



**Centro Universitario de la Defensa  
en la Escuela Naval Militar**

**TRABAJO FIN DE GRADO**

*Herramientas de OSINT para la obtención de información en  
Twitter sobre el conflicto Ucrania - Rusia*

**Grado en Ingeniería Mecánica**

**ALUMNO:** Álvaro Hernández Ruiz

**DIRECTOR:** Norberto Fernández García

**CURSO ACADÉMICO:** 2021-2022

**Universida<sub>de</sub>Vigo**





# Centro Universitario de la Defensa en la Escuela Naval Militar

## TRABAJO FIN DE GRADO

*Herramientas de OSINT para la obtención de información en  
Twitter sobre el conflicto Ucrania - Rusia*

**Grado en Ingeniería Mecánica**  
Intensificación en Tecnología Naval  
Cuerpo General

UniversidadeVigo



# RESUMEN

Hoy en día, las redes sociales son una de las aplicaciones más populares en Internet, estimándose que el 53.6% de la población mundial las utiliza activamente. Es por ello que, en los últimos años han cobrado interés las herramientas de inteligencia de fuentes abiertas (OSINT) que pretenden extraer información relevante (inteligencia) de este tipo de fuentes de información pública. En este trabajo se pretende hacer un estudio de las herramientas OSINT existentes y aplicar alguna de ellas a un caso de estudio de interés en el ámbito de la Defensa. En particular, se ha decidido trabajar con datos de la red social Twitter (la tercera en popularidad en España) utilizando la herramienta T-hoarder para capturar información sobre el reciente (febrero 2022) conflicto entre Ucrania y Rusia. La información capturada se analiza estadísticamente y mediante la herramienta Gephi de análisis de grafos. Aunque el estudio realizado tiene un alcance limitado, los resultados obtenidos permiten concluir el interés de usar este tipo de técnicas para la obtención de información de utilidad.



# AGRADECIMIENTOS

Agradecer en primer lugar a mi tutor, Norberto Fernández García, por la inconmensurable ayuda a la hora de la realización de este proyecto, guiándome durante toda la realización del mismo.

Agradecer también a mi familia y a mi novia Laura que siempre me ha estado apoyando durante toda mi etapa en esta escuela, sin ella no hubiese sido quien soy y haber llegado a donde estoy ahora mismo.

Por último y no menos importante, a mi gran familia que es la Promoción 422-152, que no creo que haya existido una promoción así en tiempo y que de la cual me voy a llevar el mejor de mis recuerdos. Muchas gracias por todo chavales.









## CONTENIDO

Contenido .....	1
Índice de Figuras .....	3
1 Introducción y objetivos .....	7
1.1 Introducción y motivación .....	7
1.2 Objetivos y metodología .....	10
1.3 Estructura de la memoria .....	10
2 Estado del arte .....	11
2.1 La red social Twitter .....	11
2.1.1 Inicios .....	11
2.1.2 Los primeros años de Twitter .....	11
2.1.3 Twitter en la actualidad y características .....	12
2.1.4 Twitter API .....	14
2.2 OSINT .....	14
2.2.1 Historia de la metodología OSINT y características .....	14
2.3 Herramientas para OSINT .....	15
2.3.1 Google Dorks .....	15
2.3.2 Shodan .....	16
2.3.3 Maltego .....	16
2.3.4 The Harvester .....	17
2.3.5 Recon-ng .....	18
2.3.6 Creepy .....	18
2.3.7 DNSDumpster .....	19
2.3.8 Tinfoleak .....	19
2.3.9 T-warder .....	20
2.3.10 SNAcrabe .....	21
2.4 Comparativa de herramientas .....	21
3 Desarrollo del TFG .....	23
3.1 Proceso de Instalación de T-warder .....	23
3.1.1 Twitter API y credenciales .....	23
3.1.2 VirtualBox y Linux .....	26
3.1.2.1 Creando la máquina virtual .....	27
3.1.2.2 Máquina virtual ya existente .....	28
3.1.3 T-warder .....	28

3.2 Funciones de T-hoarder .....	30
3.2.1 Información sobre perfiles de Twitter .....	30
3.2.2 Realizar una búsqueda en Twitter .....	32
3.2.3 Obtener tuits en tiempo real (streaming) .....	33
3.3 Procesado de datos con Gephi .....	33
3.3.1 Instalación de Gephi .....	34
3.3.2 Procesado de datos .....	34
4 Caso de uso .....	37
4.1 Comparativa de datos obtenidos mediante API Streaming y API Search .....	37
4.1.1 API Streaming .....	37
5 Conclusiones y líneas futuras .....	55
5.1 Conclusiones .....	55
5.2 Líneas futuras .....	55
6 Bibliografía .....	57
Anexo I: Código python para el análisis de los datos obtenidos con t-hoarder .....	63

## ÍNDICE DE FIGURAS

Figura 1-1 Estadística del uso de Internet [1].....	7
Figura 1-2 Estadística de las redes sociales más utilizadas en España en 2021 [2] .....	8
Figura 1-3 Estructura de un tuit.....	8
Figura 1-4 Ejemplos de tuits sobre el conflicto de Nagorno Karabaj .....	9
Figura 1-5 Esquema del método OSINT [8] .....	9
Figura 2-1 Odeo, el origen de Twitter [10] .....	11
Figura 2-2 Primera interfaz web de Twitter [11] .....	12
Figura 2-3 Estadística de las redes sociales más famosas de España en 2009 [12] .....	12
Figura 2-4 Perfil en Twitter de la Armada Española [14] .....	13
Figura 2-5 Primavera Árabe (2011) [16].....	13
Figura 2-6 Tuit sobre el paso de buques rusos por el Mar Mediterráneo [17].....	13
Figura 2-7 Invasión de Bahía Cochinos [20].....	14
Figura 2-8 Interfaz Google Dorks [21].....	16
Figura 2-9 Interfaz de Shodan [26] .....	16
Figura 2-10 Interfaz de Maltego [30] .....	17
Figura 2-11 Interfaz de TheHarvester [26].....	17
Figura 2-12 Interfaz de Recon-ng [35].....	18
Figura 2-13 Interfaz de Cree.py [36].....	19
Figura 2-14 Interfaz de DNSDumpster [42].....	19
Figura 2-15 Interfaz de Tinfoleak [47].....	20
Figura 2-16 Interfaz de T-hoarder [48] .....	20
Figura 3-1 Página de creación de una cuenta de Twitter.....	23
Figura 3-2 Email de verificación y página para introducir el código.....	24
Figura 3-3 Interfaz de la página de desarrolladores de Twitter.....	24
Figura 3-4 Página de datos a introducir.....	25
Figura 3-5 Página de verificación de la aplicación de Twitter.....	25
Figura 3-6 Página de creación de la aplicación de Twitter .....	25
Figura 3-7 API key y API key secret .....	26
Figura 3-8 Limitaciones de los niveles de la API de Twitter [54] .....	26
Figura 3-9 Creación de una nueva máquina virtual [56].....	27
Figura 3-10 Máquina virtual preparada para instalación del sistema operativo.....	27
Figura 3-11 Interfaz de instalación de Ubuntu [57] .....	27
Figura 3-12 Página de descarga de KaliLinux .....	28

Figura 3-13 Comando de instalación de Tweepy .....	28
Figura 3-14 Comandos de instalación de T-hoarder .....	28
Figura 3-15 Comandos de creación de las carpetas necesarias .....	29
Figura 3-16 Comandos de creación del archivo .....	29
Figura 3-17 Archivo formato .key con las claves .....	29
Figura 3-18 Comando para ejecutar el programa .....	29
Figura 3-19 Datos a introducir para ejecutar el programa.....	29
Figura 3-20 Pantalla de inicio de la herramienta.....	30
Figura 3-21 Comando de creación del archivo de usuarios .....	30
Figura 3-22 Selección de la opción de analizar tuits de usuarios.....	31
Figura 3-23 Archivo de texto con los tuits .....	31
Figura 3-24 Selección de la opción de obtener la información básica de los usuarios .....	31
Figura 3-25 Archivo de texto con la información de los usuarios .....	31
Figura 3-26 Comando para la creación del archivo de texto en la carpeta del trabajo .....	32
Figura 3-27 Datos a introducir para iniciar la búsqueda .....	32
Figura 3-28 Archivo de texto con los tuits .....	32
Figura 3-29 Archivo de texto con las palabras clave .....	33
Figura 3-30 Datos a introducir para iniciar la recepción de tuits en tiempo real .....	33
Figura 3-31 Archivo con los tuits recogidos en tiempo real .....	33
Figura 3-32 Comando de descarga de Gephi .....	34
Figura 3-33 Comandos de descompresión y ejecución .....	34
Figura 3-34 Selección del procesado de datos en la herramienta.....	34
Figura 3-35 Interfaz gráfica de Gephi .....	35
Figura 3-36 Nodos de los datos diferenciados .....	35
Figura 3-37 Grafo final.....	36
Figura 4-1 Archivo de datos exportado a Excel .....	37
Figura 4-2 Gráfica de número de tuits por usuario (API streaming).....	38
Figura 4-3 Gráfica con los 20 usuarios que más han tuiteado (API streaming).....	38
Figura 4-4 Producto del código de Python .....	39
Figura 4-5 Tabla de datos del código de Python .....	39
Figura 4-6 Gráfica de usuarios con más <i>retuits</i> .....	40
Figura 4-7 Grafica de los usuarios más mencionados .....	40
Figura 4-8 Gráfica de los usuarios más influyentes (API streaming) .....	41
Figura 4-9 Grafo con los <i>retuits</i> de la búsqueda realizada .....	41
Figura 4-10 Parte morada del grafo.....	42
Figura 4-11 Parte azul del grafo .....	42

Figura 4-12 Parte verde del grafo .....	43
Figura 4-13 Parte negra del grafo .....	43
Figura 4-14 Parte rosada del grafo .....	44
Figura 4-15 Comunidades celeste y naranja.....	44
Figura 4-16 Lista de usuarios de la comunidad naranja.....	45
Figura 4-17 Lista de usuarios de la comunidad morada.....	45
Figura 4-18 Usuarios de la comunidad azul .....	45
Figura 4-19 Lista de usuarios de la comunidad verde.....	45
Figura 4-20 Usuarios obtenidos de la búsqueda (API search) .....	46
Figura 4-21 Gráfica con los usuarios que más han tuiteado (API search) .....	46
Figura 4-22 Los 20 usuarios con más tuits (API search).....	47
Figura 4-23 Tabla de datos del código de Python .....	47
Figura 4-24 Gráfica con los usuarios que más <i>retuits</i> tienen (API search).....	48
Figura 4-25 Gráfica de los usuarios con más menciones (API search).....	48
Figura 4-26 Gráfica de los usuarios más influyentes (API search).....	49
Figura 4-27 Representación del grafo de <i>retuits</i> (API search).....	49
Figura 4-28 Comunidad verde del grafo .....	50
Figura 4-29 Comunidad azul del grafo.....	50
Figura 4-30 Comunidad rosa del grafo.....	51
Figura 4-31 Comunidad verde celeste del grafo.....	51
Figura 4-32 Estructura central del grafo.....	52
Figura 4-33 Usuarios de la comunidad naranja.....	52
Figura 4-34 Usuarios de la comunidad rosa .....	52
Figura 4-35 Usuarios de la comunidad negra.....	53





# 1 INTRODUCCIÓN Y OBJETIVOS

## 1.1 Introducción y motivación

A día de hoy, Internet se ha convertido en una necesidad indispensable para la sociedad actual, lo que conlleva una total dependencia de este servicio para usuarios comunes, empresas, organizaciones, etc. En el año 2021, como se puede ver en la Figura 1-1, el 60% de la población mundial utilizaba Internet de los cuales el 53.6% de la población mundial son usuarios activos de redes sociales [1].



