



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



**TFG del Grado en Ingeniería
Informática**
IG Study



Presentado por
Miguel Pérez López de Echazarreta
en Universidad de Burgos — 3 de julio
de 2019

Tutor: Dr. Jose Manuel Galán Ordax
y Dr. José Ignacio Santos Martín



UNIVERSIDAD DE BURGOS
ESCUELA POLITÉCNICA SUPERIOR
Grado en Ingeniería Informática



Dr. Jose Manuel Galán Ordax, profesor del departamento de Ingeniería Civil, área de Organización de Empresas.

Expone:

Que el alumno D. Miguel Pérez López de Echazarreta, con DNI 71301255J, ha realizado el Trabajo final de Grado en Ingeniería Informática titulado IG Study.

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Burgos, 3 de julio de 2019

Vº. Bº. del Tutor:

D. Jose Manuel Galán Ordax

Resumen

Se va a realizar un estudio sobre usuarios de Instagram que cumplan un requisito particular, en nuestro caso ser de la misma ciudad. Se utiliza Instagram, por su gran uso a día de hoy y los pocos estudios que hay por el momento.

Se creará una herramienta que pudiera servir a empresas para una mejor dispersión de un mensaje o producto, gracias al estudio sobre personas relevantes, dividiendo por comunidades y particularidades de estas que aprendamos gracias a un estudio sobre los hashtags que usan.

La red va a requerir de un número manejable y suficiente de usuarios a estudiar que conformen la red.

Los resultados finales serán expuestos sobre una web. Mostrarán como se dividen las comunidades, viendo si hay alguna característica específica que une a todos ellos, como que viven en el mismo pueblo o barrio, y como cada comunidad tiene unas necesidades especiales que requieren de una forma de dispersión diferente.

Descriptores

Instagram, Estudio de redes sociales , teoría de grafos ...

Abstract

It is going to be done a research about Instagram users that fulfill a specific requirement, in our case it will be to be all of them from the same city. It will be used Instagram, because of the great use that people is doing of it nowadays and the little research that has been done for the moment.

It will be develop a tool that might be used by enterprises for improving the spread of a message or a product, based on the study of relevant users, dividing by communities and the particular characteristics of them that we learn thanks to the study of the hashtags that they use.

The network need to have a sufficient and easy to handle number of users for the study.

The final results will be exposed on a web. They will show how the communities are divided, watching if there is any specific characteristic that all of them fulfill, like for example, that all of them are from the same zone or village, and that each community has some special needs that requires a different spreading in the network.

Keywords

Instagram, Social media research, Graph theory . . .

Índice general

Índice general	III
Índice de figuras	v
Índice de tablas	vi
Introducción	1
Objetivos del proyecto	3
2.1. Objetivos globales	3
2.2. Objetivos técnicos	3
2.3. Objetivos personales	4
Conceptos teóricos	5
3.1. Instagram	5
3.2. Teoría de grafos	7
3.3. Pagerank	8
3.4. Modularidad	9
Técnicas y herramientas	11
4.1. Técnicas de Desarrollo	11
4.2. Lenguajes de programación	13
4.3. Entornos de Desarrollo	14
4.4. Administradores de Paquetes	16
4.5. Herramientas de Documentación	17
4.6. Bibliotecas	18
4.7. Herramientas para el control de versiones	18

4.8. Herramientas de diseño gráfico	20
Aspectos relevantes del desarrollo del proyecto	21
5.1. Definición del proyecto	21
5.2. Creación general de la red	23
5.3. Estudio por comunidades	25
5.4. Búsqueda de los usuarios más relevantes	30
5.5. Estudio de los Hashtags	31
5.6. Creación de una Web	32
5.7. Realización del proyecto	33
5.8. Creación de una Web	33
Trabajos relacionados	35
6.1. AutoLiker	35
Conclusiones y Líneas de trabajo futuras	37
7.1. Conclusiones	37
7.2. Líneas de trabajo futuras	38
Bibliografía	39

Índice de figuras

3.1. Foto, con pie de foto, hashtags y personas que han dado me gusta	6
3.2. Ejemplo de Perfil, con descripción de perfil y seguidores y seguidos	7
3.3. Fórmula del PageRank básica	8
4.4. Archivo JSON que nos devuelve la API	14
5.5. Logo del proyecto	23
5.6. Tendencia de ganancia de personas como nodos de la red	25
5.7. Crea un árbol de decisión para ver los puntos óptimos de corte	26
5.8. Imagen de un grafo neutro	28
5.9. Force Atlas 2 básico	29
5.10. Force Atlas 2 básico con colores	30

Índice de tablas

Introducción

Existen muchos estudios sobre redes sociales y relaciones entre las personas en ellas. Twitter suele ser la más usada. Instagram está tomando la cabeza a día de hoy en las redes sociales y sin embargo aun así no hay estudios sobre ella. Este ha sido el gran motivador para elegir esa red social. He encontrado una posible necesidad a la hora de promocionar específicamente a la gente los productos y viendo sus publicaciones y la de sus amigos quiero encontrar quien es la persona más influyente a su alrededor para elegirla como persona a la que ofrecer un producto y que se pueda expandir con la mayor efectividad por la red. Para este proyecto se ha elegido la red de Burgos y las relaciones que existen entre los usuarios. Esta red tiene un número de usuarios manejables y suficiente para el estudio. Se puede hacer el estudio sobre otros valores que no sean puramente geográficos. El producto final se ofrece en una aplicación web. Permite seleccionar los términos que tienen en común las personas a estudiarse, en mi caso Burgos y la universidad y los pueblos de la provincia, hacer un estudio de las palabras usadas y hashtags. Una división por comunidades en la red y una clasificación en la comunidad de la persona más influyente sobre la que enfocarse para vender nuestro servicio o producto.

Objetivos del proyecto

El propósito del proyecto es estudiar la red de gente de Burgos y sus relaciones en la red social Instagram. Además, se pretende desarrollar de una interfaz web para poder hacer análisis de otras redes, no necesariamente de gente de Burgos, sino usando como factor común otro valor como un deporte o una afición, no solo de manera geográfica.

2.1. Objetivos globales

- Automatizar el proceso de extracción de los datos de las relaciones de Instagram.
- Introducir la información en una base de datos para almacenarla
- Automatizar el estudio de la información.
- Mantener el usuario que ha permitido la extracción de los datos para que un futuro sea posible hacer otro estudio en mayor profundidad.
- Desarrollar una aplicación web que permita:
 - La extracción de datos
 - La impresión por pantalla de los resultados

2.2. Objetivos técnicos

- Montar un servidor para la obtención de los datos (se requiere de unos días)

- Desarrollar una aplicación utilizando Python para la extracción de relaciones de usuarios e interacciones entre ellos.
- Utilización de Scrum como metodología para la planificación del proyecto, con los sprint definidos en Github.
- Utilización de Github como servidor web donde subir las versiones.
- Utilización de Flask para el desarrollo de la aplicación Web
- Utilización de LaTeX como herramienta con la compilar la documentación.
- Uso de Networkx como principal herramienta de estudio de redes.

2.3. Objetivos personales

- Adquirir conocimientos nuevos que no se han profundizado en la universidad.
- Aportar a la universidad una investigación para un mejor desarrollo del sector de la investigación en la universidad.
- Aprender a llevar un proyecto de duración media, con la distribución de cargas de trabajo.
- Aprender a desarrollar una aplicación entera, con el Frontend y el Backend.

