos pueden acceder a todos, por lo que el número de caminos en los que cada uno puede ser intermediario es el mismo.

A simple vista observamos que solamente la mitad de los participantes actúan como intermediarios, por tanto tendrán posiciones ventajosas con respecto a la otra mitad del grupo.

Entre los que actúan como intermediarios se puede observar diferencias significativas, siendo los actores representados en la figura por los números 2 y 6 los que más porcentaje de intermediación tienen. Por tanto sus posiciones serán entendidas como valiosas por el hecho de posibilitar la dispersión de los recursos en la red y hacer de puente para la comunicación entre pares de nodos.

	Intermediacion Normalizada
Promedio	5.53
Suma	66.36
Desviacion Estandar	8.46
Varianza	71.53
Maximo	28.18
Minimo	0
Coeficiente de Variacion	152.94%

Figura 4.9: *Medidas descriptivas de la red en base a la centralidad de intermediación*

Ya anunciábamos que la centralidad de intermediación estaba distribuida de manera desigual, esto se evidencia aún más analizando las medidas descriptivas de la red. Todas las medidas apuntan a una variabilidad bastante alta de la intermediación. La desviación típica es de 8.46, es alta con respecto al promedio, incluso lo supera, lo que se refleja también en el coeficiente de variación siendo superior al 100%.

4.6 Conclusiones

Como se ha visto durante el desarrollo de este punto del trabajo, la generación de grafos es una parte sumamente importante y necesaria para generar conocimiento a través de los datos, no tanto quizás en este sentido la matriz de adyacencia, que simplemente hemos usado como modo de almacenamiento temporal de datos.

Sobre las medidas de centralidad se entiende que cada una de ellas por separado no da una representación veraz de la importancia que tienen los actores dentro del grupo de Facebook, imposibilitando su evaluación. Por eso para tener una visión general y verdadera debemos examinar cada una de las medidas por separado para finalmente fusionarlas en una sola interpretación.

Una de las posibles opciones para evaluar a los miembros del grupo sería ponderar las posiciones de los participantes en las tres medidas, así cuanto más arriba en el ranking de cada medida estuviera el alumno mayor nota se le asignaría.

Las medidas de centralización de la red son interesantes a fin de comparar nuestra red con otras redes, en el caso de que nos ocupa un profesor podría comparar mediante estas la clase A y la clase B de la misma materia.

5. Visualización

Se presenta en este apartado la parte gráfica o visual del desarrollo de la aplicación. Consta de una página de inicio o registro, tras esta se muestra la primera de las visualizaciones que se corresponde con el apartado 5.2 de esta misma sección, en ella

y en el resto de las visualizaciones disponemos de una barra de selección que nos permite navegar entre las distintas visualizaciones.

Todas las visualizaciones pertenecen a la librería de grafos D3, nosotros nos centramos en la programación del contenido y la estructura de los archivos de datos necesarios con el fin de adaptar cada visualización a nuestros requerimientos.

5.1 Registro en la aplicación

Inicialmente al lanzar la aplicación en el servidor web nos encontramos la página de inicio o de registro:

Figura 5.1: *Página de inicio. Registro en la aplicación*

Los datos requeridos se presentan diferenciados en tres bloques:

Access to the aplication

Necesitamos identificar la aplicación que tenemos en Facebook para desarrolladores, como ya comentamos en el apartado 2.1.1 Crear la App, es necesario disponer de una aplicación para lanzar la consulta a los servidores de Facebook, en ese mismo

apartado mostramos como acceder al panel de la aplicación para rescatar los campos que en este punto se solicitan.

Como se puede ver en la imagen anterior necesitaremos el identificador de la aplicación, la clave secreta de la misma y la versión de la API con la que queremos trabajar.

Access to the Facebook group

El siguiente bloque nos permite identificar el grupo de Facebook sobre el que queremos realizar la consulta. Para ello necesitamos el identificador del grupo y un token de acceso.

En Facebook todos los nodos, edges o campos cuentan con un identificador, una clave que los representa de forma unívoca. Para obtener el identificador del grupo deberemos acceder al mismo desde Facebook y en la barra del navegador encontraremos el identificador del grupo al que estamos accediendo.

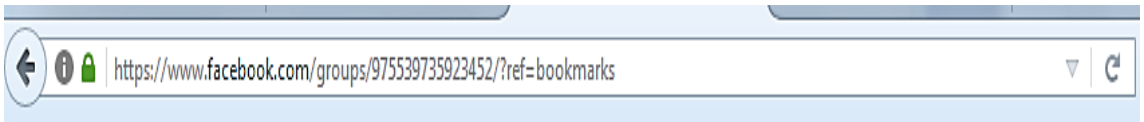


Figura 5.2: *Acceso al grupo de Facebook, obtención del identificador del grupo*

En cuanto al token de acceso, ya comentábamos en el apartado 2.1.3 API Graph que tenemos que acceder al explorador del API Graph y seleccionar la aplicación, en nuestro caso "Social Count" para obtener el token de acceso.

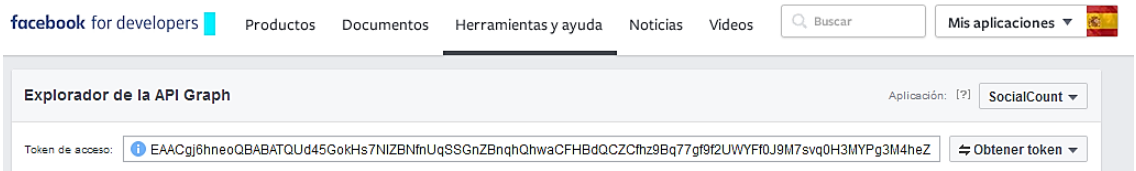


Figura 5.3: *Obtención del token de acceso*

Indicate de group's name

Por último para terminar el registro se nos pide que asignemos un nombre al grupo, este nombre es un parámetro bastante importante, pues se comportará como identificador de la información del grupo dentro del desarrollo, almacenamiento, extracción y visualización de los datos solicitados a Facebook.

5.1.1 Posibles casos de error en el registro

Teniendo en cuenta que habitualmente en los formularios de registro se introducen datos erróneos vamos a observar la respuesta de la aplicación ante los posibles casos.