Figura 1-1 Estadística del uso de Internet [1]

Como se puede observar en la Figura 1-2, Facebook, Instagram y Twitter se han convertido en las redes sociales más utilizadas en España, de donde se puede obtener información que puede ser beneficiosa para una persona u organización.



Figura 1-2 Estadística de las redes sociales más utilizadas en España en 2021 [2]

Twitter es una red social con la que se puede compartir información de varios tipos de una manera rápida, sencilla y gratuita [3]. En esta red social se pueden encontrar contenidos de diverso tipo, desde mensajes personales a publicidad. Es una red social utilizada desde usuarios comunes hasta empresas para publicitar sus productos o servicios. El tuit (término procedente de *tweet* en inglés) es la denominación utilizada para definir un mensaje publicado en Twitter, al que se le pueden adjuntar incluso vídeos o imágenes. Una característica básica de Twitter es que un tuit tiene como máximo 240 caracteres, lo cual hace que esta red social sea muy característica por el intercambio de mensajes rápidos y directos. Estas características, unidas al hecho de que un mensaje se puede replicar (*retuitear*) con mucha facilidad, hacen de Twitter una red adecuada para la difusión de noticias y otra información de actualidad. En la Figura 1-3, podemos ver un ejemplo de tuit, destacando sus principales elementos.



Figura 1-3 Estructura de un tuit

Como podemos observar en la Figura 1-3, además del texto del tuit, se incluye información de autoría (autor, fecha) y, opcionalmente, menciones a otros usuarios de la red (indicadas con la arroba). También se pueden crear los denominados como *hashtag*, que son etiquetas para clasificar contenido de la misma índole en un mismo grupo. Este tipo de herramienta nos permite seguir temas de actualidad de una manera muy directa [4]. Un ejemplo claro en el ámbito militar podría ser la guerra de Nagorno Karabaj, guerra que empezó en septiembre de 2020 [5], y que pudo ser seguida por los usuarios de esta red, teniendo actualizado al instante cualquier cambio en los hechos, como observamos en la Figura 1-4.



Figura 1-4 Ejemplos de tuits sobre el conflicto de Nagorno Karabaj

Centrándonos en el ámbito de Twitter, sabemos que es una red social que maneja una gran cantidad de información cada segundo. De media se escriben 656 millones de tuits al día [6], de los cuales podemos obtener algún tipo de información que sea beneficiosa para los intereses de Defensa. Usaremos esta red social para la realización del trabajo debido a su popularidad, a sus características propicias para la difusión de noticias, así como por la disponibilidad de interfaces de programación (APIs) y herramientas que permiten acceder a la información públicamente disponible. Por ello es una red social propicia para su uso en aplicaciones de inteligencia de fuentes abiertas.

La inteligencia de fuentes abiertas, o en inglés *Open source intelligence* (OSINT) es la inteligencia que se produce a partir de información disponible públicamente y que se recoge, explota y difunde de manera oportuna a una audiencia adecuada con el fin de abordar un requisito de inteligencia específico [7], tal como se puede observar de forma esquematizada en la Figura 1-5.

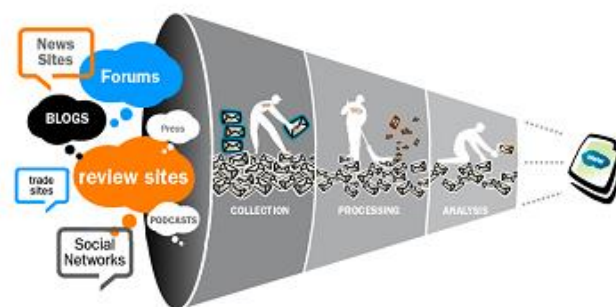


Figura 1-5 Esquema del método OSINT [8]

Este término surge debido a la importante revolución de información, en cuanto a calidad y cantidad que surge en fuentes de información abiertas, por ello la comunidad de inteligencia está ampliando sus esfuerzos y sistemas de explotación para producir inteligencia de fuentes abiertas.

La producción de inteligencia mediante fuentes abiertas es una valiosa disciplina de inteligencia que debe integrarse en la asignación de tareas de inteligencia, la recopilación, el procesamiento, la explotación y la difusión para estar plena y completamente informados.

Además, permite compartir la información ya que la inteligencia de fuentes abiertas se produce sin el uso de fuentes y métodos sensibles. Los productos de inteligencia de fuentes abiertas pueden ser compartidos con todo el mundo debido a la naturaleza no clasificada de la información [7], siempre que

los creadores de información hayan aceptado difundir sus mensajes en público, al no dejar su cuenta en modo privado donde solo ve la información quien el usuario quiera.

## 1.2 Objetivos y metodología

Los objetivos a los que se quieren llegar son los siguientes:

- Revisar el estado del arte de técnicas y herramientas de OSINT.
- Comparar las herramientas existentes y seleccionar una de ellas.
- Aplicar la herramienta en un caso de estudio de interés en el ámbito de la Defensa.

En cuanto a la metodología, se utilizará una metodología ágil, manteniendo reuniones periódicas con el tutor para resolver cuestiones y tomar decisiones sobre la evolución a corto plazo del trabajo a realizar.

## 1.3 Estructura de la memoria

Empezando con el capítulo 1, comenzaremos hablando de los objetivos establecidos y la metodología que utilizaremos en el proyecto, además de una breve introducción a Twitter y a la inteligencia de fuentes abiertas.

En el capítulo 2 de la memoria vamos a describir la historia de Twitter y de cómo surge. Además, nos adentraremos dentro de la metodología OSINT y en las herramientas disponibles para poder aplicar esta metodología a la red social Twitter.

Siguiendo con el capítulo 3, seleccionaremos una de las herramientas que hemos plasmado en el anterior capítulo y mostraremos cómo instalar y configurar dicha herramienta. Además, se realizarán una serie de pruebas para ilustrar cómo funciona.

En el capítulo 4, desarrollaremos el caso de uso. En particular, realizaremos una serie de búsquedas con unas determinadas palabras clave y *hashtags* para obtener información sobre el conflicto que está ocurriendo actualmente (marzo 2022) entre Rusia y Ucrania. Analizaremos la información obtenida mediante grafos y técnicas estadísticas básicas, mostrando e interpretando (en la medida que sea posible) los resultados.

Por último, en el capítulo 5, se llegan a una serie de conclusiones tras los resultados obtenidos en el capítulo 4, además de aportar unas líneas futuras para el desarrollo de este trabajo.

## 2 ESTADO DEL ARTE

### 2.1 La red social Twitter

#### 2.1.1 Inicios

El proyecto de Twitter surge en el año 2004 de las manos del programador estadounidense Noah Glass y del creador de Blogger.com, Evan Williams, bajo el nombre de Odeo [9] (véase Figura 2-1).



Figura 2-1 Odeo, el origen de Twitter [10]

El objetivo de Odeo era dejar un mensaje de audio en formato MP3 y que se quedara alojado en la nube. Es decir, la red social que conocemos hoy en día nació como una plataforma de podcast y de radio por Internet con 14 empleados. Los creadores de Odeo ofrecieron su producto a Apple y, tras unos meses, Apple lanzó en iTunes una plataforma similar de podcasts a la de Odeo, por lo que la empresa empezó a caer en picado.

Tras ver como empezaba la caída de la empresa, los directores decidieron dar un giro a la idea original. Decidieron crear una app de mensajes cortos como el SMS (*Short Message Service*, en inglés) que estaba implementado dentro de la plataforma de Odeo. Así es como surge la plataforma Twitter.

#### 2.1.2 Los primeros años de Twitter

Twitter aparece en escena el 15 de Julio de 2006. El primer tuit de la plataforma fue escrito por Jack Dorsey diciendo que estaba ajustando su cuenta de Twitter. Más tarde mediante la empresa Obvious Corporation, Biz Stone, Jack Dorsey y Evan Williams adquirieron las acciones de Odeo ya que sus creadores perdieron el interés por el proyecto [9]. Como podemos observar en la Figura 2-2, la interfaz de Twitter en aquel momento era bastante sencilla.

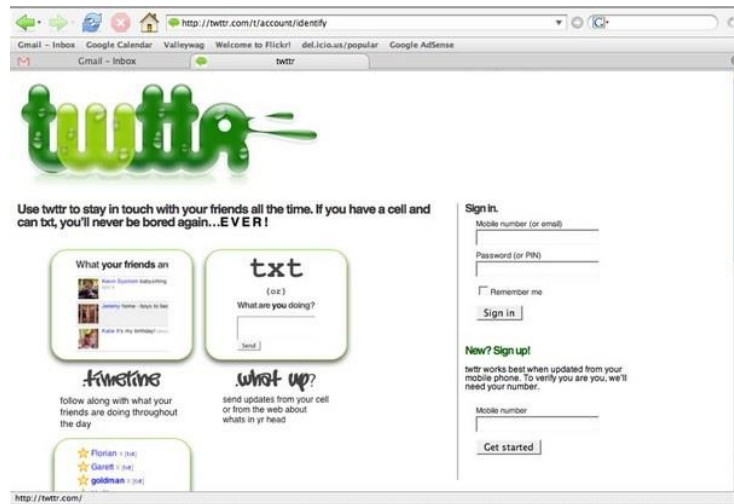


Figura 2-2 Primera interfaz web de Twitter [11]

En 2007, la red social comenzó a alcanzar una gran popularidad y ese mismo año ganó el premio South by Southwest Web Award en la categoría de blog.

A partir del año 2009 la plataforma decide incluir un servicio de publicidad, implementando también en ese año las versiones en español, italiano, francés y alemán. A partir de ese momento se nota una gran expansión debido al auge de los teléfonos inteligentes (*smartphones* en inglés) y a la posibilidad de escribir los tuits desde el teléfono móvil. Como podemos observar en la Figura 2-3, Twitter ya está surgiendo como una de las mayores redes sociales del momento en España.