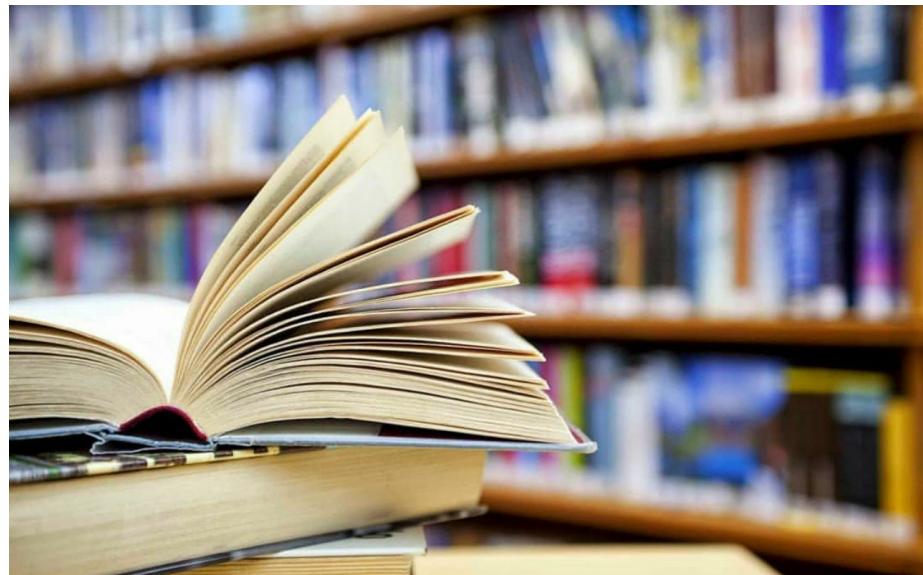
Conceptos teóricos

Para una completa compresión del proyecto hay que tener claros y entender algunos conceptos. A continuación, voy a explicar los más importantes:

3.1. Instagram

Red Social bastante reciente que es mayormente utilizada por una generación joven. Tiene un enfoque para los dispositivos móviles. Aunque existe una versión de servidor web, no permite todas las opciones. Las acciones principales que se pueden realizar son:

- Subir una fotografía con un pie de foto.
- Dar un “me gusta” a la fotografía.
- Dejar comentarios en las fotos.
- Escribir Hashtags que pueden ayudar a encontrar temas relacionados



Liked by **davix_a16** and **93 others**

proyectoburgos Descripción de una foto



#Hashtags #Burgos

March 27 · See Translation

Figura 3.1: Foto, con pie de foto, hashtags y personas que han dado me gusta

- Tener una descripción sobre nuestra persona.
- Seguir a gente y que nos sigan.
- Posibilidad de hacer privado el perfil para tener dominio sobre quien puede seguirnos y ver las fotos que hemos subido.



Figura 3.2: Ejemplo de Perfil, con descripción de perfil y seguidores y seguidos

3.2. Teoría de grafos

Es una rama de las matemáticas y las ciencias de la computación que estudia las propiedades de los grafos. Formalmente, un grafo $G = (V, E)$ es una pareja ordenada en la que V es un conjunto no vacío de vértices y E es un conjunto de aristas. Donde E consta de pares no ordenados de vértices, tales como $x, y \in E$ entonces se dice que X e Y son adyacentes; y en el grafo se representa mediante una línea no orientada que une dichos vértices. Si el grafo es dirigido se le llama dígrafo, se denota D , y entonces el par (X, Y) es un par ordenado, esto se representa con una flecha que va de X a Y y se dice que $(X, Y) \in E$. Existen varios tipos de Grafos: [11]

- Grafos Simples: Un grafo es simple si a lo sumo existe una arista

uniendo dos vértices cualesquiera.

- Multigrafo: Tiene aristas multiples
- Grafos conexos: Un grafo es conexo si cada par de vértices está conectado por un camino; es decir, si para cualquier par de vértices (a, b), existe al menos un camino posible desde a hacia b.
- Grafos completos: Un grafo es completo si existen aristas uniendo todos los pares posibles de vértices.
- Grafos bipartitos: sus vértices son la unión de dos grupos de vértices.
- Árboles: Un grafo que no tiene ciclos y que conecta a todos los puntos.
- Grafos ponderados o etiquetados: Cada arista tiene un peso específico.
- Diámetro: es la mayor distancia entre todos los pares de puntos de la misma.

En las redes sociales, el estudio se centra en la asociación y medida de las relaciones y flujos entre las personas, grupos, organizaciones, computadoras, sitios web, así como cualquier otra entidad de procesamiento de información/conocimiento. Los nodos en la red en este caso son personas y grupos mientras que los enlaces muestran relaciones o flujos entre los nodos. El análisis de redes sociales proporciona herramientas tanto visuales como matemáticas para el estudio de las relaciones humanas.

3.3. Pagerank

Pagerank es un algoritmo utilizado para dar una importancia a un nodo en una red, dependiendo tanto de los conectores de entrada como de salida.

$$\text{PR}(A) = (1 - d) + d \sum_{i=1}^n \frac{\text{PR}(i)}{C(i)}$$

Figura 3.3: Fórmula del PageRank básica

- $PR(A)$ es el PageRank de la página A.
- d es un factor de amortiguación que tiene un valor entre 0 y 1.
- $PR(i)$ son los valores de PageRank que tienen cada una de las páginas i que enlazan a A.
- $C(i)$ es el número total de enlaces salientes de la página i (sean o no hacia A).

El primer documento sobre el proyecto, que describe el PageRank y el prototipo inicial del motor de búsqueda de Google, se publicó en 1998. Sus propiedades son muy discutidas por los expertos en optimización de motores de búsqueda. [16] Nosotros lo vamos a utilizar para ver quién es la persona más importante dependiendo de cómo dividimos el grupo y sobre cual es sobre la que queríamos hacer las propuestas para llegar mejor a la gente.

3.4. Modularidad

La modularidad de un vértice en un grafo cuantifica qué tanto está de agrupado (o interconectado) con sus vecinos. Se puede decir que si el vértice está agrupado como un clique (grafo completo) su valor es máximo, mientras que un valor pequeño indica un vértice poco agrupado en la red. Se suele representar formalmente como C_i . [15] Muchas redes complejas tienen una estructura modular subyacente, es decir, subunidades estructurales (comunidades o grupos) caracterizada por nodos altamente interconectados. La modularidad se ha introducido como una medida para evaluar la calidad de clusterizaciones. [13]

Existen distintos tipos de algoritmos para buscar una mejor optimización de la modularidad [12]:

- Método del mínimo corte
- Algoritmo de Girvan-Newman
- Louvain
- Métodos basados en el Clique

En este proyecto se va a utilizar el método de Louvain (Blondel) para calcular la modularidad y dividir el grafo. Se usa este método por el gran

tamaño que presenta. Su O grande es de $O(n \log n)$. Es un método que utiliza una heurística para calcular una buena modularidad en redes de gran tamaño. [1]

Con la modularidad vamos a ser capaces de distinguir los distintos comunidades que existen en Burgos a grandes rasgos, por la gran interconexión de sus nodos (personas)

Técnicas y herramientas

Esta parte de la memoria tiene como objetivo presentar las técnicas metodológicas y las herramientas de desarrollo que se han utilizado para llevar a cabo el proyecto. Se comentarán los aspectos más destacados de cada opción, con un repaso somero a los fundamentos esenciales y con referencias bibliográficas.