- El token de acceso o Access token tiene un tiempo de validez limitado, por lo que puede ser que el token introducido haya expirado y tengamos que generar uno nuevo.
  La aplicación respondería con el siguiente aviso:

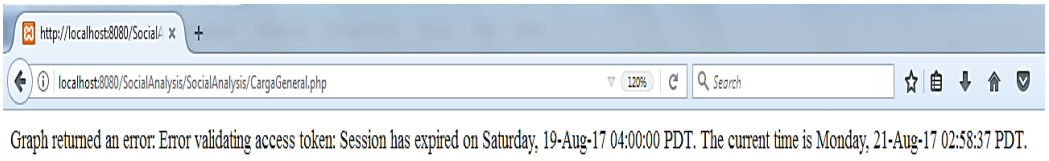


Figura 5.4: *Error token de acceso expirado*

En caso de que el token de acceso no fuese válido se generaría el siguiente aviso:

Figura 5.5: *Error token de acceso*

- Como el Access token se genera para un determinado objeto, en este caso el grupo, al introducir mal el identificador del grupo la aplicación muestra el siguiente error:

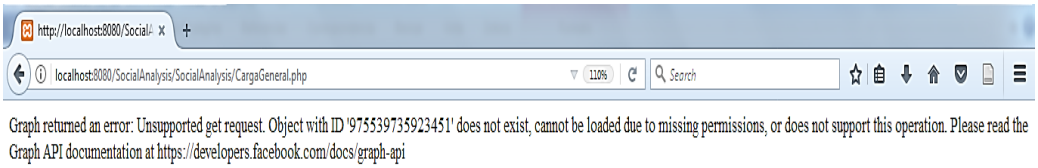


Figura 5.6: *Error el identificador introducido no existe*

- Introducir mal el identificador de la aplicación no genera error alguno, es la clave secreta de la aplicación "App password" la que necesariamente ha de ser correcta, pues es la que identifica la aplicación. Si esta no es correcta se muestra lo siguiente:

Figura 5.7: *error clave secreta de la aplicación no valida*

En cuanto a la versión de la API, simplemente debemos indicar una versión por su puesto existente y a poder ser actual, la lista de las versiones actualmente en uso podemos encontrarla en el apartado 2.1.3 del presente documento.

El nombre del grupo no genera ningún error, ya que es libre elección del usuario.

5.2 Estadísticas generales

Inmediatamente después de realizar el registro en la aplicación, se carga la primera visualización, un dashboard interactivo que ofrece estadísticas generales sobre la utilización del grupo de Facebook.

En todas las visualizaciones disponemos en la parte superior la barra que nos permite cambiar de página y por lo tanto de visualización. También se muestra el nombre del grupo que se está analizando.

Figura 5.8: *Nombre del grupo y barra de selección*

Puesto que en el desarrollo de las medidas de centralidad dejamos excluidos los datos referentes a la publicación de posts y todo tipo de publicación que un usuario realizase en sus propias publicaciones, aquí se incluirán.

Se trata de dar una visión general sobre todo lo publicado en el grupo sin realizar ninguna excepción.

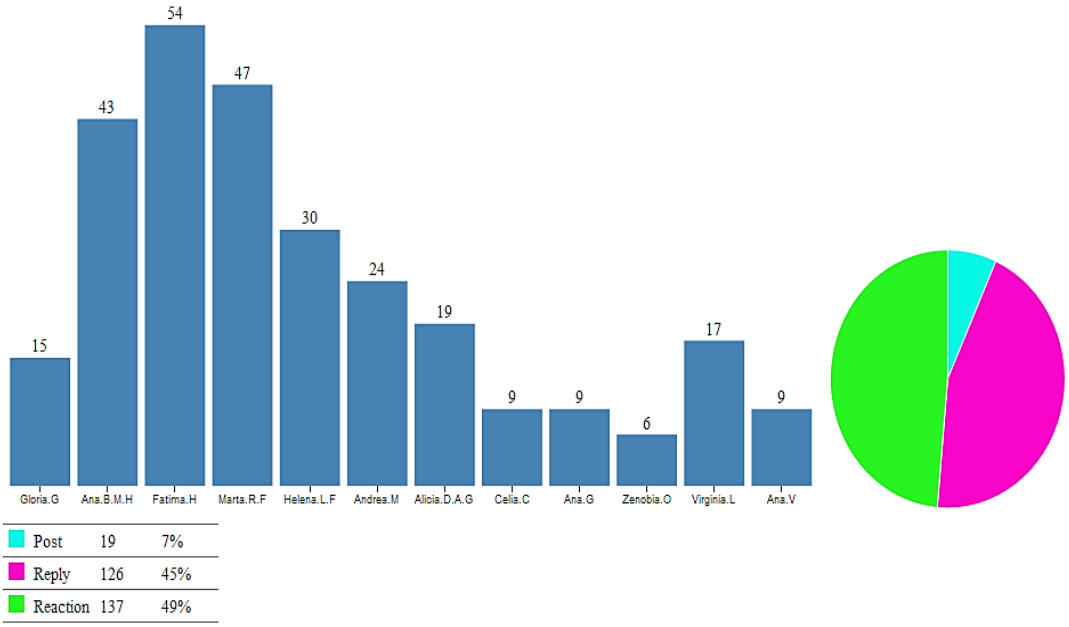


Figura 5.9: *Visualización estadísticas generales del grupo*

Desde esta perspectiva observamos que se han publicado 19 post, 126 repies o rereplies y 137 reacciones a publicaciones, con los porcentajes sobre el total de todas las interacciones. Los post representan el 7% de todo lo que se ha publicado, los comentarios el 45% y las reacciones el 49%. Las reacciones cuentan con el mayor porcentaje como era de esperar, debido a su inmediatez y facilidad.

En el pie del diagrama de barras se coloca el nombre de cada participante del grupo, definido mediante el nombre de pila completo y las iniciales de los apellidos. Cada barra representa por tanto a un participante, en la parte superior de cada una de ellas se coloca el número de publicaciones totales generadas por cada miembro.

El diagrama circular se divide en sectores mediante colores, cada sector simboliza un tipo de comunicación. Siempre representa los datos que aparecen en la leyenda. En esta imagen simboliza como se reparte la generación de publicaciones.