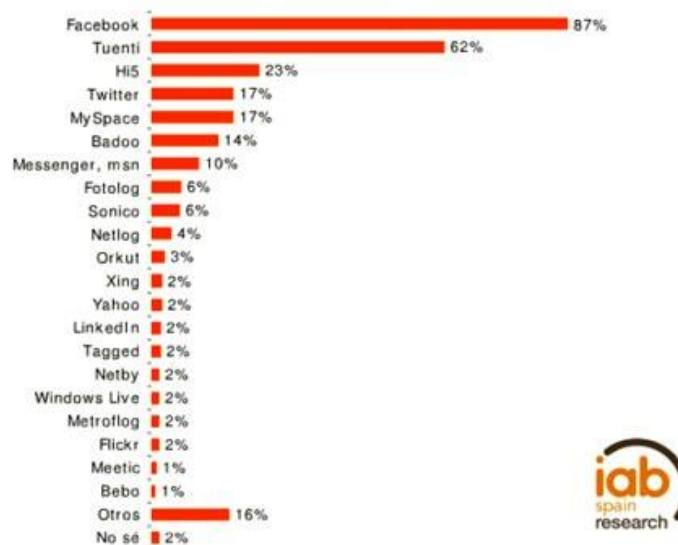


Figura 2-3 Estadística de las redes sociales más famosas de España en 2009 [12]

### 2.1.3 Twitter en la actualidad y características

La red social de los 240 caracteres dio su estallido global entre 2011 y 2014 y en la actualidad es una de las tres redes sociales más grandes del mundo junto con Instagram y Facebook [13]. Tal es su importancia que famosos, empresas, organizaciones (como la Armada Española, véase Figura 2-4) y marcas crean cuentas oficiales desde las cuales promocionan sus productos o servicios o difunden información y novedades llegando potencialmente a millones de usuarios. Surgen a raíz de esto los *'Influencers'*, usuarios no famosos de Twitter que se convirtieron en estrellas gracias a esta red y que disponen de cuentas con millones de seguidores por lo que las marcas les pagan para promocionar sus productos.



Figura 2-4 Perfil en Twitter de la Armada Española [14]

Además, debido a su importancia, Twitter se ha convertido en ocasiones en una alternativa sin censura a los medios de comunicación clásicos, como pasó en la primavera árabe, donde mientras los medios de comunicación censuraban todo tipo de información, miles de usuarios de la plataforma tuiteaban en tiempo real sobre lo que allí sucedía realmente [15], como se puede observar en la Figura 2-5.



Figura 2-5 Primavera Árabe (2011) [16]

Otro ejemplo de uso de Twitter más relacionado con los propósitos de este trabajo lo observamos en el actual conflicto que se mantiene entre Rusia y Ucrania. Como podemos observar en la Figura 2-6, vemos que puede ser de utilidad, por ejemplo, a la hora de ubicar donde se encuentran los barcos de un bando u otro.



Figura 2-6 Tuit sobre el paso de buques rusos por el Mar Mediterráneo [17]

### 2.1.4 Twitter API

Twitter nos ofrece una interfaz de programación de aplicaciones o API (*Application Programming Interfaces* en inglés) pudiendo solicitar información pública en Twitter. Hay funcionalidad como la recepción y envío de mensajes directos en la que se requieren permisos adicionales [18]. Se puede escribir en lenguaje de programación como C++, JavaScript o Python entre otros. Las API de Twitter ofrecen una gran variedad de puntos de conexión (*endpoints* en inglés) divididos en cinco grupos:

- Cuentas y usuarios: Permite silenciar o bloquear a otros usuarios, administrar seguidores, configurar cuentas de forma programática.
- Tuits y respuestas: Los desarrolladores tienen a su disposición los tuits y respuestas públicas, permitiendo también publicar a través de la API. Se puede acceder a los tuits mediante palabras clave o solicitando una muestra de tuits de cuentas específicas.
- Mensajes directos: Acceso a conversaciones mediante mensajes directos de usuarios que han dado permiso de forma explícita a una aplicación específica.
- Anuncios: Ayuda a los desarrolladores a crear y administrar de manera automática campañas de anuncios en Twitter. Los desarrolladores pueden identificar tuits públicos para ver temas de interés lo que da a las empresas herramientas para ejecutar las campañas publicitarias con el objetivo de atraer audiencias en Twitter.
- Herramientas de editor: Permiten a los desarrolladores integrar en páginas web las cronologías de Twitter, el botón para compartir u otros contenidos.

## 2.2 OSINT

### 2.2.1 Historia de la metodología OSINT y características

Se considera que la OSINT moderna nace en 1942, cuando surge la rama *Research and Analysis* de la oficina de servicios estratégicos de Estados Unidos, que se encargaba de recopilar información de fuentes abiertas, haciendo traer periódicos del bando enemigo gracias a la red de embajadas y consulados, escuchando las emisiones de radio extranjeras o accediendo a las librerías y fuentes de información oficiales [19].

Podemos ver ejemplos claros con la invasión de Bahía Cochinos en 1961 (véase Figura 2-7) por parte de los estadounidenses. Varios miembros de las fuerzas armadas estadounidenses habían hablado en medios abiertos sobre la invasión de la misma, dándole todas las claves al ejército de Fidel Castro para contraatacar a esta ofensiva por parte de los estadounidenses.



Figura 2-7 Invasión de Bahía Cochinos [20]



Con la llegada de Internet se pudo convertir a cada usuario en una fuente de información, dando lugar a un aumento de las posibilidades de obtención y haciendo que la fase más difícil del ciclo de inteligencia sea el procesamiento de la información recibida, no tanto su mera obtención.

La OSINT moderna presenta las siguientes características [19]:

- **Eficiencia:** Basándonos en parámetros de coste/beneficio, OSINT es la manera en la que con menos recursos obtenemos mayores resultados.
- **Rapidez:** Acceder a fuentes de información abiertas hace que el ciclo de inteligencia sea más ágil.
- **Desintermediación:** Cuando obtenemos información de fuentes abiertas, normalmente obtenemos la información directamente del usuario y sin pasar por intermediarios.
- **Accesibilidad:** Tiene un coste económico pequeño para los usuarios llenar Internet de información, el mismo coste que a los buscadores de información hacer OSINT, ya que cualquier persona con los recursos necesarios pueden hacer uso de este tipo de inteligencia.
- **Volumen:** Con la llegada de una red que conecta a todo el mundo, cualquier individuo puede ser una fuente de información, con lo que se llega a generar tal volumen de datos (*big data*) que el procesamiento de la información se ha convertido en el mayor reto y no la obtención de esta.

## 2.3 Herramientas para OSINT

En esta sección presentaremos algunas de las principales herramientas de código libre que pueden servir de ayuda a la hora de implementar aplicaciones de OSINT.

### 2.3.1 Google Dorks

Google Dorks [21] se basa en utilizar operadores de Google para búsquedas avanzadas. Google clasifica de manera automática el contenido de páginas web, obteniendo información de diverso tipo. Usualmente esta herramienta la utilizan los hackers para encontrar agujeros de seguridad en la configuración y el código de sitios web. Es normal utilizarla a través de la terminal de Linux, tal y como observamos en la Figura 2-8.

Esta herramienta surge en 2002, cuando Johnny Long comenzó a recopilar las consultas que funcionaban para descubrir vulnerabilidades o desvelar información sensible u oculta. Es una herramienta muy útil para recopilar información que nos proporciona Google, como poder filtrar toda la información por un autor y encontrar todo lo escrito por ese autor [22].

Los hackers utilizan Google para descubrir sitios web vulnerables para después hackearlos utilizando estas vulnerabilidades. Google Dorks es la base para hackear utilizando Google.

Google Dorks aprovecha la equivocación humana para poder acceder a una página web, como por ejemplo dejar un documento Word que incluye credenciales con usuarios y contraseñas, y este documento al estar en un directorio web será indexado por Google [23].



Figura 2-8 Interfaz Google Dorks [21]

### 2.3.2 Shodan

Shodan [24] es un motor de búsqueda, como podemos observar en la Figura 2-9, que permite encontrar diferentes equipos conectados a Internet mediante diferentes filtros. Es utilizado habitualmente por los hackers para buscar servidores, rúteres, webcams, etcétera.

Según narra el creador John Matherly en [25], Shodan se creó para que las empresas pudieran rastrear internamente su software, pasando totalmente lo contrario, los investigadores de seguridad lo utilizaron para encontrar todo tipo de software en dispositivos conectados a Internet.

Con este motor de búsqueda se han logrado realizar ataques informáticos de robo de dinero y de propiedad intelectual [25]. Principalmente los usuarios que están en la plataforma son investigadores de ciberseguridad, pero los hackers pueden acceder a la misma mediante otros mecanismos que hacen acceder a la web de manera remota sin ser detectado.

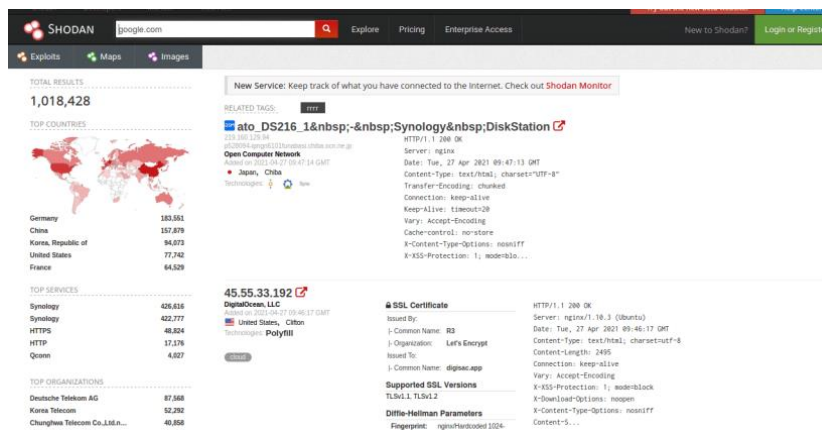


Figura 2-9 Interfaz de Shodan [26]

### 2.3.3 Maltego

Maltego es un software que se centra principalmente en el análisis forense y que se ha desarrollado para que el análisis de URL y la minería de datos a partir de dominios IP sean más sencillos [27].

El software fue lanzado por Maltego Technologies en 2008 y desde entonces no ha dejado de crecer debido a que tiene a una interfaz bastante intuitiva (véase Figura 2-10) para un usuario nuevo [28]. Además, el alcance que tiene la herramienta para recopilar información es bastante grande y debido a esta gran ventaja podemos detectar amenazas cibernéticas antes de que se pueda sufrir un ataque.