4.1. Técnicas de Desarrollo

Metodología ágil

He decidido usar una metodología ágil para este proyecto. Hay bastantes motivos que he usado para decantarme por este tipo de metodologías:

- El desarrollo es progresivo y admite muy bien la adaptación a posibles cambios. Se puede entregar el progreso al cliente después de cada fase dependiendo de esta.
- Se comienza con unos objetivos y requisitos bajos sobre los que se construye el proyecto. A medida que se van cumpliendo objetivos se van incluyendo nuevos objetivos.
- El que haya una comunicación con el cliente nos permite el que el resultado final sea lo esperado por el cliente puesto que da indicaciones y muestra los posibles cambios a realizar.
- Nos permite buscar las fortalezas en los conocimientos de las personas implicadas más que en los procesos que hemos usado en el proyecto.

- Es mejor tener una buena comunicación con el equipo de trabajo en vez de que la documentación sea excesiva.
- Los roles se pueden dividir bien entre las distintas personas que participan en el proyecto.

Solapar las fases de desarrollo en lugar de hacerlas de forma secuencial o en cascada.

Scrum

Scrum es un marco de trabajo dentro de las metodologías ágiles, el cual vamos a usar. [17] Existen tres roles principales que en este proyecto se dividen de manera un poco distinta a una empresa por la falta de personas a quien asignar los cargos:

- Product Owner: es la persona que tiene la idea que quiere que se desarrolle. Suele ser el cliente que requiere que el producto se haga.
- Se Scrum Master: Es el enlace que existe entre el Product Owner y el Equipo Scrum. Es una persona con experiencia que ayuda al entendimiento de ambas partes y las ayuda tanto a desarrollar y explicar bien sus ideas como a enfocar el trabajo.
- Equipo Scrum: es el grupo de personas encargadas de llevar el proyecto adelante. Lo ideal es que el equipo sea de entre 5 a 9 personas y con distintas áreas de conocimiento abarcadas por estas personas. Las personas del equipo tienen que tener en mente que no trabajan individualmente para una mejor comunicación del desarrollo y que se lleve a cabo con fluidez.

Solapar las fases de desarrollo en lugar de hacerlas de forma secuencial o en cascada. Como he explicado antes, en este proyecto el profesor será el Product Owner, el cliente, del proyecto. El alumno toma el rol de Equipo Scrum dado a que realizará el proyecto solo. El Scrum Master es un rol diluido entre el profesor y el alumno puesto que tienen que entenderse para saber bien cuales son los requisitos y cuales las metas del proyecto. El proyecto se dividirá en varios sprints, en los cuales al principio se dará una importancia máxima a la extracción de los datos que tarda varios días, pues de una buena primera fase dependerá el éxito de que nuestro proyecto llegue a buen puerto. Después se hará una investigación y paulatinamente se desarrollará la investigación y la aplicación web.

4.2. Lenguajes de programación

Python

Es un lenguaje de programación interpretado. Está muy extendido, por su facilidad para aprenderlo, ya que, según su filosofía, se hace un especial hincapié en una sintaxis que favorece el código legible. Ha sido utilizado Python como lenguaje de programación en la versión 2.7. Actualmente se encuentra en la versión 3.7. Se ha utilizado esa versión de Python porque es la que mejor congeniaba con todas las bibliotecas. Alguna biblioteca tarda un poco más en ser actualizada y de esta manera no había ningún problema. La razón por la que me he decantado por Python a la hora de elegirlo como lenguaje para el proyecto es por su facilidad de uso, la gran comunidad que hay detrás apoyándolo, es de código abierto y multiplataforma. Elegí este lenguaje porque tiene una gran cantidad de librerías desarrolladas para el análisis de redes (Networkx, Python-louvain) y porque la API que uso para la extracción de los datos está desarrollada en python. [9]

Flask

Flask es un framework minimalista escrito en Python que te permite crear aplicaciones web rápidamente y con un mínimo número de líneas de código. [2] He elegido este framework para el desarrollo de la aplicación Web donde muestro los datos por la facilidad de aprendizaje que presenta. Está basado en Python, lo cual me permite una gran conexión entre las distintas partes del proyecto. He elegido Flask por varios motivos:

- No necesito crear un servidor web para tener la aplicación. Flask se ejecuta sobre el localhost.
- Utiliza las mismas estructuras para cualquier proyecto y eso da sencillez para su uso.
- Es open source, al igual que Python.
- Me da compatibilidad con las distintas partes del proyecto al tenerlo en Python.
- Me permite soportar alguna prueba unitaria.

HTML

HTML es un lenguaje usado para la crear páginas web. [3] Es conocido por su gran simplicidad y las grandes posibilidades que da cuando se junta con css y js. Admite muy bien el uso de posibles templates. Por el momento es el lenguaje más estandarizado para páginas web puesto que la mayoría de los navegadores han adoptado esta tecnología. He utilizado HTML para el desarrollo de la página web donde se muestra nuestra aplicación.

JSON

JSON (acrónimo de JavaScrip Object Notation) es un formato de texto ligero para el intercambio de datos. Es un formato de texto ligero para intercambio de datos. [4] Es el formato que más se utiliza en las APIs cuando se usa el método GET. En la API utilizada nos devuelve los datos como ficheros JSON que hay que pasar una base de datos primeramente para su almacenamiento, y a csv después para el estudio de Networkx.

```
{"story_type": "60", "pk": "00000000000000000000000000000000", "counts": {}, "args": {"media": [{"image": "https://scontent-mad1-1.cdninstagram.com/vp/911f50e92f92f3209f361602543208bf503fd58/r551.2888-15/e35/cb.126_1800_1800/s150x150/49678893_548902136422577_3779189567463136882_n.jpg?_nc_ht=scontent-mad1-1.cdninstagram.com|cache_key=V1K2PHJUzHmHtxdijyPME3NAU2030_Lc", "id": "198235731398694074_8138542895"}], "links": [{"start": 0, "end": 8, "id": "1805470758", "type": "user"}, {"start": 8, "end": 15, "id": "1386454085", "type": "user"}, {"count": 30400_6399_116_0808_888888888888, "text": "mt.133 like'd loveyou.tos's post", "profile_id": 1885470758, "profile_image": "https://scontent-mad1-1.cdninstagram.com/vp/911f50e92f92f3209f361602543208bf503fd58/r551.2888-15/e35/cb.126_1800_1800/s150x150/49678893_548902136422577_3779189567463136882_n.jpg?_nc_ht=scontent-mad1-1.cdninstagram.com|cache_key=V1K2PHJUzHmHtxdijyPME3NAU2030_Lc", "link": "https://www.instagram.com/p/BXWzjLwDgBv/"}], "next_max_id": 1}
```

Figura 4.4: Archivo JSON que nos devuelve la API

MYSQL

MySQL es un sistema de gestión de bases de datos relacional desarrollado bajo licencia dual GPL/Licencia comercial por Oracle Corporation y está considerada como la base datos open source más popular del mundo, y una de las más populares en general junto a Oracle y Microsoft SQL Server, sobre todo para entornos de desarrollo web. [7]

No se requiere la instalación de ningún software específico, sino que desde línea de comandos puede ser manejada. Permite una muy buena integración con Python. Se ha utilizado la mysql.connector.

4.3. Entornos de Desarrollo

Sublime Text

Sublime Text 3 es un editor de código, escrito en Python y C++, que abarca gran cantidad de lenguajes de programación. Cambia la interfaz dependiendo del lenguaje usado, lo cual facilita la escritura del lenguaje

usado. Para facilitar el desarrollo de un proyecto, admite la instalación de diferentes plug-ins. Aunque fue la primera opción de desarrollo de mi código, al poco de empezar vi que tenía otras necesidades que eran satisfechas por IDEs más específicos para Python. Esto ocurre al desarrollar un proyecto tan enfocado a un solo lenguaje de programación. Como dice el dicho, el que mucho abarca poco aprieta, y este caso así fue.