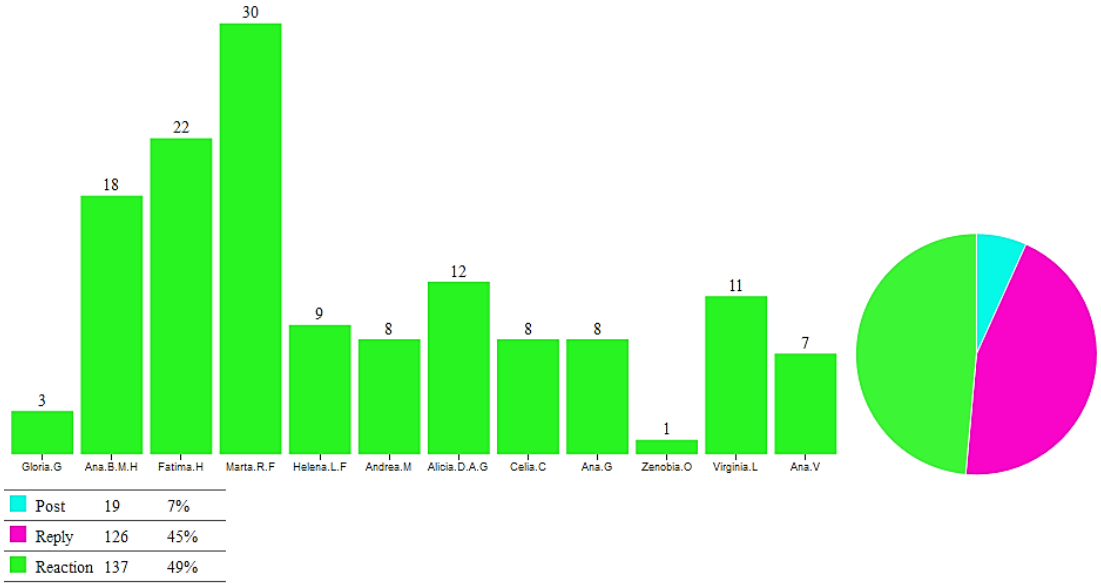


Figura 5.10: *Visualización de estadísticas en la publicación de reacciones*

Posicionándonos sobre cada sector, el diagrama de barras muestra en el mismo color que el sector seleccionado como se distribuyen por ejemplo para el caso de la figura 45 la publicación de reacciones entre los participantes, mostrando en lo alto de cada barra el número de publicaciones de este tipo para cada miembro.

Para finalizar en cuanto a los posibles modos de interactuar con la visualización, la siguiente imagen (figura 46) representa las estadísticas de un único participante. Tendríamos para ello que posicionar la flecha del ratón encima del participante que nos interese. En este caso su barra se vuelve de un color más claro al resto, la leyenda indica sus estadísticas y los porcentajes están referidos al total de sus propias publicaciones. También el diagrama de sectores representa los datos que se muestran en la leyenda.

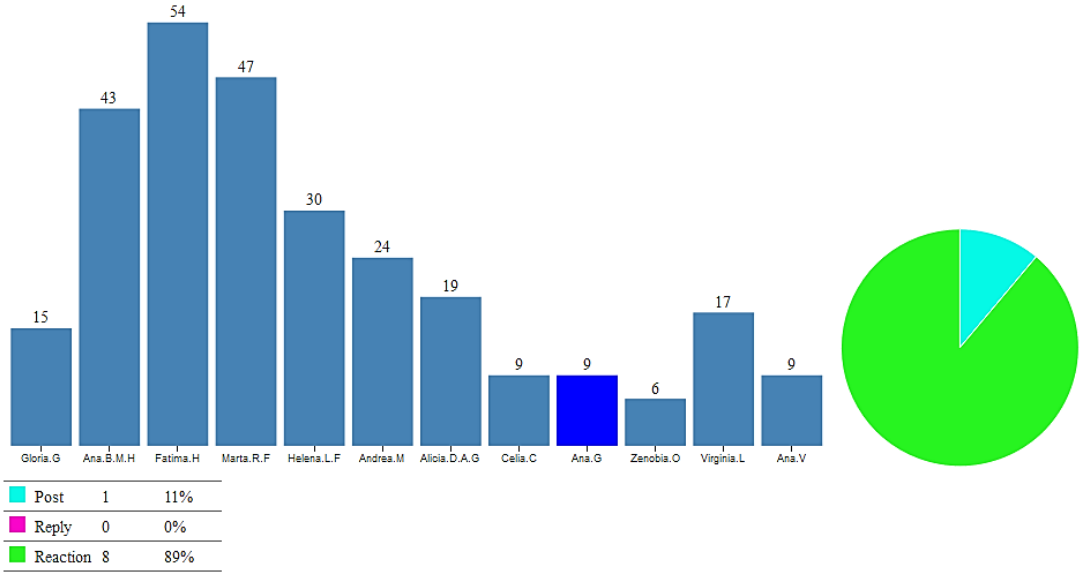


Figura 5.11: *Visualización de estadísticas para un participante*

Para completar las estadísticas generales, se genera también un archivo Excel en el que se presenta una tabla detallada de todas las estadísticas que se presentan en la visualización, pero con el porcentaje referido al total de cada tipo de publicación.

	NOMBRE	Post	% Post	Reply	% Reply	Reaction	% Reaction
1	Gloria.G	3	15.79%	9	7.14%	3	2.19%
2	Ana.B.M.H	4	21.05%	21	16.67%	18	13.14%
3	Fatima.H	4	21.05%	28	22.22%	22	16.06%
4	Marta.R.F	2	10.53%	15	11.9%	30	21.9%
5	Helena.L.F	2	10.53%	19	15.08%	9	6.57%
6	Andrea.M	1	5.26%	15	11.9%	8	5.84%
7	Alicia.D.A.G	1	5.26%	6	4.76%	12	8.76%
8	Celia.C	0	0%	1	0.79%	8	5.84%
9	Ana.G	1	5.26%	0	0%	8	5.84%
10	Zenobia.O	1	5.26%	4	3.17%	1	0.73%
11	Virginia.L	0	0%	6	4.76%	11	8.03%
12	Ana.V	0	0%	2	1.59%	7	5.11%

Figura 5.12: *Estadísticas generales con respecto al total de publicaciones*

En esta tabla podemos observar que los participantes 2 y 3 están a la cabeza del grupo en cuanto a la publicación de post y también de comentarios. En el caso de las reacciones solo les supera el actor 4. En general se puede considerar a los participantes 2 y 3 como los más activos dentro del grupo, pueden ser catalogados como actores independientes debido a la publicación de posts, es decir; no esperan a

comentar otras publicaciones, que también lo hacen, sino que toman las riendas y generan nuevos debates posibilitando así el desarrollo de la comunicación entre los miembros del grupo.