Otra de las ventajas que tiene la aplicación es que el software se puede ejecutar en cualquier sistema operativo además de poseer servidores con TAS (*Telephony Application Server* en inglés), propios para asegurarse de la privacidad. Una de las limitaciones que existen es que la versión gratuita que tiene es bastante limitada, haciendo así que para obtener todas las funciones disponibles tengamos que pagar por ello [29].

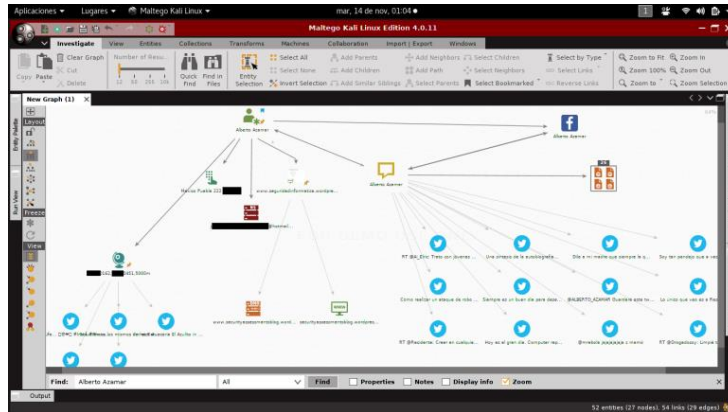


Figura 2-10 Interfaz de Maltego [30]

### 2.3.4 The Harvester

The Harvester [31] es una herramienta por línea de comandos que obtiene información pública en la web, haciéndolo de dos maneras: activa o pasiva. De manera activa, la herramienta interactúa con el objetivo probando todas las soluciones posibles para acceder al DNS y obteniendo capturas de información de los subdominios encontrados. De manera pasiva, la herramienta obtiene la información mediante los motores de búsqueda, como Google o Bing, que vienen integrados en la herramienta.

Esta herramienta fue creada por Christian Martorella, quien trabaja en la empresa Edge-Security. Viene incluida en las distribuciones de Linux tales como Kali y Bugtrug, pero también se puede descargar desde su repositorio en GitHub para su instalación, y ejecutarla desde una consola, como podemos observar en la Figura 2-11 [32].

Se encarga de recolectar información que se encuentra en Internet para futuros ataques en los que se utiliza la ingeniería social en y para los que puede ser de interés disponer información sobre correos electrónicos o números de teléfonos. Se nutre de distintos motores de búsquedas como puede ser Google, Shodan o Bing además de redes sociales como Twitter o LinkedIn.

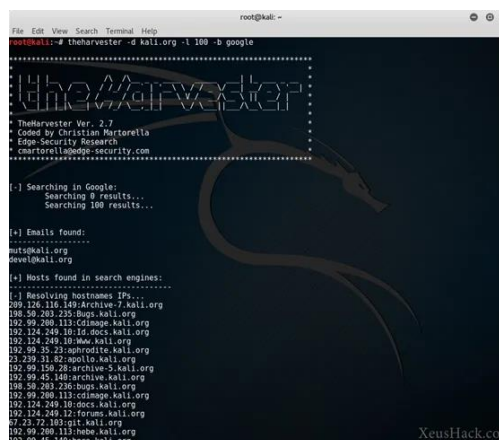


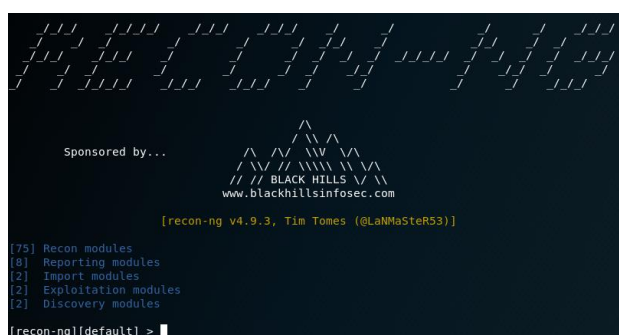
Figura 2-11 Interfaz de TheHarvester [26]

### 2.3.5 Recon-ng

Recon-ng (véase en la Figura 2-12) es una herramienta abierta OSINT desarrollada en Python que automatiza la recolección de información desde varias fuentes haciendo así que el tiempo de recolección se reduzca drásticamente con respecto a hacerlo de manera manual [33].

Esta herramienta capta información tanto de Google, Shodan o redes sociales como Instagram o Twitter. Nos permite llamar a la API de Twitter para la recolección de información. Permite identificar y busca información sobre hostnames, IPs, localizaciones, etcétera.

Una de las características de Recon-ng es que es una herramienta de código abierto, por lo tanto, además de ser gratuita puede ser modificada para mejorarla. Otra de las características que tiene Recon-ng es que puede apuntar a un solo dominio y encontrar todos los subdominios de ese dominio, lo que facilita el trabajo para los *pentesters* (usuarios que atacan diferentes entornos o sistemas con el objetivo de detectar y prevenir posibles agujeros de seguridad [34]).



```

Sponsored by...
          ^
        ^ ^
       ^ ^ ^
      ^ ^ ^ ^
     ^ ^ ^ ^ ^
    ^ ^ ^ ^ ^ ^
   ^ ^ ^ ^ ^ ^ ^
  ^ ^ ^ ^ ^ ^ ^ ^
 ^ ^ ^ ^ ^ ^ ^ ^ ^
^ ^ ^ ^ ^ ^ ^ ^ ^ ^
BLACK HILLS
www.blackhillsinfosec.com

[recon-ng v4.9.3, Tim Tomes (@LaNMaSteR53)]

[75] Recon modules
[8] Reporting modules
[2] Import modules
[2] Exploitation modules
[2] Discovery modules
[2]

[recon-ng][default] >

```

Figura 2-12 Interfaz de Recon-ng [35]

### 2.3.6 Creepy

Creepy [35] es una herramienta de geolocalización. Obtenemos información relacionada con ubicaciones mediante fuentes abiertas. Es posible extraer información de cuentas de Twitter, Facebook, Flickr, etcétera.

Extrae la información de geolocalización de los usuarios que se puede obtener desde varias fuentes distintas. Además, tiene una interfaz bastante sencilla, como se puede observar en la Figura 2-13, ya que solo es necesario introducir el nombre de usuario en el buscador y pulsar el botón de localizar.

Esta plataforma fue creada por Ioannis Kakavas con propósitos educativos y para los profesionales de seguridad en la red. Hablando de Twitter recoge información de posicionamiento de tuits con [37]:

- Información GPS cuando se proporciona desde algunos clientes para teléfono móvil.
- Tuits que tienen asociada una ubicación.
- Obtener la posición vinculada a la IP desde la que se hizo el tuit.
- Recoge información GPS de tuits que vienen desde *Four Square* (red social en la que intervienen diferentes servicios de localización [38]) y de los metadatos de las fotos publicadas en otros servicios de compartición de fotografías.

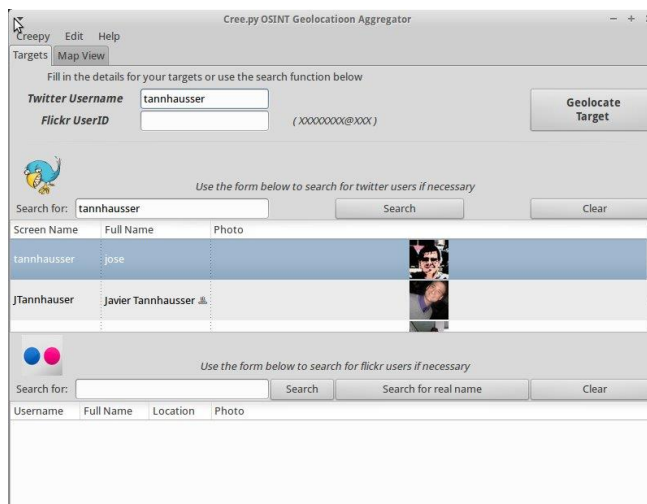


Figura 2-13 Interfaz de Cree.py [36]

### 2.3.7 DNSDumpster

Herramienta que nos da información sobre un dominio [39]. Obtiene la información de diferentes motores de búsqueda. Presenta un uso sencillo ya que basta con poner el nombre del objetivo, como se puede observar en la Figura 2-14.

No utiliza la fuerza bruta como otras, esta herramienta utiliza fuentes abiertas para encontrar la información de una manera más rápida. Los datos los obtiene de plataformas como Alexa Top 1 Million, Google, Bing o de otras herramientas OSINT como Shodan, mencionada anteriormente [40].

Es una herramienta muy importante cuando realizamos análisis de red, ya que la recopilación de información es una parte muy importante. Sin hacer la recopilación de información necesaria, no se puede continuar con la evaluación de red. Se necesita recuperar y recopilar información del dominio como direcciones IP, registros DNS, registros MX, etcétera [41].



Figura 2-14 Interfaz de DNSDumpster [42]

### 2.3.8 Tinfoloak

Herramienta que automatiza la obtención de datos de Twitter y facilita su análisis posterior para que se extraiga la inteligencia, como se observa en Figura 2-15. Introduciendo un nombre de usuario, unas coordenadas geográficas o palabras clave analiza el *timeline* de Twitter y extrae una gran cantidad de datos y muestra información útil y estructurada para el análisis de inteligencia [43].

Esta herramienta está implementada en distribuciones de Linux como Kali [44], CAINE [45] o BlackArch [46]. La herramienta fue creada por Vicente Aguilera Díaz, cofundador de Internet Security Auditors, una de las empresas líder en ciberseguridad en España.

MEDIA	APP	REPLY TO	RT	LIKE	SOURCE USER	RT USER	TWEET
	SocialFlow		498	664	BBC NEWS BREAKING @BBCBreaking 07/20/2017 15:04:52		view
	TweetDeck	@BBCBreaking	878	814	BBC NEWS BREAKING @BBCBreaking 07/20/2017 12:16:11		view
	StrappyTV.com		796	1648	BBC NEWS BREAKING @BBCBreaking 07/20/2017 12:08:44		view
	TweetDeck		2149	2495	BBC SPORT @BBCSport 07/08/2017 17:12:31	BBC NEWS BREAKING @BBCBreaking 07/08/2017 17:28:40	view

Figura 2-15 Interfaz de TinFoleak [47]

### 2.3.9 T-hoarder

Se trata de una aplicación que utiliza Python y que permite interactuar con la API de Twitter de manera sencilla para obtener información y guardarla de forma fácil y reutilizable. Además, tenemos dos métodos de utilización: mediante instalación directa en Windows o Linux o mediante una imagen virtual para ejecutarla en VirtualBox en donde ya viene instalado todo lo necesario para trabajar con la aplicación [48].

Esta herramienta fue creada por Mariluz Congosto, doctora en telemática por la Universidad Carlos III de Madrid que desde 2008 es investigadora en dicha universidad especializándose en el análisis de datos de redes sociales [49].