PyCharm

PyCharm es un entorno de desarrollo (IDE) específico para Python desarrollado en la república checa. Perteneces a la empresa JetBrains. Tiene una gran posibilidad de elegir herramientas que ayudan al buen desarrollo de la aplicación, tales como:

- Escritura.
- Revisión.
- Comentarios.
- Ejecutar el programa en línea de comandos propia.
- Instalación de frameworks específicos.

Con la licencia de estudiante de manera gratuita y siendo el IDE que mejor se adaptaba a mi proyecto, terminé decantándome por esta herramienta.

Bash

Bash es un programa informático, que tiene como función el interpretar las distintas órdenes sobre la consola. Para la extracción de datos se ha montado un servidor en un ordenador personal en casa para tenerlo encendido durante los días que durase la extracción. El servidor corre con un Linux.

Jupyter Notebook

Jupyter es una aplicación de código abierto sobre un explorador web. que permite la creación de programas en Python. Permite dividir el programa en distintas partes para irlo corriendo por separado. Es muy usado en la educación porque puedes enseñar distintas cosas en un solo fichero. La mayoría de la gente lo usa para testear distintos problemas que puede dar un programa. La ejecución por piezas simula una ejecución en vivo. Se instala junto con Anaconda o desde la línea de comando con pip.

Google Colaboratory

Colaboratory es un programa Web que pertenece al motor de búsqueda de Google. El funcionamiento es parecido a Jupyter Notebook. Las ventajas de Colaboratory, son que funciona sobre Drive, en un ordenador de sus servidores, por lo tanto, no necesita de instalación y puedes acceder a tus proyectos desde cualquier sitio. Se guarda todo en Drive. Se ha usado porque ciertos cálculos requerían de mayor fuerza de computación que no tenía en el ordenador personal.

Gephi

Software de código abierto desarrollado en Java que se utiliza para el estudio de redes. Se caracteriza por la falta de necesidad a la hora de programar y por la buena visualización de la red que ofrece. Se pueden elegir distintas características para el estudio tales como:

- Apariencia
- Distribución
- Grafo
- Contexto
- Filtros
- Estadísticas

4.4. Administrados de Paquetes

Pip

Pip pertenece al software Python. Es el encargado de administrar e instalar los distintos paquetes y funciones que queramos obtener junto con Python. Se usa en la línea de comandos, con un funcionamiento similar al que se usa con el APT-GET en Linux. Tiene cuatro posibilidades a la hora de usarlo:

- Instalar.
- Instalar una lista de requerimientos.

- Instalar un paquete específico para una versión de Python, quitando la versión que tuviéramos en nuestro ordenador.
- Desinstalar un paquete.

En los gits de internet muchas veces la gente pone los requerimientos en una lista para facilitar a la gente con el git los nombres y que lo puedan instalar automáticamente.

4.5. Herramientas de Documentación

Libre Office

Libre Office es un software de código abierto que se utiliza para crear documentos escritos. Se ha optado para usar esta herramienta para facilitar la corrección de errores en la documentación. Da facilidades tales como escribir notas en las páginas, y así, cuando no sea posible una reunión presencial con los tutores, estos puedan ir aportando las anotaciones pertinentes.

LaTeX

LaTeX es un sistema de composición de textos, orientado a la creación de documentos escritos que presenten una alta calidad tipográfica. Por sus características y posibilidades, es usado de forma especialmente intensa en la generación de artículos y libros científicos que incluyen, entre otros elementos, expresiones matemáticas. LaTeX es un sistema de composición de textos, orientado a la creación. [8] LaTeX se usará para la maquetación final del proyecto. La he elegido por la gran calidad que ofrece a la hora de imprimir tanto la memoria como los anexos.

Texmaker

Texmaker es uno de los editores gratuitos que existen para LaTeX. Ofrece casi todas las herramientas necesarias para una buena edición del documento que estemos escribiendo. Texmaker es un editor gratuito distribuido bajo la licencia GPL. [10]

Zotero

Zotero es un programa de gestión de enlaces bibliográficos. Funciona instalando un plugin en nuestro ordenador y cuando vemos un artículo que

nos interesa lo marcamos. Cuando hacer el informe, o memoria en este caso, con un plug-in en el editor de texto, se encarga automáticamente de insertar las referencias a las páginas previamente guardadas.

4.6. Bibliotecas

Instagram-API-python

Es un proyecto desarrollado por LevPasha. Tiene una licencia MIT. Esta biblioteca nos acerca la interacción con la API de Instagram a Python. Está basado en un proyecto anterior de PHP aunque en este caso es todo Python. La hemos elegido por lo bien que se adapta a nuestros requisitos para la extracción de los datos.

Networkx

Social network analysis software (SNA software) [6] Es un software de análisis de Redes Sociales. Funciona en Python. Es muy conocido y usado por su versatilidad y por el gran número de algoritmos que lleva integrados. Se usa como base para el estudio de los datos, una vez obtenidos estos.

Python-Louvain

Es una biblioteca paralela a Networkx que aplica el algoritmo de Louvain. Sirve para aplicar un algoritmo muy específico de clustering de grandes redes. Por grandes redes se entiende que tiene al menos varios miles de nodos. Sus desarrolladores teóricos son: Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Renaud Lefebvre en Journal of Statistical Mechanics: Theory and Experiment 2008. El desarrollo de la versión de Python para que fuese unida a Networkx es de Thomas Aynaud . El repositorio de Github es: <https://github.com/taynaud/python-louvain> La documentación se encuentra en: <https://python-louvain.readthedocs.io/en/latest/>

4.7. Herramientas para el control de versiones

Git

Git es un sistema para el control de versiones. El control de versiones se caracteriza porque nos permite regresar a algún punto anterior en el tiempo

del proyecto. En el supuesto de que hubiera algún problema en el proyecto, podríamos ver el cuándo y el por qué se hicieron tales casos. En un caso extremo, trabajar en una línea paralela desde ese punto avanzando. Así podemos evitar pérdidas de informaciones. En la actualidad es el sistema más utilizado gracias a su simplicidad y la eficiencia que muestra. La licencia que usa es de software libre GNU LGPL v2.1

Github

Github es una plataforma online que alberga proyectos desarrollados mediante un sistema de control de versiones git. Existen alternativas tan potentes como Github, tales como Bitbucket y Gitlab. Sin embargo, me he decantado por ella por la simplicidad que ofrece, por la familiarización previa que ya tenía de varias asignaturas durante la carrera y porque permite integrar extensiones paralelas que mejoran su funcionamiento. Es una aplicación gratuita para todos los proyectos de código abierto, aunque existe una versión Pro de pago con más privilegios. Esta versión extendida es la que voy a usar puesto que la dejan gratuita para estudiantes y docentes. Hay cuatro elementos básicos que conocer de la página con los que es conveniente familiarizarse, puesto que han sido usados bastante en el proyecto:

- Milestones: Son distintos momentos importantes marcados en el proyecto. En mi caso marcarán los distintos Sprints.
- Issues: Son tareas asociadas al Milestone (Sprint) que hay que hacer.
- Commits: Son actualizaciones en el proyecto que subimos al git para tener las distintas versiones. Sirve tanto para añadir distintos elementos, como para modificarlos o eliminarlos. Cada vez que una tarea se resuelve se puede ir actualizando.
- Forks: Son bifurcaciones en un punto del proyecto para trabajar sobre el de manera paralela. A veces se hacen para tener solo el punto de partida igual y después desarrollar distintas cosas. Otras veces es el mismo equipo el que hace el fork para resolver algún problema y cuando lo tiene resuelto lo unen al árbol principal del que se separaron.