5.3 Centralidad de grados

Una segunda visualización muestra de forma interactiva la centralidad basada en grados de los miembros del grupo, representa en forma de estrella las conexiones directas entre los participantes.

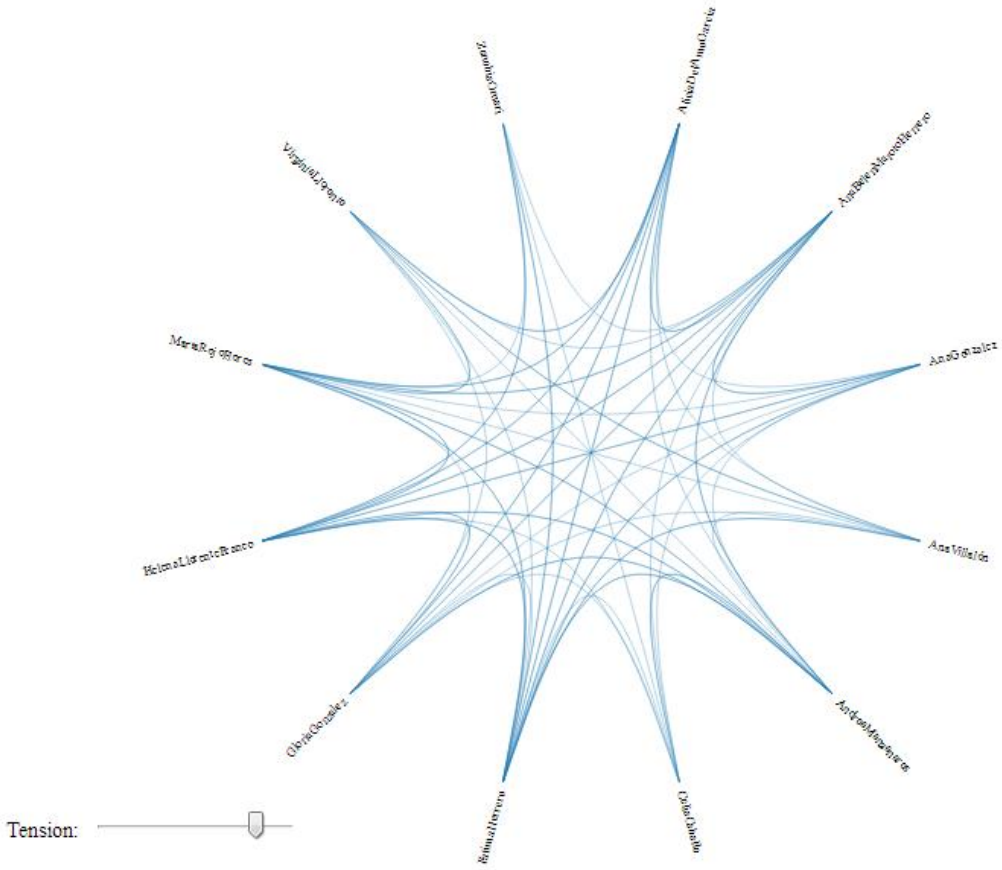


Figura 5.13: *Visualización centralidad de grados*

Dispone de un selector que modifica la tensión de los enlaces para mejorar su visualización.

Posicionándonos sobre cualquiera de los participantes sus enlaces cambian de color, el número de enlaces de color rojo representa el grado de entrada, los enlaces de color verde representan total (si no hubiera enlaces en rojo) o parcialmente los grados de salida del participante.