Para trabajar con esta herramienta será necesario crear una aplicación desde la plataforma de desarrolladores de Twitter y obtener así unas claves que se generan automáticamente y que permiten acceder a la API.

Una de sus características principales es que es una aplicación de código abierto, por lo que se puede descargar y utilizar de manera gratuita además de poder ser modificada para mejorarla o restringirla.

Una vez que hemos obtenidos los tuits, podremos generar un grafo de relaciones entre usuarios ya que T-hoarder tiene la opción de devolver dichas relaciones en un archivo con formato gdf que se puede importar en un software de análisis de grafos llamado Gephi [49].

```

numeroteca@cinco ~/sites/t-hoarder_kit/scripts (master *) $ ./t_hoarder_kit.sh
-----> Welcome to t-hoarder kit <-----
-----
t-hoarder_kit is up to date
working in /home/numeroteca/sites/t-hoarder_kit
-----
-----> Environment data <-----
-----
Enter the file name with the application keys: forthoarder.key
Enter a twitter user: numeroteca
Enter experiment name: pruebas
    
```

Figura 2-16 Interfaz de T-hoarder [48]

### 2.3.10 SNscrape

SNscrape [51] es una herramienta para servicios de búsqueda de información de redes sociales. Investiga información de las redes sociales como los perfiles de los usuarios, los *hashtags* o las búsquedas y devuelve los tuits que incluyan lo que hayamos filtrado.

Fue creado por Martin Beck en 2018 y la gran característica que tiene es que no es necesaria una cuenta de desarrollador de Twitter para obtener tuits, lo que conlleva no utilizar la API y por lo tanto no tiene ninguna limitación de búsqueda de tuits. [52]



Figura 2-17 Logo de SNscrape [51]

## 2.4 Comparativa de herramientas

Analizando las herramientas anteriormente mencionadas, observamos que la gran mayoría de herramientas descritas están orientadas al análisis de ciberseguridad y *pentesting*, que, aunque es un área de interés, no es el ámbito concreto de este trabajo. En nuestro caso necesitamos herramientas en las que realicemos un análisis de Twitter, es decir, lo que se denomina como una minería de tuits para su posterior análisis y procesado.

Teniendo esto en cuenta las herramientas más adecuadas para la realización del trabajo serían Tinfoleak, SNscrape y T-warder. Tras su instalación se probaron todas las herramientas y su utilización en la recopilación de tuits y se ha decidido elegir la herramienta T-warder.

Se escoge la herramienta T-warder debido a su facilidad de uso y a la mayor flexibilidad a la hora de recopilar información respecto a Tinfoleak, que se centra en obtener información de la actividad de un usuario. Además, T-warder ofrece funcionalidades para operar con la herramienta Gephi, lo que va a facilitar el análisis del grafo de interacciones entre usuarios.





## 3 DESARROLLO DEL TFG

### 3.1 Proceso de Instalación de T-warder

#### 3.1.1 Twitter API y credenciales

Para solicitar acceso a la API de Twitter, será necesario tener creada una cuenta de la misma red social. A la hora crear la cuenta, nos van a solicitar una serie de datos como puede ser el correo electrónico, el nombre de usuario que se quiere usar y, por último, la fecha de nacimiento, tal y como se puede observar en la Figura 3-1.



The image shows a screenshot of the Twitter account creation interface. The title is "Crea tu cuenta". There are three main input fields: "Nombre" (with a red arrow pointing to "NOMBRE DE USUARIO"), "Correo electrónico" (with a red arrow pointing to "CORREO ELECTRÓNICO"), and "Fecha de nacimiento" (with a red oval around the date pickers and a red arrow pointing to "FECHA DE NACIMIENTO"). Below the date pickers is a "Siguiente" button. There is also a "Usar teléfono" option and a disclaimer about public information.

Figura 3-1 Página de creación de una cuenta de Twitter

Posteriormente, la plataforma nos va a enviar un correo para verificar que la dirección introducida es la correcta y tendremos que introducir el código de verificación que se incluye en el correo en la página de registro. Ambas páginas las podemos ver en la Figura 3-2.

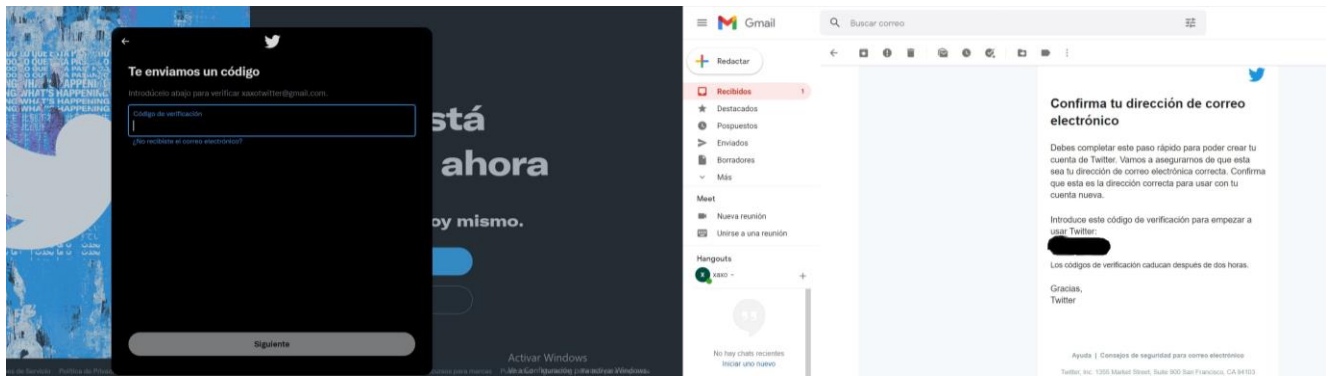


Figura 3-2 Email de verificación y página para introducir el código

Una vez que se ha realizado el paso de verificación nos pedirá introducir una contraseña y por último confirmar el idioma en el que queremos usar Twitter.

Tras haber finalizado la creación de una cuenta de Twitter, procederemos a solicitar acceso al apartado de desarrollador de Twitter.

El siguiente paso será dirigirnos a la página de desarrolladores de Twitter [53] para la solicitar permiso para utilizar la API de Twitter (véase Figura 3-3). Una vez hayamos accedido a la página de desarrolladores, buscaremos en la sección derecha de la página el botón de *Create an App*.

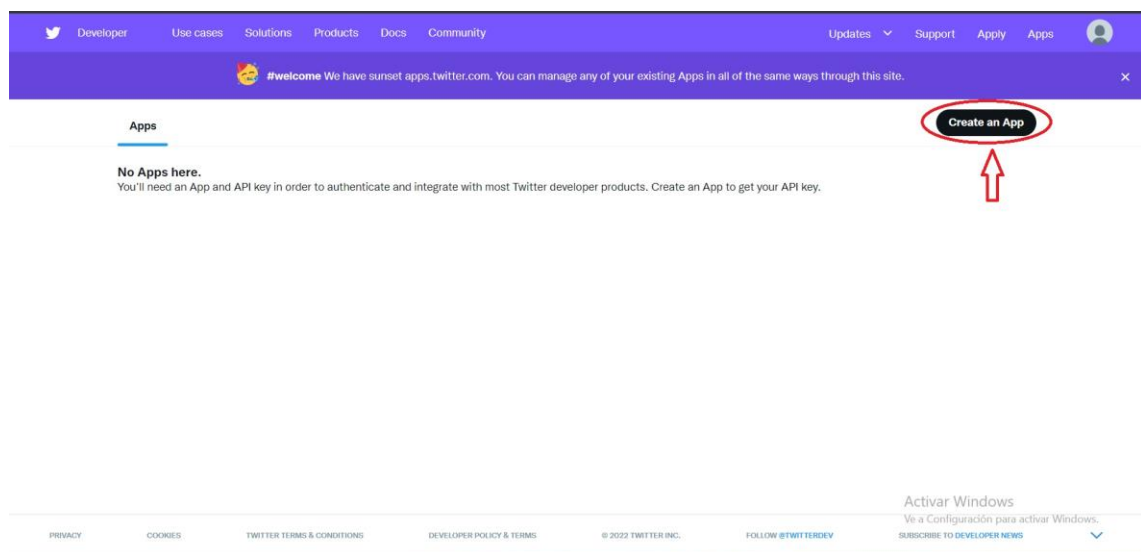


Figura 3-3 Interfaz de la página de desarrolladores de Twitter

Al realizar clic, nos saldrá una pestaña donde nos van a pedir más datos como nuestro nombre o el propósito de crear la aplicación, como observamos en la Figura 3-4. En nuestro caso, queremos las credenciales para uso de estudiante, por lo tanto, seleccionaremos *Student* en la pestaña despegable. A lo hora de seleccionar este apartado es solo meramente informativo para saber si eres estudiante o cualquier otra profesión.

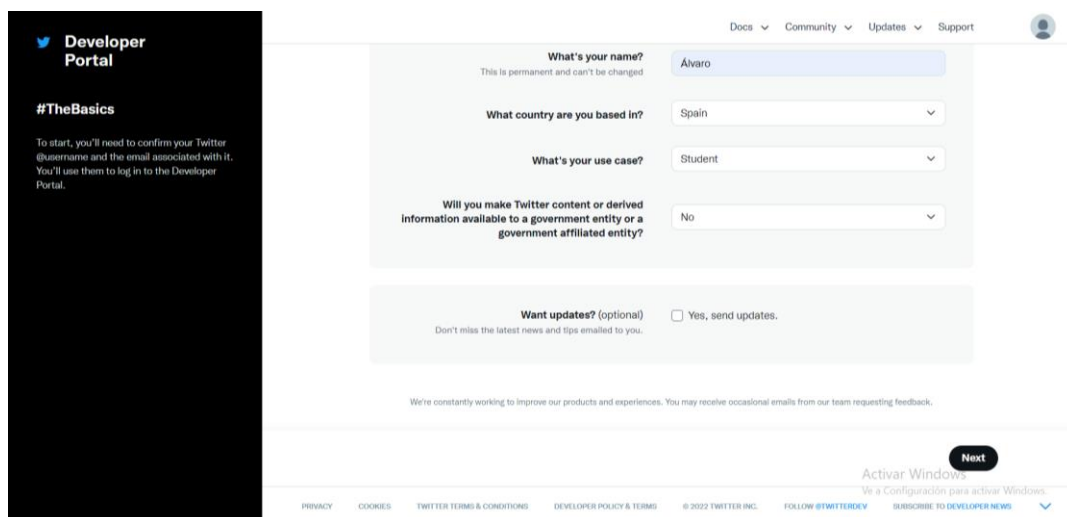


Figura 3-4 Página de datos a introducir

Una vez que le damos al botón *Next*, nos saltará una página sobre política de privacidad donde le daremos a aceptar. Al darle clic, nos saldrá una pantalla tal y como observamos en la Figura 3-5, donde nos pide que vayamos al correo electrónico y verifiquemos que somos nosotros. Se verifica de nuevo ya que anteriormente hemos verificado nuestra dirección para crear el usuario de Twitter y esta vez lo utilizamos para autorizar la elevación de la cuenta a una cuenta de desarrollador.