GitKraken

Es una aplicación de escritorio que está conectado con Github. Permite de manera muy sencilla el hacer los commits, ocupándose ello de actualizar los ficheros. Permite el escribir un título al commit y una descripción. Tiene

una parte gráfica que nos muestra la línea temporal de los commits, cada persona que lo hizo y las distintas bifurcaciones. Se puede pedir que nos descargue todo el proyecto tal y como estaba en algún commit de manera muy sencilla.

Zenhub

Zenhub es una extensión que existe para varios exploradores web. En mi caso uso la que existe para el motor de Mozilla. Es una forma gráfica de aplicar la metodología Scrum tal y como lo haríamos en una pared. Lo bueno de usar esta extensión es que a medida que vas moviendo las tareas por la pantalla, lo va actualizando en el repositorio. La elección de esta herramienta ha sido por considerarla casi un obligado en la gestión del proyecto con una metodología ágil como Scrum.

4.8. Herramientas de diseño gráfico

Gimp

Gimp es un software profesional de edición gráfica. Se uso principalmente para el diseño de gráficas y retoque fotográfico. Es el rival directo de Photoshop, pero de código abierto. La elección se ha basado en el precio desorbitado de la licencia para un uso no profesional, y en el conocimiento previo que tenía de él. Se ha usado en el proyecto para el diseño del poster final y del logo del proyecto.

Aspectos relevantes del desarrollo del proyecto

En este apartado se van a mostrar los puntos más relevantes del ciclo de vida de este proyecto. Se detallarán como se han realizado y resuelto los distintos problemas y dificultades que han ido surgiendo a medida que se iba avanzando en el desarrollo del proyecto. Se dividirá este capítulo en varias secciones en el periodo temporal. Se explicará el porqué de las decisiones tomadas.

5.1. Definición del proyecto

La propuesta del proyecto se les realizó a los tutores que ayudaron a la matización del mismo. En un par de tutorías fueron cincelados detalles importantes sobre el enfoque de este. Al final se concluyó que queríamos conseguir:

- Un estudio de Instagram: Puesto que tanto Twitter, como Facebook han sido base de muchísimos estudios dado su magnitud durante muchos años, queríamos hacerlo sobre Instagram porque encontramos muy pocas investigaciones. Esto se debe a que, aunque lleva unos años en el mercado, ha sido en estos últimos años cuando ha sufrido una gran subida de los números en lo que usuarios se refiere.
- Una aplicación web que ofrezca un estudio de una red con unas características específicas a posibles empresas que quieran hacer propuestas comerciales exitosas. Con un estudio minucioso de la red puede ofrecer

de una determinada manera más especializada un producto, adaptándose a:

- Hashtags: Muestran una forma de hablar específica de un grupo altamente conectado. También nos muestran sus intereses particulares.
- Personas influyentes: No necesariamente la persona con mayor numero de seguidores es la más influyente. Abordar a las personas influyentes por comunidades ofrece ventajas de cercanía y abaratamiento de los costes de contratación para la promoción de nuestro producto.

Para llevar estos puntos a cabo tenía que satisfacer 5 tareas principales:

- Creación general de la red: Creación de una red para estudiar con los suficientes nodos como para que diera juego a la hora de hacer una investigación sobre ella. La red debía ser de gente de Burgos (incluyendo la provincia) y gente que estudie en la Universidad de Burgos.
- Estudio por comunidades: Hacer una división por comunidades, sin puntualizar aún que algoritmo se iba a usar. Al no conocer la red era muy complicado estimar cual iba a ser el más efectivo algoritmo de clustering más efectivo.
- Estudio de los hashtags: Se propusieron diferentes posibles estudios en este punto, aunque el éxito de alguno de ello dependía mucho de una posible extracción de datos durante un periodo de tiempo bastante que al final fue inviable y quedará como posible línea futura de estudio.
- Búsqueda de los usuarios más relevantes: con la red definida, hacer una investigación de los usuarios más relevantes a los que poder hacer una propuesta para que se expanda por la red de manera más exitosa.
- Creación de una aplicación web: La aplicación web nos sirve para mostrar los resultados de la investigación de la red.
- Realización de la documentación: Parte fundamental del proyecto en el cual se incluyen los detalles y explicaciones pertinentes del proyecto.

Por tema de privacidad no se subirá la red extraída al repositorio ni adjunta con el código. Con los puntos anteriores definidos nos pusimos manos a la obra a la realización material de estas ideas. Se empezó a estudiar cómo se podía obtener la red.



Figura 5.5: Logo del proyecto

5.2. Creación general de la red

Puesto que no habría proyecto si no se tienen datos sobre los que trabajar, el primer punto a trabajar con gran rapidez fue la obtención de la red. Del éxito de este primer punto dependía que fructificara el resto del proyecto. Antes de nada, se creó un perfil de Instagram que sería el encargado de obtener mediante la API los datos. Instagram tiene la opción de etiquetar las fotos con una ubicación, sin embargo, eso no nos ayudaba para buscar gente de Burgos puesto que mucha gente que pasa de manera temporal por Burgos puede subir ese tipo de fotos. Para buscar a la gente tuvimos que recurrir a las biografías que ponen. La gente suele escribir ahí distinta información:

- Donde vive: Se suele poner la localidad o a veces algún apodo. En nuestro caso podría ser, por ejemplo: “La ciudad de Cid” o “Burgati”

- Donde estudia: Los adolescentes escriben el instituto y los universitarios la universidad y la carrera. cual iba a ser el más efectivo algoritmo de clustering más efectivo.
- Equipos a los que pertenecen: Un ejemplo de esto son los equipos infantiles de Futbol y Baloncesto. Por ejemplo: Vadillos CF, Burgos Promesas
- Equipos de los que son aficionados: Los principales equipos deportivos de Burgos reúnen a gran cantidad de fans de la ciudad. Por ejemplo: San Pablo Burgos, Rugby Aparejadores Burgos, o Burgos CF
- Frases con las que se sienten identificados: Descartamos esto puesto que no nos sirve de nada.
- Descripción de aficiones o profesiones: Tampoco podemos sacar nada de utilidad con esta información, así que es descartada.

Se agrego de manera manual a 15 personas que servirían como punto de partida para la búsqueda del resto de burgaleses. La búsqueda se realizó de la siguiente forma:

- 1. De los seguidores de los 15 primeros, se entraba en cada uno de los perfiles mirando la descripción, y si tenía alguna palabra clave que nos mostrase que esa persona es de Burgos, se le agregaba.
- 2. Una vez se acabó con los 15 primeros, se buscaba entre los nuevos perfiles encontrados y así sucesivamente. Se hizo una búsqueda fue por niveles hasta que obtuvimos un número relevante.

Hay que tener en cuenta que esta búsqueda duró bastante, más de 2 semanas porque la API da de sí lo que da de sí. En esta etapa tuve que familiarizarme con la API de Python y como eran devueltos los datos, al igual que escribir un programa para la detección de palabras en las biografías. También que fuera buscando entre la gente aun no investigada. Es importante tener en cuenta que en los perfiles la red es dirigida, puedes seguir a alguien y que no te siga de vuelta y viceversa, se miró cuál de las 2 listas era más interesante mirar para ganar dinamismo y tiempo. Se llegó a la conclusión que cuando los perfiles son pequeños, tienden a seguir más que los que les siguen, por lo tanto, nos interesa buscar entre los seguidores de la gente de Burgos. Un ejemplo gráfico de esto es: En mi cuenta personal a mí solo me siguen mis amigos, puesto que no soy alguien famoso, y yo sigo a mis amigos más

futbolistas y celebridades que no me van a seguir de vuelta. Es por esto por lo que las probabilidades de que me siga mayor porcentaje de gente relevante es mayor (nos quitamos de en medio famosos que no nos interesan). La tendencia de ganancia de las personas que van a formar parte de la red tiene una progresión bastante estable. Supongo que es porque el número de personas que puede investigar la api es estándar sin importar cuantos más haya en la lista de espera.