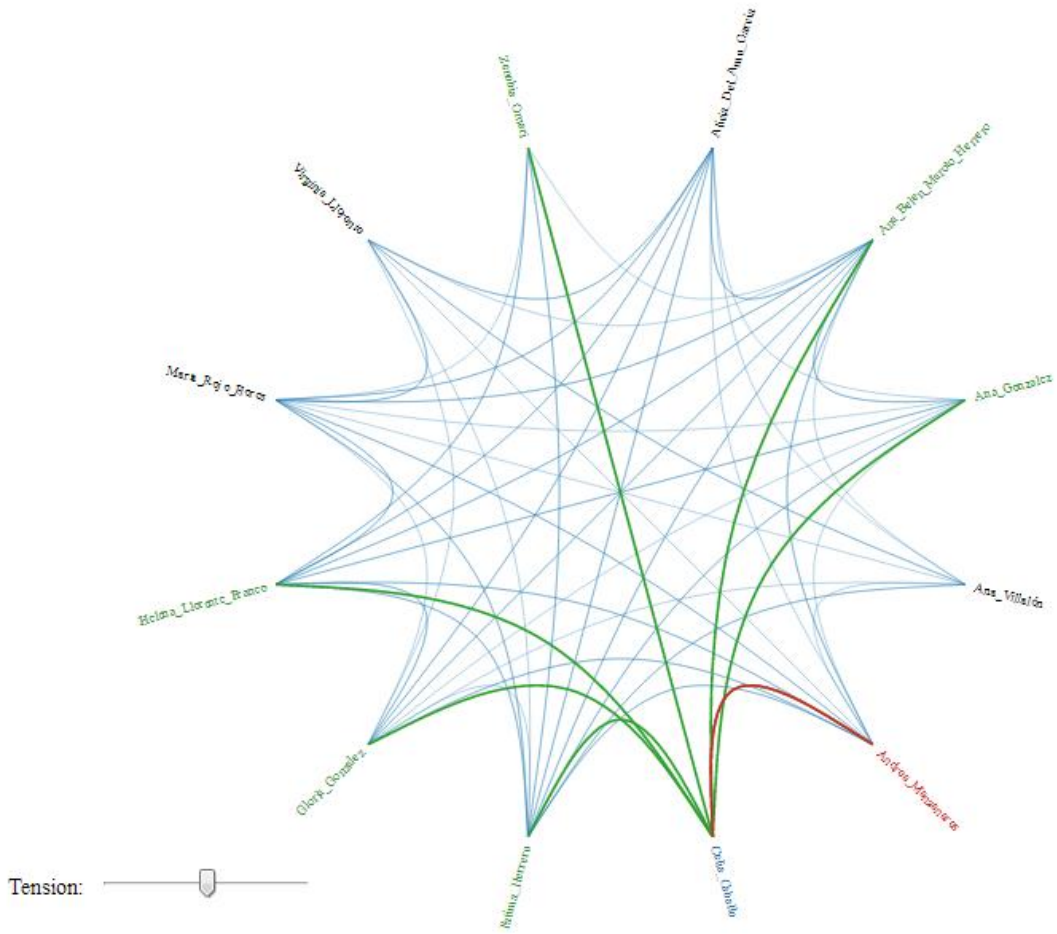


Figura 5.14: *Visualización centralidad de grados para un participante*

Este participante tendría grado de entrada 1, pues solo tiene un enlace en rojo, en cuanto al grado de salida, podemos afirmar que como mínimo será 6, pues tiene seis enlaces en verde indicando que él se comunicó con ellos pero ellos nunca lo hicieron con él, por lo que faltaría comprobar si el nodo del que recibe comunicación (enlace en rojo) también la recibe de este, que en este caso sí. Por tanto su grado de salida sería 7.

5.4 Cercanía

La siguiente visualización representada por un diagrama en forma de árbol muestra para cada participante del grupo los nodos o participantes con lo que puede comunicarse directa o indirectamente (a través de intermediarios). El número que acompaña a cada nodo es el coste del camino más corto para llegar hacia él. Los nodos están colocados de forma que aquellos cuyos caminos geodésicos tienen menor coste aparecen en la parte superior, hasta finalizar en aquel nodo con mayor coste.

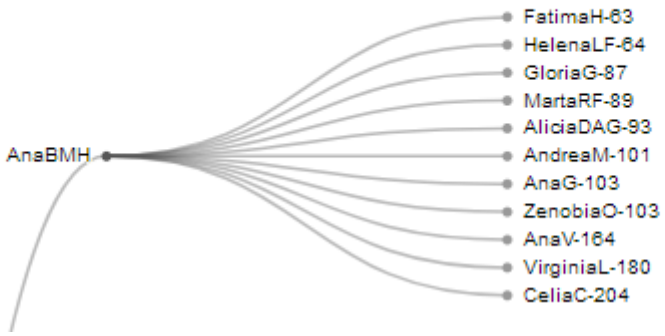


Figura 5.15: *Visualización centralidad de cercanía para un nodo de la red*

Viendo el árbol al completo se pueden establecer conclusiones sobre que nodos son los que poseen mayor número de conexiones, siendo estos más centrales en la red que otros, y pudiendo llegar a obtener más del conjunto de los recursos del grupo.

También si dos miembros del grupo están conectados al mismo número de nodos, haciendo la suma de los caminos podríamos saber quién de los dos supera al otro en cuanto a centralidad.

En cuanto a las posiciones que ocupan los nodos, se podría ver que si un nodo ocupa en la mayoría de las veces posiciones altas, será porque el nodo ha buscado relacionarse directamente con él, o bien porque recibe mejores conexiones que otros nodos, indicando entonces que es un nodo bien conectado.

Esta visualización cuenta con un desplegable de selección que permite ver desde el árbol de todo el grupo, solamente las conexiones de un participante o bien una única conexión entre dos miembros del grupo.

Cada vez que se realiza una selección se procede a la recarga de la página de manera dinámica, y el tamaño que ocupa se redimensiona para ocupar toda la página.

5.5 Grafos de Dijkstra

Los grafos de Dijkstra dan una visión similar a la del apartado anterior, indicando las conexiones de un cierto nodo por los caminos geodésicos al resto de nodos que está conectado, pero ahora visualizamos cuales de estos caminos geodésicos se hacen de forma indirecta y a través de quien.

También se permite la selección del nodo que nos interesa y la página se recarga de forma dinámica.