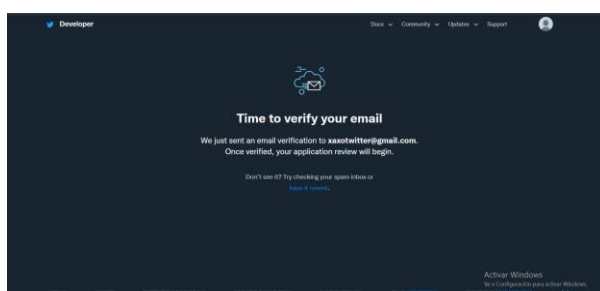


Figura 3-5 Página de verificación de la aplicación de Twitter

Ahora para continuar, deberemos darle un nombre a la aplicación, como vemos en la Figura 3-6. Una vez hayamos creado la aplicación podremos tener acceso a las credenciales de la API de Twitter.

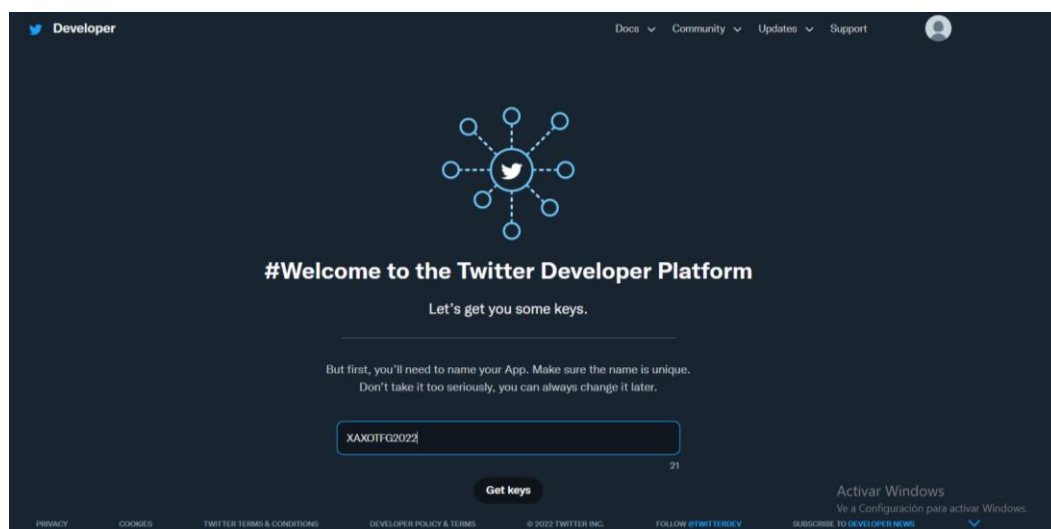


Figura 3-6 Página de creación de la aplicación de Twitter

Al crear la aplicación, nos proporcionarán tres claves, de las cuales nos quedaremos con dos de ellas: con la *API key* y con la *API key secret*, tal como se observa en la Figura 3-7, que son las necesarias para utilizar la herramienta T-hoarder.

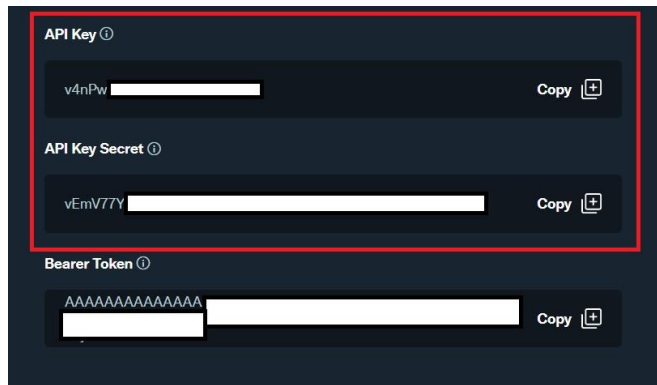


Figura 3-7 API key y API key secret

Hay que destacar que actualmente existen tres niveles de acceso a la API de Twitter: *Essential*, *Elevated* y *Academical*. La principal diferencia entre ellos es el nivel de procesamiento de tuits en un mes, siendo el mayor el nivel *Academical* con 10 millones de tuits al mes, le sigue el nivel *Elevated* con un procesamiento de 2 millones de tuits al mes y por último está el nivel *Essential* con 500000 tuits al mes [46].

En la Figura 3-8, también se puede observar que el nivel *Essential* no tiene acceso a los *endpoints* de la versión anterior de la API de Twitter debido a que el propio Twitter lo limita por tema de seguridad, por lo tanto, para poder seguir con la realización del trabajo se solicitó a Twitter el nivel *Elevated* para poder analizar más tuits.

	Essential	Elevated	Elevated+ (coming soon)	Academic Research
Getting access	<a href="#">Sign up</a>	Apply for additional access within the developer portal	Need more? <a href="#">Sign up for our waitlist</a>	<a href="#">Apply for additional access</a>
Price	Free	Free		Free
Access to Twitter API v2	✓	✓		✓
Access to standard v1.1	✗	✓		✓
Access to premium v1.1	✗	✓		✓
Access to enterprise	✗	✓		✓
Project limits	1 Project	1 Project		1 Project
App limits	1 App per Project	3 Apps per Project		1 App per Project
Tweet caps	Retrieve up to 500k Tweets per month	Retrieve up to 2 million Tweets per month		Retrieve up to 10 million Tweets per 100 days <small>month: Configuración para activar Win</small>

Figura 3-8 Limitaciones de los niveles de la API de Twitter [54]

### 3.1.2 VirtualBox y Linux

Para la instalación de la herramienta, utilizaremos como sistema operativo Linux. En nuestro caso vamos a utilizar una máquina virtual para no tener que realizar ninguna partición del disco duro e instalar el sistema operativo.

Para utilizar una máquina virtual necesitamos instalar el programa de la empresa Oracle *VirtualBox* [55]. Posteriormente, se procederá a instalar el sistema operativo en la máquina virtual. Para la instalación del sistema operativo se puede realizar de dos maneras: creando nosotros mismos la máquina virtual o ejecutando una máquina virtual existente, como veremos ahora a continuación.

### 3.1.2.1 Creando la máquina virtual

Dentro de *VirtualBox*, le hacemos clic a “crear una nueva máquina virtual”, tal y como se observa en la Figura 3-9.



Figura 3-9 Creación de una nueva máquina virtual [56]

Al hacer clic, nos saldrá una serie de pestañas donde podremos seleccionar el tamaño de memoria RAM que queremos darle a la máquina virtual y cuánto espacio queremos darle a la misma [56]. Por defecto, *VirtualBox* nos pondrá los valores mínimos recomendados para que la máquina virtual funcione de manera correcta. Al finalizar la selección de valores nos saldrá la siguiente pestaña, como se puede observar en la Figura 3-10, y tendremos la configuración ya preparada para seleccionar la imagen con el sistema operativo e instalarlo en la máquina virtual.

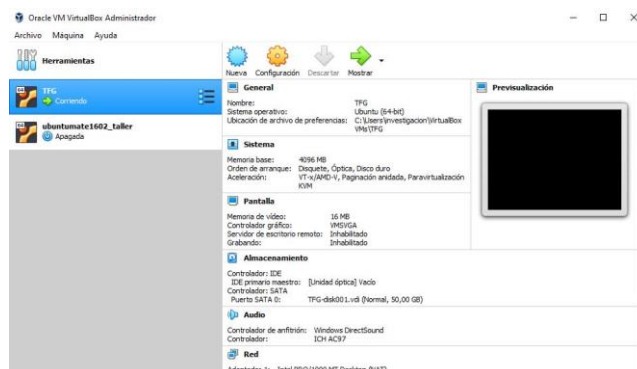


Figura 3-10 Máquina virtual preparada para instalación del sistema operativo

Cuando abrimos *VirtualBox*, nos abrirá una pestaña donde tendremos que indicar a la máquina virtual dónde se encuentra la imagen iso del sistema operativo en nuestro ordenador. Una vez seleccionada la imagen iso, se procederá a instalar el sistema operativo, como vemos en la Figura 3-11.



Figura 3-11 Interfaz de instalación de Ubuntu [57]

Una vez instalado, ya podremos trabajar con la máquina virtual.

### 3.1.2.2 Máquina virtual ya existente

Hoy en día, los propios desarrolladores de Linux distribuyen en sus páginas web los archivos imagen de una máquina virtual ya creada con el sistema operativo configurado [58], como se puede observar en la Figura 3-12. De hecho, en el caso de la herramienta T-warder, los desarrolladores proporcionan la imagen de una máquina virtual con la herramienta instalada.

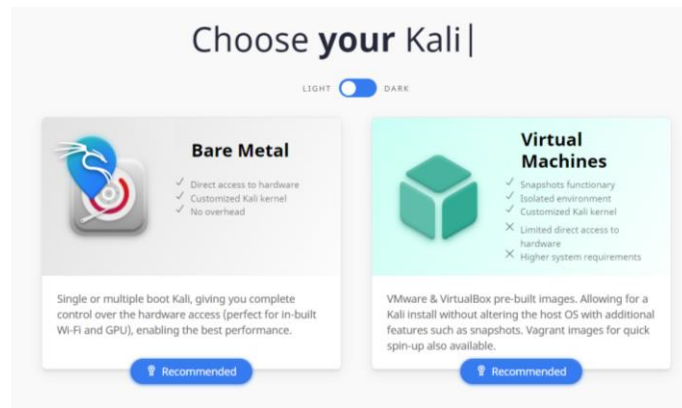


Figura 3-12 Página de descarga de KaliLinux

La gran ventaja de las máquinas virtuales ya creadas es que, a la hora de realizar la instalación, es un método mucho más fácil, ya que solo es necesario abrir el archivo que contiene la máquina virtual y ya se podría trabajar sobre él. Además, se pueden instalar programas y guardar posteriormente la máquina virtual para distribuirla a quién la necesite.

### 3.1.3 T-warder

En el caso de que se elija instalar la herramienta en un sistema (en lugar de utilizar la máquina virtual pre-configurada) el proceso de instalación bastante simple. Antes de la instalación es necesario tener instalado Python 2.7 (en particular 2.7.12 o posterior). Un inconveniente que tiene T-warder es que no se puede utilizar con Python 3.x [59].