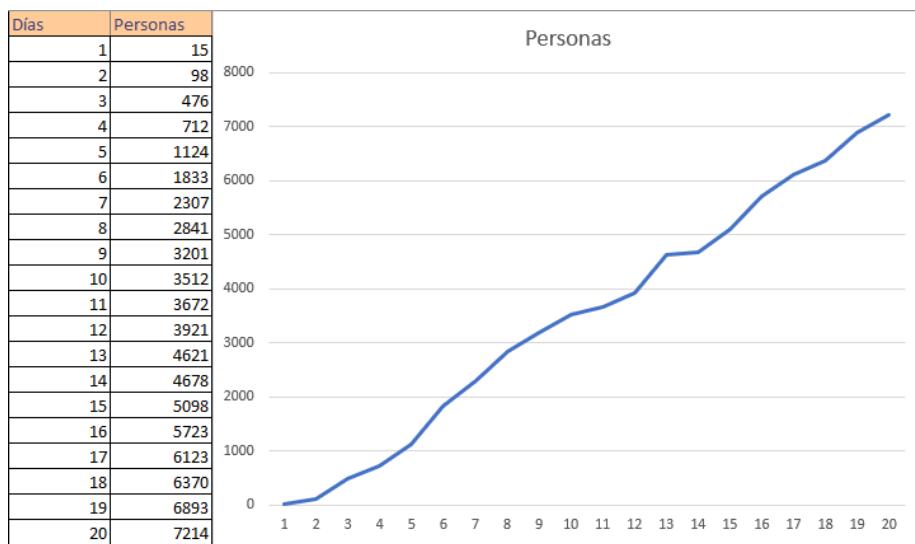


Figura 5.6: Tendencia de ganancia de personas como nodos de la red

5.3. Estudio por comunidades

Hay que dividir la red en comunidades que tengan una alta conectividad entre sus nodos. Para ello existen diversos algoritmos de los que podemos hacer uso. Aunque hablamos de unos cuantos, aunque al final la discusión se centró en solo en los dos siguientes. Cada uno tiene sus pros y sus contras:

Girvan-Newman

El funcionamiento de este algoritmo es sencillo. Calcula el betweeness de los distintos nodos de la red, y cuando encuentra el mayor valor, “corta” esa conexión. Así puede ir dividiendo la red en comunidades que dependen menos del betweeness. Cuando el grado de modularidad empieza a disminuir en las comunidades para. Los pros de usar este método son muchos, los expongo brevemente ahora:

- Se estudia en la carrera, por lo tanto, su funcionamiento es conocido y no requiere de tiempo de aprendizaje
- Los resultados que da son sorprendentemente buenos.
- Viene integrada en la biblioteca Networkx, que usaremos para el estudio de redes.

Las contraprestaciones son pocas, pero de mucha relevancia:

- - Tiene un costo temporal alto ($O(m^2n)$) en una red de n vértices y m aristas, por lo que es poco práctico para redes de más de unos cuantos miles de nodos. [5]

Para una red de Facebook de unos mil nodos funciona bien, y más o menos ahí estaría el límite. [5]

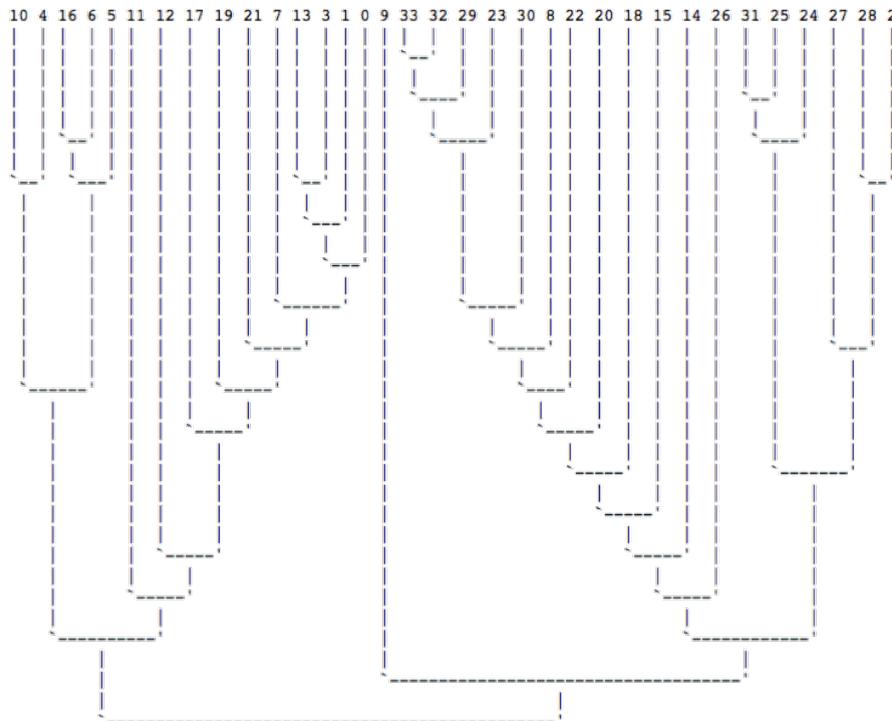


Figura 5.7: Crea un árbol de decisión para ver los puntos óptimos de corte

Louvain

Louvain es un algoritmo que entra en los conocidos como: “Maximización de la modularidad”. El funcionamiento es el siguiente: de manera iterativa optimiza comunidades locales hasta que la modularidad global no puede ser mejorada a través de modificaciones a la distribución de comunidades actual. [14]

El mayor pro que tiene es:

- Las redes que puede soportar en un tiempo aceptable puede llegar a ser de varios cientos de miles de nodos con sus respectivas aristas.

Las contraprestaciones son:

- No viene integrado de manera directa en Networkx, sino que requiere de una biblioteca paralela (Python-Louvain).
- A veces tiene problema identificando comunidades menores de cierta escala, dependiendo del tamaño de la red.
- También presenta deficiencias cuando actúa sobre una red de alta modularidad, cercana al máximo absoluto. Esto no nos afecta a nuestra red.

Nuestra red es de 7493 nodos y 811132 aristas. Teniendo en cuenta estos valores, la decisión final por la red que manejamos es de usar Louvain como algoritmo para realizar las comunidades en un tiempo rápido. Como apunte sobre me gustaría añadir que, aunque si bien se ha dicho que es una contraprestación el tener que instalar Python-Louvain como una nueva biblioteca y tener que investigarla para aprender a usarla, la instalación es sencilla y se acopla perfectamente al funcionamiento de Networkx.

Con Gephi se ha mostrado por pantalla los grafos de la red.