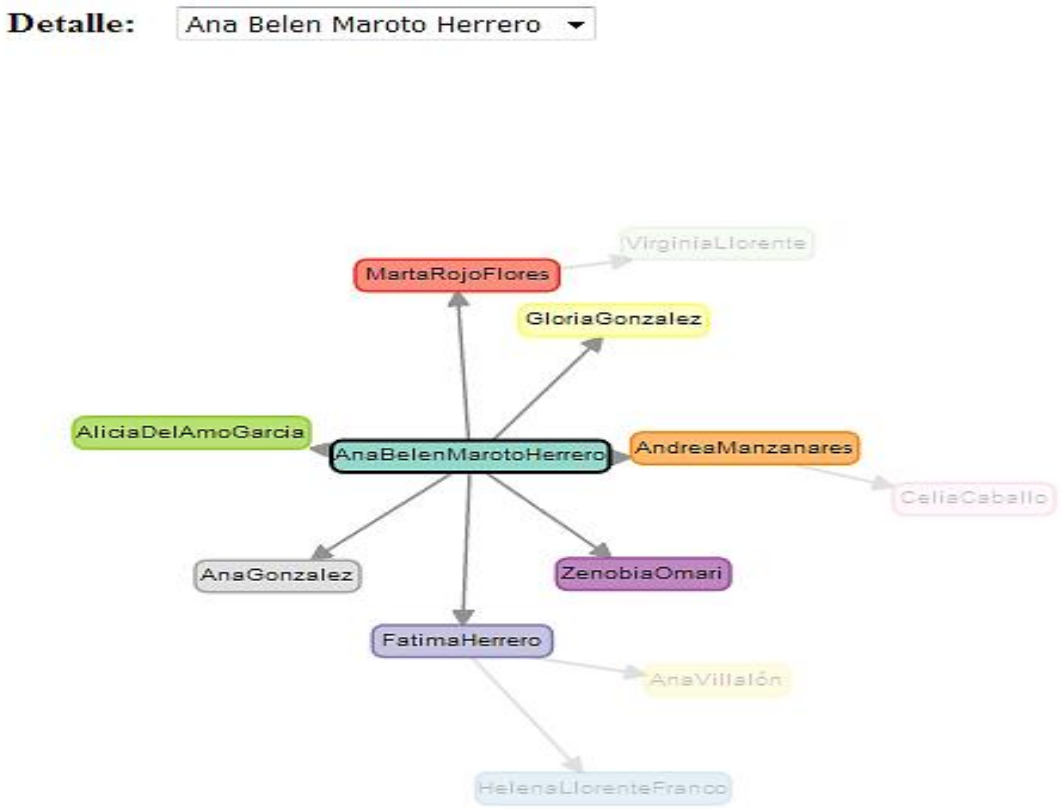


Figura 5.16: *Visualización grafo de Dijkstra*

En esta figura estamos visualizando el nodo que indica el desplegable de selección, posicionándonos sobre él se encienden sus conexiones directas y observamos que para comunicarse de forma eficaz (por el camino con menor coste posible) con ciertos nodos (los que están apagados) necesitaría intermediarios.

Haciendo un recorrido por todos los grafos de Dijkstra de todos los nodos, obtendríamos una idea de que nodos participan en la red como intermediarios.

5.6 Matriz de enlaces

En la última visualización se presenta la matriz de enlaces o matriz de adyacencia que introdujimos en el apartado 4.2 del presente documento. Esta representa las conexiones entre los participantes del grupo excluyendo posts y auto publicaciones.

Para representarla inicialmente hemos divido a los miembros de la red en tres grupos en base a su grado de participación en el mismo. A cada grupo se le ha asignado un color, verde para los que tienen una participación mayor o igual a la mínima y menor que la mínima más un tercio del rango, morado para los que están entre el máximo anterior y menor que la mínima más dos tercios del rango, y azul para los que superan el máximo del grupo anterior. Dentro de cada grupo los participantes están ordenados de mayor a menor participación.

En este caso los valores que almacena se traducen en colores de mayor o menor intensidad dependiendo del valor de la comunicación establecida, entendiendo por valor de la comunicación el número de veces que se comunicaron y mediante que recursos lo hicieron (mensajes, o reacciones). El color blanco es el que representa la ausencia de comunicación entre dos miembros.

Ahora tenemos que hacer distinción en función de si la relación es entre dos miembros que pertenecen al mismo grupo o son miembros de distinto grupo:

- Si son del mismo grupo, el valor de la comunicación se representa de menor a mayor mediante la gama de colores del grupo, cuanto más intenso sea el color en cuestión mayor flujo y valor de comunicación representa.
- Si son de distinto grupo, el valor de la comunicación se representa de igual forma pero en escala de grises hasta llegar al negro.

La visualización permite ordenarlos por orden alfabético (por nombre), por cluster (color en representación de la participación), agrupando así aquellos que tienen una participación similar y por frecuencia.

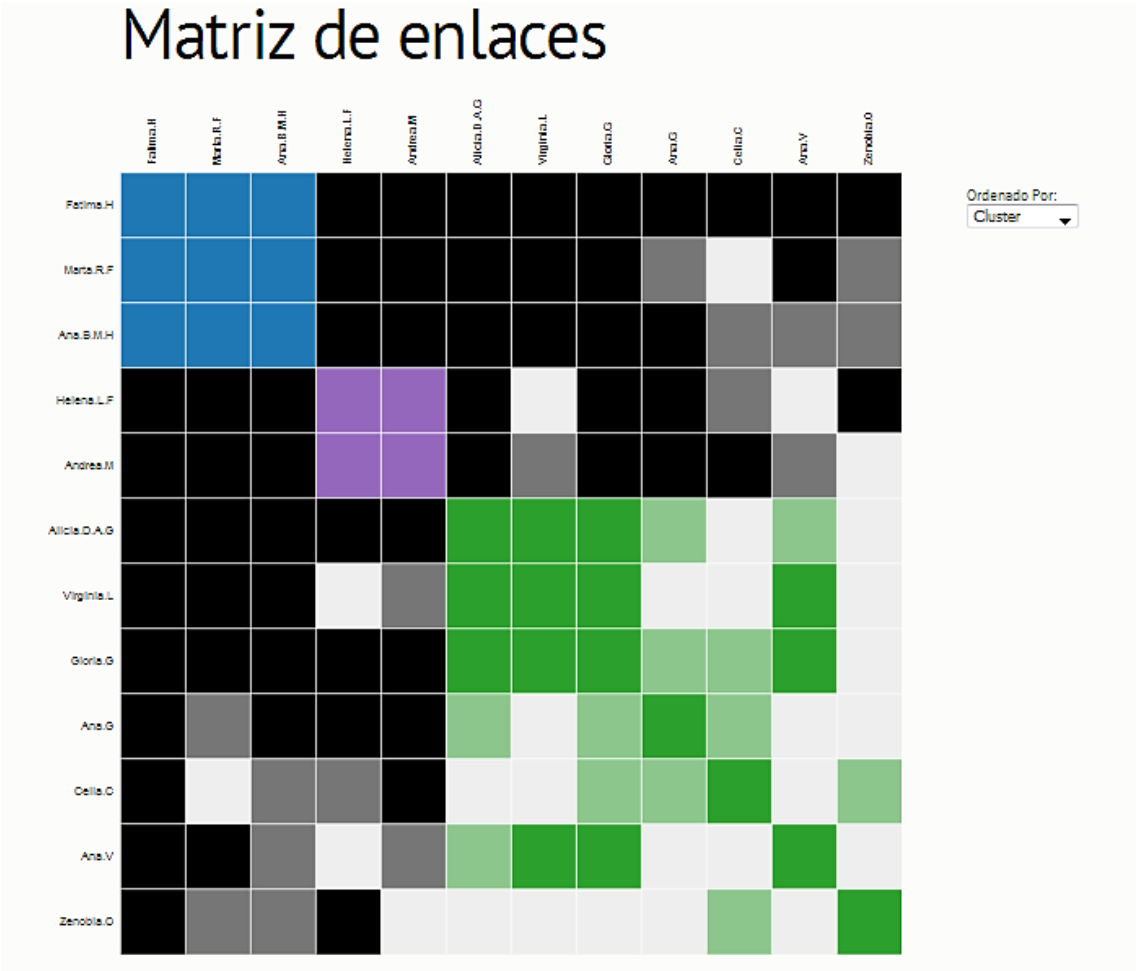


Figura 5.17: *Matriz de enlaces ordenada por cluster*

En esta ordenación podemos diferenciar los tres grupos en que hemos dividido la red con el fin de mostrar su grado de participación con respecto al resto, se observa que el grupo de color verde cuenta casi con más de la mitad de los participantes, por tanto entiende que la distribución de la participación en la red es bastante desigual.