También es necesario tener instalada la librería Tweepy [60], algo que es bastante sencillo como se muestra en la Figura 3-13, ya que utilizando un único comando en la consola se lleva a cabo la instalación.

```
tfg@tfg:~$ pip install tweepy
```

Figura 3-13 Comando de instalación de Tweepy

Una vez instalado Tweepy, procederemos a instalar T-warder mediante el comando que podemos observar en la Figura 3-14, haciendo que descarguemos la carpeta con todos los archivos necesarios para que funcione.

```
tfg@tfg:~$ git clone https://github.com/congosto/t-warder_kit
```

Figura 3-14 Comandos de instalación de T-warder

Cuando finaliza la instalación, es necesario crear dos carpetas, llamadas *keys* y *store*, dentro del directorio de la herramienta, por lo que utilizaremos los comandos de la Figura 3-15 para crear las carpetas necesarias y empezar a trabajar con la herramienta.

```
tfg@tfg:~$ cd t-hoarder_kit
tfg@tfg:~/t-hoarder_kit$ mkdir keys
tfg@tfg:~/t-hoarder_kit$ mkdir store
```

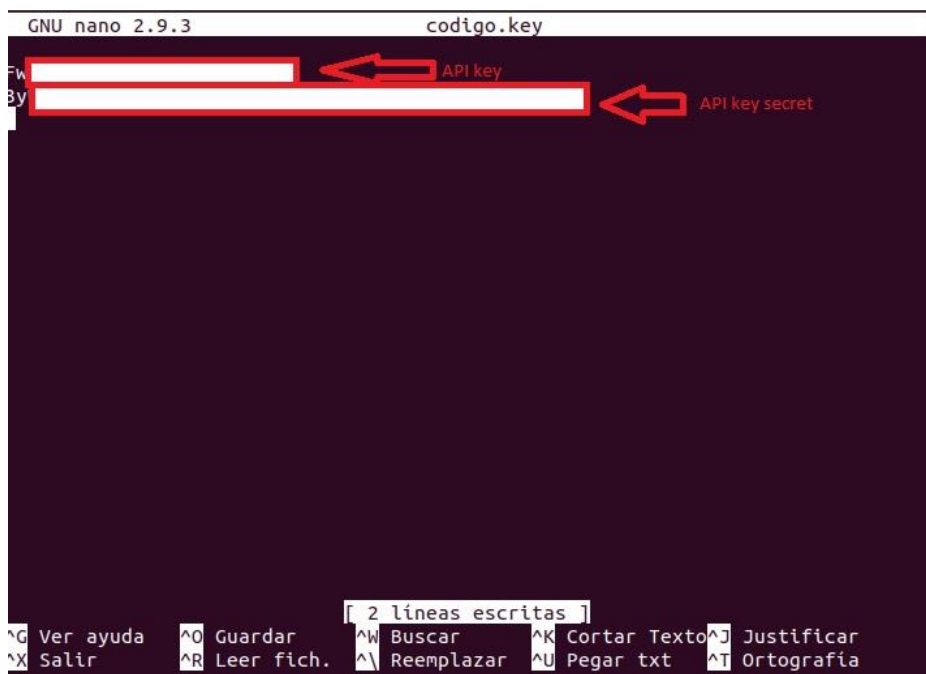
Figura 3-15 Comandos de creación de las carpetas necesarias

Antes de empezar a trabajar, al ejecutar la herramienta nos solicitará la *API key* y la *API key secret*, anteriormente obtenidas. Tendremos que crear un fichero *.key* en la carpeta de *keys*, tal y como se observa en la Figura 3-16. En él guardaremos las claves.

```
tfg@tfg:~/t-hoarder_kit$ cd keys
tfg@tfg:~/t-hoarder_kit/keys$ nano codigo.key
```

Figura 3-16 Comandos de creación del archivo

Una vez abierto el archivo, introducimos las claves sin dejar ningún hueco o fila libre y sin dejar ningún espacio en las líneas, como observamos en la Figura 3-17.



The screenshot shows the GNU nano 2.9.3 editor with the file 'codigo.key' open. Two lines of text are entered: 'API key' on the first line and 'API key secret' on the second line. Red arrows point from the text labels to the corresponding lines in the editor. The status bar at the bottom indicates '2 líneas escritas' (2 lines written).

Figura 3-17 Archivo formato *.key* con las claves

Tras realizar todo este proceso, crearemos una carpeta en el directorio *store* con el nombre del proyecto que queramos llevar a cabo. Esta carpeta será donde se va a almacenar toda la información que recopilamos con la herramienta. Cuando finalicemos todo este proceso, ejecutaremos el programa con el comando de la Figura 3-18.

```
tfg@tfg:~/t-hoarder_kit/scripts$ bash t_hoarder_kit.sh
```

Figura 3-18 Comando para ejecutar el programa

Cuando ejecutemos el programa, tendremos que introducir los datos que podemos ver en la Figura 3-19.



The screenshot shows the program prompts: 'Enter the file name with the application keys: codigo.key', 'Enter a twitter user: tfgxaxo', and 'Enter experiment name: xaxotfg2022'. Red arrows point from labels to the input fields: 'NOMBRE DEL ARCHIVO CON LOS CÓDIGOS' points to 'codigo.key', 'NOMBRE DE USUARIO' points to 'tfgxaxo', and 'NOMBRE DE LA CARPETA CREADA EN STORE' points to 'xaxotfg2022'.

Figura 3-19 Datos a introducir para ejecutar el programa

Al realizar el anterior proceso nos saltará una ventana como observamos en la Figura 3-20, donde introduciremos la opción 1 para finalizar la instalación de la herramienta. Al seleccionar esta opción lo que hacemos es autorizar el acceso por parte de la herramienta a la API de Twitter. Cuando finalicemos este proceso, estaremos listos para empezar a trabajar con T-hoarder, como veremos en la próxima sección.

```

-----
Enter the file name with the application keys: codigo.key
Enter a twitter user: tfgxaxo
Enter experiment name: xaxotfg2022
-----
Working with:
  app: codigo.key
  user: tfgxaxo
  experiment: xaxotfg2022
-----
What function do you want to run?
-----
1. Get a user token access
2. Get users information (profile | followers | following | relations | tweets
| role | list)
3. Make a query on Twitter
4. Get tweets on real time
5. Generate the declared relations graph (followers or following or both)
6. Generate the dynamic relations graph (RT | reply | mentions)
7. Processing tweets (entities| classify| users | spread)
8. utils (sort| remove-duplicate-tweets| convert-to-csv | user-cards| add-commu
nities |spread-by-communities| get_photos-community| twarc2thoarder)
9. Exit
--> Enter option: █

```

Figura 3-20 Pantalla de inicio de la herramienta

## 3.2 Funciones de T-hoarder

Como observamos en la Figura 3-20, la herramienta T-hoarder nos ofrece una serie de funciones. En este apartado se procederá a describir cada función que tiene la herramienta y qué datos podemos obtener a partir de ellas.

### 3.2.1 Información sobre perfiles de Twitter

En esta función, elaborando un listado de perfiles de Twitter en un archivo de texto, podremos sacar información de estas cuentas, como quién le sigue (*followers* en inglés) o a quién sigue (*followees* en inglés), los tuits que publica o la información básica de la cuenta, como cuándo se creó o el nombre real del usuario, si lo tiene habilitado.

A continuación, se expondrá un ejemplo de cómo sacar la información de las cuentas. Primeramente, tendremos que crear el archivo de texto con las cuentas a analizar y guardarlo en la carpeta de proyecto que hemos creado anteriormente en la carpeta *store*, es decir, la carpeta de nuestro trabajo, tal y como podemos ver en la Figura 3-21.

```

tfg@tfg:~/t-hoarder_kit/store/xaxotfg2022$ nano user_list.txt
GNU nano 2.9.3 user list.txt
fabrizioromano
realmadrid
marca
pelica999
mane_ugarte12

```

Figura 3-21 Comando de creación del archivo de usuarios

A continuación, vamos a pedir que nos dé primeramente los tuits que han publicado los usuarios, por lo que al abrir la herramienta seleccionaremos la opción 2. Al seleccionar esta opción, como se puede ver en la Figura 3-22, nos solicita qué información queremos obtener de las cuentas. Como queremos ver qué tuits han publicado escribiremos la palabra *tweets*.



```
--> Enter option: 2
Enter input file name with the list of users or list of profiles (each user in
a line): user_list.txt
Enter an option (profile | followers | following |relations | tweets| role | li
st) : tweets
```

Figura 3-22 Selección de la opción de analizar tuits de usuarios

Una vez finalizado el proceso, se generará un archivo de texto en donde se almacenan los tuits más recientes de los usuarios, como podemos observar en la Figura 3-23. Al abrir el archivo, nos saldrá los tuits para cada uno de los usuarios que hemos puesto en el archivo de usuarios (en orden) y dentro de cada usuario, ordenará los tuits de forma cronológica.

```
tfg@tfg:~/t-hoarder_kit/store/xaxotfg2022$ ls
nato-tweets-query.txt      user_list_log.txt      user_list.txt
nato-tweets-query.txt.log user_list_tweets.txt
GNU nano 2.9.3 user_list_tweets.txt

#Id tweet          date      author text      app      id user followers  follow$
1494341394512560129 2022-02-17 16:02:06 @FabrizioRomano @Cmoffiziell >$
1494333319583977475 2022-02-17 15:30:01 @FabrizioRomano Excl: AC Milan$
1494329759441903627 2022-02-17 15:15:52 @FabrizioRomano Done deal, Ger$
1494307015786696706 2022-02-17 13:45:30 @FabrizioRomano Frank Lampard $
1494268474050424833 2022-02-17 11:12:21 @FabrizioRomano Chelsea are pl$
1494231896347189248 2022-02-17 08:47:00 @FabrizioRomano Antonio Conte $
1494117809156153344 2022-02-17 01:13:39 @FabrizioRomano Two goals toni$
1494077764638547980 2022-02-16 22:34:32 @FabrizioRomano Antonio Conte $
1494029584278425611 2022-02-16 19:23:05 @FabrizioRomano Inter CEO Maro$
1494012821872463872 2022-02-16 18:16:28 @FabrizioRomano There are no c$
14939982920566817668 2022-02-16 17:18:44 @FabrizioRomano AC Milan direc$
1493972643426541569 2022-02-16 15:36:49 @FabrizioRomano Bayern talente$
1493962135075962882 2022-02-16 14:55:04 @FabrizioRomano OL president A$
1493959755295035392 2022-02-16 14:45:36 @FabrizioRomano Barça manager $
1493922002062749702 2022-02-16 12:15:35 @FabrizioRomano @hadrien_greni$
1493916403161870345 2022-02-16 11:53:20 @FabrizioRomano Papu Gómez: "I$
1493915650393317381 2022-02-16 11:50:21 @FabrizioRomano French striker$
1493904992461213698 2022-02-16 11:08:00 @FabrizioRomano After spending$
1493877058132398002 2022-02-16 09:17:00 @FabrizioRomano Pep Guardiola $
1493858087529598978 2022-02-16 08:04:00 @FabrizioRomano Kyllan Mbappé:$
1493810155703740560 2022-02-16 04:51:09 @FabrizioRomano @mallnovskyl1$
149377936897306633 2022-02-16 02:43:07 @FabrizioRomano Pep Guardiola:$
```