El grafo con los nodos de manera aleatoria presenta una visión así:

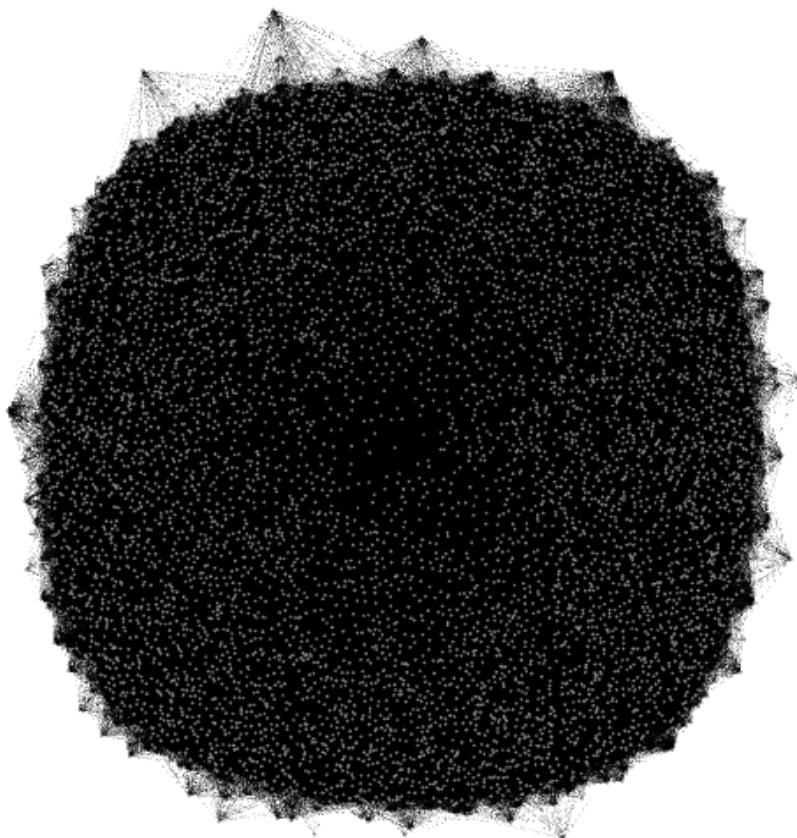


Figura 5.8: Imagen de un grafo neutro

Tras intentar usar varias distribuciones, se ha llegado a la conclusión que quien mejor lo muestra es Force Atlas2. Por tanto, recurriendo a Force Atlas 2 la imagen se nos queda de esta forma. Sin saber bien quien es cada nodo podemos vislumbrar una forma más o menos clara. Se nota que a la derecha hay una comunidad muy marcada en comparación con el gran tamaño del resto de nodos.

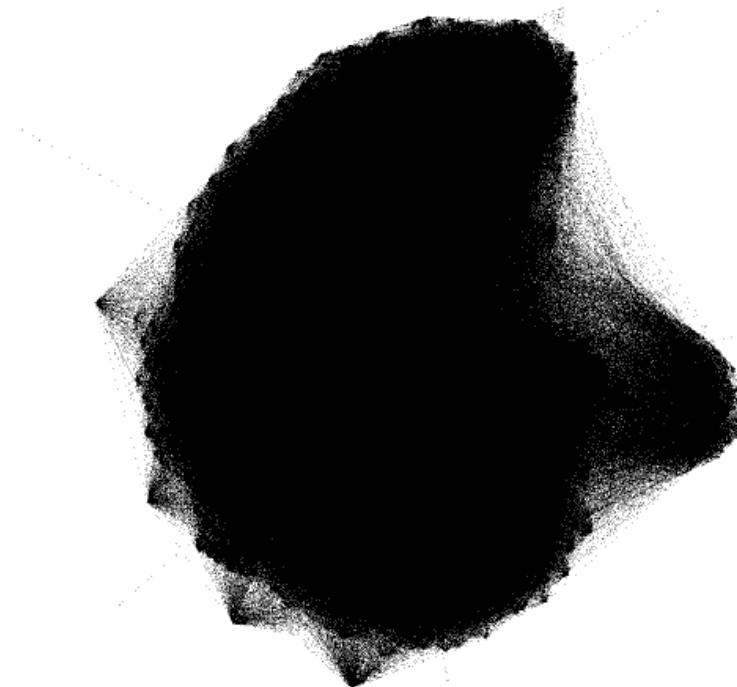


Figura 5.9: Force Atlas 2 básico

Aunque cuando mejor podemos hacernos a la idea de cómo queda la red distribuida por comunidades es al asignarle un color a cada una de ellas. Vemos que estábamos en lo correcto sobre la comunidad de la derecha separada ligeramente del grueso de los nodos.



Figura 5.10: Force Atlas 2 básico con colores

Haciendo un estudio un poco más exhaustivo se sabe por qué esta comunidad separada no tiene tanta relación con el resto. Todo el mundo de esa comunidad pertenece a Aranda de Duero. Al incluir pueblos se ve perfectamente como ellos han creado su propia red interconectada. Viendo esto podríamos deducir que si a “pequeña escala” provincial hay diferencias tan importantes en la modularidad, en un territorio nacional o incluso internacional, se podría llegar a dibujar un mapa y las relaciones entre países basados en las conexiones que existen entre ellos.

5.4. Búsqueda de los usuarios más relevantes

Para la búsqueda de los usuarios más relevantes se pusieron sobre la mesa distintos aspectos a tener en cuenta. Cuál es la variable que queremos dar más relevancia, puesto que viendo como están fallando las medidas

básicas que se usan actualmente para dar patrocinios, igual es momento de ir más allá con la información privilegiada que tenemos. Las marcas en la actualidad contratan a la gente basándose en el número de seguidores que alguien presenta. Esta visión no solo es errónea por la falta de cercanía que puedan crear dado a que no tienen personalización dependiendo de quien vaya a ver el anuncio, sino que además que ha quedado demostrado que existen redes muy extensas de bots que se pueden comprar para mostrar un número de seguidores totalmente falso que solo busca engrandecerse de manera embusteria. Nosotros somos más partidarios de mostrar el producto de manera más personalizada, y mucho más cercana. Al final te marca mucho más lo que te dice un amigo sobre un producto o servicio, que lo que te pueda decir un anuncio de la televisión que es igual para todo el mundo. Cada persona necesita que le hablen en el lenguaje que está acostumbrado a oír y con las experiencias que vive en su día a día. Se pueden aplicar distintas técnicas de centralidad:

- Betweeness: Da valores a las aristas, siendo de mayor peso las que son más usadas para crear caminos entre los distintos nodos. La que mayor betweeness tiene, tiene un mayor número de caminos que conectan nodos de distintas comunidades normalmente.
- PageRank: Algoritmo usado por los motores de búsqueda para dar importancia a los páginas dependiendo de cuantas y como de importantes eran las páginas que le apuntaban.
- Closeness: Es la suma reciproca de la distancia de los caminos más cortos de un nodo con el resto de los nodos de la red.
- Hubs y autoridades: Muestra quien son los nodos que mayor número de aristas tienen distribuidas por la red y quien tiene mayor número de aristas viniendo de la red.

Por la que me he decantado es por PageRank como medida de centralidad más adecuada para este estudio. Es capaz de valorar cada nodo con la importancia que tiene quitando peso a posibles bots que solo sigan a gente y así restando importancia los “influencers” que requieran de estas técnicas para aumentar su número de seguidores.