También observamos que tanto el grupo azul como el morado cuyos miembros aportan más al conjunto mediante su participación establecen relaciones de alto valor, cosa que no ocurre en el grupo con menor participación, donde se observa que los tres con mayor participación si establecen relaciones de calidad entre ellos y no así con el resto de su grupo, incluso siendo inexistente la comunicación con algunos.

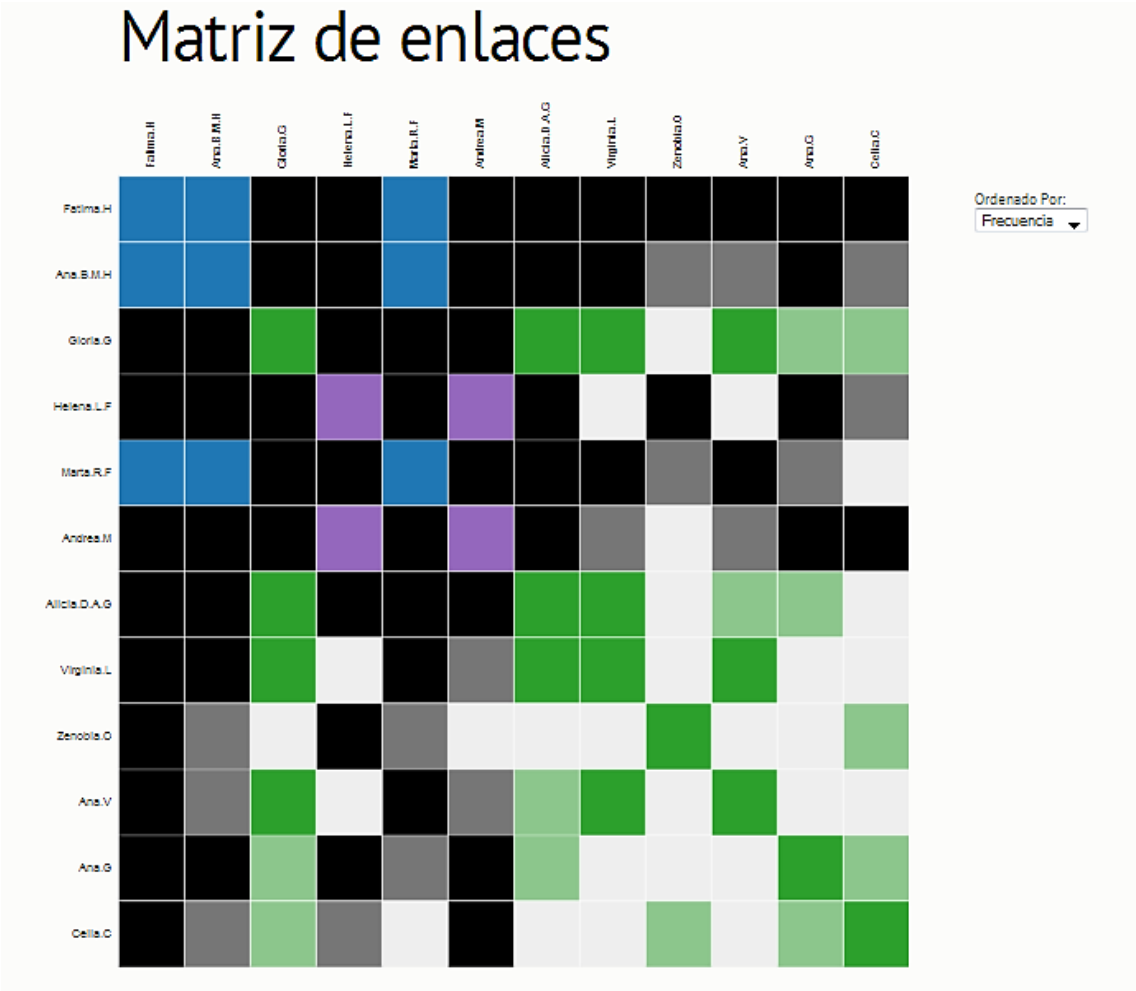


Figura 5.18: *Matriz de enlaces ordenada por frecuencia*

La ordenación por frecuencia coloca a los participantes haciendo un recuento del valor de todas las comunicaciones.

Por ejemplo, el miembro que ocupa la primera fila se comunica con toda la red, y además la comunicación es de calidad. En el caso del miembro que ocupa la quinta fila, es un miembro con alta participación dentro de la red, sin embargo observamos que la calidad de sus conexiones es más baja que la de otros participantes con menor participación en la red.

5.7 Conclusiones

Una vez finalizada la parte de visualización de los datos se concluye que la librería D3.js permite generar visualizaciones muy atractivas y que presentan al usuario gran cantidad de información.

No obstante, hay que tener en cuenta que la exploración de los grafos que conforman la librería es un paso muy importante, pues de su selección dependerán los resultados obtenidos, estos deben expresar la información de forma clara y a poder ser

interactiva, características que enriquecerán y facilitarán al usuario la extracción de conclusiones.

La mayor dificultad encontrada para generar los grafos reside en acondicionar los datos procedentes de la base de datos al formato que espera el código que genera la visualización. Como resumen del trabajo realizado en esta parte se mencionan los archivos que han sido creados por el alumno: “index.html”, “visualizacion.php”, “DashBoard.php”, “flare-imports.json”, “flare.csv”, “objects.json” y “miserables.json”.

6. Conclusiones

Para concluir con el trabajo fin de grado se presentan las conclusiones generales, personales y posibles trabajos futuros.

6.1 Conclusiones generales

En este trabajo fin de grado se diseñó una aplicación que permite evaluar a los alumnos participantes en un grupo privado de Facebook mediante su participación en el mismo.

En cuanto a la parte teórica, el Análisis de Redes Sociales que fundamenta nuestro estudio, nos permitió definir los conceptos necesarios para transformar los datos en información útil, capaz de aportar conocimiento.

La parte práctica se llevó a cabo utilizando herramientas que posibilitaron: la extracción de datos de un entorno tan actual como son las redes sociales y tan popular como es actualmente Facebook gracias al API Graph y a la librería SDK de Facebook, la organización y almacenamiento de la información en base de datos mediante MySQL lo cual aportó gran versatilidad a la aplicación, pues debido a su utilización nuestra aplicación es capaz de gestionar y presentar información de diversos grupos de Facebook, el análisis de los datos y finalmente su visualización mediante la librería de grafos D3.js que nos permitió crear grafos interactivos y visualmente atractivos enriqueciendo así el análisis de la información.

Aunque ciertamente la librería de grafos D3.js que utilizamos para la fase final de visualización es la cara visible, todas y cada una de las tecnologías utilizadas juegan un papel fundamental dentro del desarrollo de la aplicación y son completamente necesarias para la implementación de la misma.

Por último el diseño de la aplicación se realizó a través del entorno de desarrollo integrado Netbeans, el cual posibilito el acceso y acercamiento al resto de tecnologías.

6.2 Conclusiones personales

El análisis y las herramientas para el tratamiento de los datos que se generan en las redes sociales ofrecen grandes posibilidades en cuanto a su uso en actividades docentes. Se ha comprobado mediante este trabajo que es posible el acercamiento de

las redes sociales al ámbito docente, entendidas como una herramienta complementaria del mismo.

En nuestro caso, se ha trabajado con un grupo de datos asumible para trabajar con él, hablando de capacidad de computación. Sin embargo, es lo suficientemente grande como para generar problemas cuando tratamos de analizar la información sin las herramientas utilizadas en este trabajo fin de grado.

6.3 Líneas futuras

La detección de subgrupos no conectados en una red, puede ser de utilidad para realizar comparaciones respecto a la medida de centralidad basada en grados de una persona que participe en grupos distintos. Sería posible detectarlos mediante la extracción de todos los miembros que componen cada red, y no solo aquellos que participan en la misma.

Con los datos que manejamos y que tenemos en la base de datos, se podrían realizar visualizaciones de tipo Timeline, de este modo se puede ver mediante las fechas de las publicaciones en que períodos temporales los usuarios participan más activamente.

Otra mejora posible seria valorar las reacciones que reciben los participantes, mediante el tipo de reacción, es decir; contabilizar positivamente reacciones de tipo LIKE, LOVE, WOW y THANKFUL y negativamente reacciones de tipo HAHA, SAD, ANGRY.

7. Bibliografía

[1] F. Esteve, «Bolonia y las TIC: de la docencia 1.0 al aprendizaje 2.0», *Cuestión Univ.*, vol. 0, n.º 5, pp. 58-67, oct. 2016.

[2] F. Llorens Cerdà y N. Capdeferro Planas, «Posibilidades de la plataforma Facebook para el aprendizaje colaborativo en línea», *OJS*, jul. 2011.

[3] F. Staff, «Los 10 sitios web más visitados en el mundo • Forbes Mexico», *Forbes Mexico*, 17-may-2017. [En línea]. Disponible en: https://www.forbes.com.mx/los-10-sitios-web-mas-visitados-en-el-mundo/. [Accedido: 25-ago-2017].

[4] «47 Incredible Facebook Statistics and Facts for 2016», *Brandwatch*, 12-may-2016. [En línea]. Disponible en: https://www.brandwatch.com/blog/47-facebook-statistics-2016/. [Accedido: 25-jul-2017].

[5] «Infographic: Facebook's Road to 2 Billion», *Statista Infographics*. [En línea]. Disponible en: https://www.statista.com/chart/10047/facebooks-monthly-active-users/. [Accedido: 25-jul-2017].

[6] E. P. V. Gómez y O. J. B. Sandoval, «Structural analysis of interactions in a social network from an educational activity in higher education», *Redes Ing.*, vol. 7, n.º 1, pp. 6-15, jun. 2016.

[7] «API Graph - Documentación», *Facebook for Developers*, 2017. [En línea]. Disponible en: https://developers.facebook.com/docs/graph-api. [Accedido: 19-jul-2017].

[8] «Aguirre, J. L. (2011) Introducción al Análisis de Redes Sociales. Buenos Aires: Documentos deTrabajo, 82, Centro Interdisciplinario para el Estudio de Políticas Públicas, Diciembre.» .

[9] Á. Rayón, «Análisis de redes sociales: el poder de la teoría de grafos», *Deusto Data*, 13-dic-2015. .

[10] «Luis Sanz Menéndez, (2003): Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes». .

[11] «Problema de los puentes de Königsberg», *Wikipedia, la enciclopedia libre*. 10-jul-2017.

[12] «HANNEMAN, Robert A. (2000): Introducción a los métodos del análisis de redes sociales. Departmento de Sociología de la Universidad de California Riverside, EE.UU. Cap_4». .

[13] «Representación de Grafos», *lizardorodriguez.wordpress.com*, 18-jun-2012. .

[14] «HANNEMAN, Robert A. (2000): Introducción a los métodos del análisis de redes sociales. Departmento de Sociología de la Universidad de California Riverside, EE.UU. Cap_6». .

[15] Jhosimar George Arias Figueroa, «CAMINO MAS CORTO: ALGORITMO DE DIJKSTRA», *Algorithms and More*, 2012. .

[16] «MySQL :: MySQL 5.7 Reference Manual :: 1.3.1 What is MySQL?» [En línea]. Disponible en: https://dev.mysql.com/doc/refman/5.7/en/what-is-mysql.html. [Accedido: 18-ago-2017].

[17] «MySQL :: MySQL 5.0 Reference Manual :: 1.4.2 Las principales características de MySQL». [En línea]. Disponible en: http://ftp.tcrc.edu.tw/MySQL/doc/refman/5.0/es/features.html. [Accedido: 18-ago-2017].

[18] «Interactive Data Visualization for the Web_ An Introduction to Designing with D3 [Murray 2013-04-05]». .