5.5. Estudio de los Hashtags

Los Hashtags como modo de sintetizar ideas de los usuarios surgió en Twitter. La opción de poderles pulsar y que te mostrase todos los tuits

sobre el tema que usaban ese hashtag ofrecía una forma de crear y agrupar cualquier ocurrencia, sin importar la relevancia de este. Rápidamente se puso en uso para apoyar campañas y darles visibilidad. Como la gente abrazó esta idea realmente bien, el resto de redes sociales han incluido esta funcionalidad con mayor o menor éxito. Por instinto la gente lo usa en cualquier sitio, convirtiéndose en una forma de sintetizar ideas, que las nuevas generaciones han incluido como forma de comunicación en su día a día a través de internet. Por estos motivos nos pareció muy interesante el incluir un estudio de los hashtags que más se usan en la comunidad que estamos estudiando puesto que eso nos da una idea de los intereses y características de las comunidades que estamos estudiando. En este punto he tenido que realizar una limpieza de los datos. Por ejemplo, una misma persona puede ser que haya incluido el mismo hashtag en mil publicaciones, pero nadie más lo éste usando y sin embargo 100 personas hayan usado un mismo hashtag. Es por eso, que, aunque un hashtag aparezca muchas veces, hay que darle un valor especial dependiendo de quien venga y de cuantas personas venga. La valoración de los hashtags se ha hecho haciendo uso de los puntos anteriores:

- Pagerank: Porque puede ser que los nodos con mayor relevancia lo sean en parte por la información que trasmiten. Queremos aprovecharnos de esa sabiduría que publican.
- Comunidades: Los hashtags por comunidades nos ayudan a encontrar palabras y gustos más específicos.

5.6. Creación de una Web

Para mostrar los resultados de la aplicación se hará uso de una web creada con Flask. Se usa Flask por:

- Tiene una forma de trabajar muy limpia.
- La curva de aprendizaje es muy corta.
- Tiene una extensa documentación.
- Al usar el lenguaje Python se integra perfectamente con el resto del proyecto.
- Su integración con todo el resto del proyecto desarrollado en Python.

Se muestran los datos tanto de los nodos más relevantes con sus respectivos hashtags como de las distintas comunidades y la información obtenida de ella.

5.7. Realización del proyecto

El proyecto está escrito en .odt para una mayor facilidad de transmisión entre el alumno y los tutores. Una vez esté la versión final pulida se pasará a LaTeX para que tenga un diseño más formal y profesional.

5.8. Creación de una Web

Como problemas encontrados estuvieron:

- Tiempo en la extracción de los datos: Se pidió ayuda al servidor del departamento de informática. Se nos concedió un espacio, pero al no poder instalar screen o no haber ninguna base de datos que pudiera utilizar, tuve que usar mi ordenador personal como servidor teniéndolo encendido durante el tiempo necesario que requeriera hasta alcanzar un buen número de datos. Desde aquí quiero agradecer a Alvar su colaboración para facilitarme el acceso.
- Problema a la hora de extraer los datos de la base de datos: Al final no se pudo automatizar esta parte y hay que hacer la extracción de manera manual para cambiarla de directorio, para poder empezar con la fase de investigación.

Trabajos relacionados

En cuanto a proyectos relacionados, no hay muchos que trabajen con Instagram como base del estudio. Lo único que se han encontrado son aplicaciones para interactuar con Instagram, más que para realizar un estudio. Por ejemplo:

6.1. AutoLiker

Autoliker es una aplicación que permite dar a "me gusta.^{en}" cualquier página personal o seguir cualquier hashtag y dar "me gusta.^a" todas las publicaciones que lo lleven. El usuario puede configurar:

- El número de publicaciones con las que interactuar
- El mínimo tiempo entre los likes
- Interactuar con las publicaciones más recientes o más relevantes

<https://github.com/thiagonespereira/AutoLiker-for-Instagram>

Conclusiones y Líneas de trabajo futuras

En este apartado vamos a exponer las conclusiones extraídas de la realización de este proyecto y las posibles líneas futuras de desarrollo para la continuidad del proyecto.

7.1. Conclusiones

Ahora que el proyecto se ha terminado, se puede decir que:

- Los distintos objetivos del proyecto se han cumplido satisfactoriamente, creando unas herramientas funcionales que dan opción para una recogida de datos y su futuro análisis.
- Se ha podido hacer uso de una gran cantidad de herramientas y conocimientos que han sido enseñados por la universidad durante estos años en la carrera. Gracias al proyecto han sido asentados.
- Otro de los objetivos personales que me había propuesto era la adquisición de técnicas que no hubiera usado nunca. He podido indagar más en la teoría de grafos, así como en la creación de la Web y el uso de APIs como la de Instagram.
- Con los problemas presentados, he aprendido a manejarme en terrenos desconocidos, teniendo que aprender a tratar con los imprevistos. Esto ofrece un conocimiento, no tanto teórico, sino más bien práctico que me ayuda a progresar en la carrera profesional, porque problemas e imprevistos hay siempre, y hay que aprender a lidiar con ellos bajo presión.

7.2. Líneas de trabajo futuras

Me gustaría aclarar que, aunque este trabajo va a ser presentado como TFG, eso no significa que sea un trabajo cerrado y perfecto. Como aprendimos en la carrera, los productos tienen que estar en continuo desarrollo, y es por esto que doy aquí unas pautas para posibles compañeros que quieran tomar mi proyecto para continuarlo. Respecto a la parte teórica y de investigación veo 2 posibles líneas de trabajo posibles:

- Estudio de los hashtags en un periodo temporal: Buscar patrones en los hashtags o las palabras y emoticonos usados en las publicación hacer una búsqueda de posibles tendencias que existan o similitudes entre los perfiles. Por ejemplo, y encontrar cosas del estilo de: usuarios que van a celebrar el Curpillos al Parral van también a la hoguera de San Juan.
- Dar valores a las aristas: Se puede hacer de distintas maneras
 - Contando las veces que una persona interactúa con las publicaciones de otra persona, ya sea con “me gustas” o con comentarios en la foto.
 - Contando las amistades reales entre personas según el número de publicaciones que tengan en las cuales aparezcan etiquetadas juntas en la misma foto.

Respecto al aspecto y practicidad del proyecto veo:

- - Aplicación Web: Aunque se ha realizado una web para el análisis de los datos y su muestra por pantalla, falta incluir pasos como el de recogida de datos por su duración, y su ensamblaje con el resto de las partes del proyecto.

Bibliografía

- [1] Guillaume J. Lambiotte R. & Lefebvre E. Blondel, V.D. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 10.
- [2] Flask. Pagina oficial de flask, 2019. [Internet; descargado 09-junio-2019].
- [3] Html. Html — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 29-mayo-2019].
- [4] Json. Json — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 01-mayo-2019].
- [5] Mehak Mohammad. Medium, 2018.
- [6] Networkx. Pagina oficial de networkx, 2017. [Internet; descargado 15-mayo-2019].
- [7] Oracle. Pagina oficial de mysql - oracle, 2018. [Internet; descargado 03-mayo-2019].
- [8] Latex Project. Pagina oficial del proyecto latex, 2018. [Internet; descargado 04-junio-2019].
- [9] Python. Pagina oficial de python, 2019. [Internet; descargado 16-junio-2019].
- [10] Texmaker. Pagina oficial del software texmaker, 2019. [Internet; descargado 20-junio-2019].
- [11] Wikipedia. Caracterización de grafos — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 12-mayo-2019].

- [12] Wikipedia. Cluster algorithms — wikipedia, la enciclopedia libre, 2018. [Internet; descargado 10-junio-2019].
- [13] Wikipedia. Cluster — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 12-junio-2019].
- [14] Wikipedia. Louvain — wikipedia , la enciclopedia libre, 2019. [Internet; descargado 07-junio-2019].
- [15] Wikipedia. Modularity — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 12-junio-2019].
- [16] Wikipedia. Pagerank — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 12-junio-2019].
- [17] Wikipedia. scrum — wikipedia, la enciclopedia libre, 2019. [Internet; descargado 13-junio-2019].