Figura 3-23 Archivo de texto con los tuits

Ahora bien, si lo que queremos es obtener la información básica de la cuenta tendremos que llevar a cabo la misma acción realizada anteriormente, pero en vez de elegir la opción *tweets* elegiremos la opción *profile*, tal y como vemos en la Figura 3-24.

```
--> Enter option: 2
Enter input file name with the list of users or list of profiles (each user in
a line): user_list.txt
Enter an option (profile | followers | following |relations | tweets| role | li
st) : profile
```

Figura 3-24 Selección de la opción de obtener la información básica de los usuarios

Al finalizar la tarea, obtendremos un archivo de texto en donde nos saldrá la información básica de forma ordenada según hayamos introducido los usuarios en el archivo de configuración (véase Figura 3-25). Este archivo nos da el nombre del usuario, el número de usuarios que sigue y que le siguen, en qué idioma publica además de la localización si la tiene activada en su cuenta.

```
tfg@tfg:~/t-hoarder_kit/store/xaxotfg2022$ ls
nato-tweets-query.txt      user_list_log.txt      user_list_tweets.txt
nato-tweets-query.txt.log user_list_profiles.txt user_list.txt
GNU nano 2.9.3 user_list_profiles.txt

#id user screen_name net      relation      followers  following  $
330262748 FabrizioRomano fabrizioromano root 6969772 1776 13636 $
14872237 realmadrid realmadrid root 40099964 65 $
15095537 marca marca root 5929279 1557 483553 21591 2008-0$
1089496158655926272 Pellica999 pellica999 root 274 335 $
427716989 mane_ugarte12 mane_ugarte12 root 392 352 14903 $
```

Figura 3-25 Archivo de texto con la información de los usuarios

Y así podemos conseguir información *followers/followees* sobre los usuarios que hemos introducido en el archivo de texto. Basta solo con cambiar lo que queremos buscar dentro de la opción 2 y obtendremos toda la información disponible sobre una o varias cuentas.

### 3.2.2 Realizar una búsqueda en Twitter

En esta opción podemos realizar búsquedas en Twitter mediante palabras clave o mediante *hashtags* en donde, al realizar la búsqueda, se obtendrán tuits que contengan esas palabras clave o el *hashtag* que hayamos filtrado. Estos tuits quedarán almacenados en un archivo de texto ordenados de forma cronológica.

En el siguiente ejemplo se expondrá cómo buscar mediante palabras clave o mediante un *hashtag*. Primeramente, será necesario crear un archivo de texto utilizando como nombre la palabra clave para localizarlo fácilmente. Como vemos en la Figura 3-26, el archivo de texto se va a guardar en la carpeta del proyecto que hemos creado anteriormente.

```
tfg@tfg:~/t-hoarder_kit/store/xaxotfg2022$ nano mbappe.txt
```

Figura 3-26 Comando para la creación del archivo de texto en la carpeta del trabajo

Una vez creado el archivo, ejecutaremos la herramienta T-hoarder y seleccionaremos la opción 3 y le damos los datos que nos pide, como la palabra clave y el archivo de texto donde queremos guardar los tuits como se puede ver en la Figura 3-27.

```
--> Enter option: 3
Enter a query (allows AND / OR connectors): mbappe
Enter output file name: mbappe.txt
```

Figura 3-27 Datos a introducir para iniciar la búsqueda

Finalizada la búsqueda, los tuits quedarán guardados en orden cronológico en el archivo de texto que hemos indicado anteriormente. El archivo de texto se queda almacenado en la carpeta del proyecto como se ve en la Figura 3-28.

```
tfg@tfg:~/t-hoarder_kit/store/xaxotfg2022$ ls
mbappe.txt
mbappe.txt.log
nato-tweets-query.txt
nato-tweets-query.txt.log
russia.txt
russia.txt.log
user_list_log.txt
user_list_profiles.txt
user_list_relation_profiles.txt
user_list_tweets.txt
user_list.txt
```

id	tweet	date	author	text	app	id user	followers	follows
1495896625177448450		2022-02-19 18:03:07	@WeightSorrow	RT @ActuFoot: \$				
1495896627816112641		2022-02-19 18:03:07	@Willyanbu	La fortuna con \$				
1495896608190615560		2022-02-19 18:03:03	@Brianxdd7	@Maxstmpdea \$				
1495896582169051139		2022-02-19 18:02:57	@JoBoLan0	RT @powellyow1 \$				
1495896580646526976		2022-02-19 18:02:57	@VvesVveskina2	@ugoat @aain \$				
1495896577131745287		2022-02-19 18:02:56	@LuisPuertoo	mbappe ven ya \$				
1495896555258499875		2022-02-19 18:02:50	@dLsk30320	@realangefred \$				
1495896549784838150		2022-02-19 18:02:49	@AFCMagicMan	@pulsesportsng \$				
1495896545150223338		2022-02-19 18:02:48	@ashatells_	RT @bderhnd: NS				
1495896527911632983		2022-02-19 18:02:44	@yusufhsankoy	RT @Fbi19Anall \$				
1495896517748686848		2022-02-19 18:02:42	@Rosanadiablita	RT @Bianco1San \$				
1495896517467779873		2022-02-19 18:02:41	@wawou16	RT @SlxxRus: \$				
1495896514288796680		2022-02-19 18:02:41	@garignton	@elpentofvcr @ \$				
1495896504243179522		2022-02-19 18:02:38	@kelidah	RT @WinamaxFC: \$				
1495896493413445645		2022-02-19 18:02:36	@jeancarloreneo	RT gde \$				
1495896482764148736		2022-02-19 18:02:33	@pulsesportsng	*Mbappe who? \$				
1495896451214544902		2022-02-19 18:02:26	@moose_olest	RT @CristianoX \$				
1495896443689967626		2022-02-19 18:02:24	@Menez_TB_	@PoliBoard Me \$				
1495896423121106895		2022-02-19 18:02:19	@yas_knr	RT @Pearja: OS				
1495896403150213760		2022-02-19 18:02:14	@jbisaro	@Michael27834 \$				
1495896368381588484		2022-02-19 18:02:06	@Mehd13Al	RT @bderhnd: NS				
1495896349653540867		2022-02-19 18:02:01	@BotComplotiste	Oui, encore mo \$				

Figura 3-28 Archivo de texto con los tuits

En el caso de querer buscar un *hashtag*, se realizará la misma operación, pero cambiando la palabra clave por el *hashtag*, y se almacena el resultado de la misma forma que en el caso anterior.

### 3.2.3 Obtener tuits en tiempo real (streaming)

Mediante esta opción, somos capaces de obtener tuits del flujo (*stream*) de datos que se están publicando en ese momento sobre algún usuario o sobre alguna palabra clave. La metodología para obtener los tuits en tiempo real es bastante parecida a la de obtener información sobre un usuario, ya que tenemos que introducir en un archivo de texto nombres de un usuario, *hashtags* o también se pueden insertar palabras clave para que cualquier tuit que los mencione se almacene en el archivo de texto.

En el siguiente ejemplo se explica de forma más detallada cómo se pueden obtener tuits en tiempo real. En primer lugar, debemos crear un archivo con las palabras clave que deseamos buscar. Como se observa en la Figura 3-29, este archivo se guarda en la carpeta del proyecto.

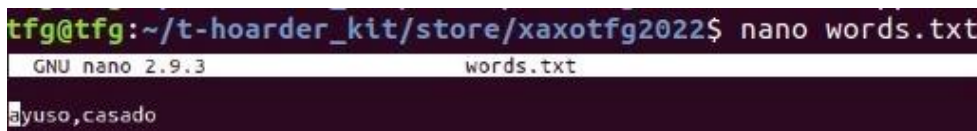


Figura 3-29 Archivo de texto con las palabras clave

Al realizar este paso, abriremos la herramienta y seleccionaremos la opción 4. Como se observa en la Figura 3-30, nos va a solicitar el archivo de texto donde hemos guardado las palabras clave, además de poner nombre al archivo donde queremos que se guarden los tuits sobre las palabras clave.

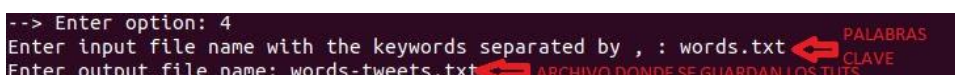


Figura 3-30 Datos a introducir para iniciar la recepción de tuits en tiempo real

La recepción de tuits finaliza de dos maneras: podemos pararla nosotros o la propia herramienta nos restringe para no excedernos del límite de tuits permitidos. Cuando finalice la recepción de tuits, podemos destacar en el archivo de texto de salida que los tuits se almacenan de forma cronológica, tal y como se guardaban en las otras opciones disponibles en la herramienta (véase Figura 3-31).

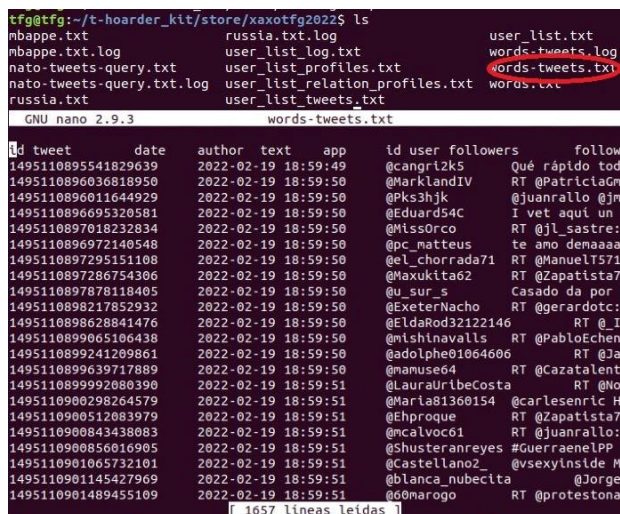


Figura 3-31 Archivo con los tuits recogidos en tiempo real

## 3.3 Procesado de datos con Gephi

Para el procesado de datos de relaciones entre usuarios, utilizaremos la herramienta de código libre Gephi, ya que la herramienta T-warder nos exporta los datos en un formato compatible con este programa. Primero procederemos a instalarlo y, posteriormente, veremos un ejemplo del uso combinado de T-warder y Gephi.

### 3.3.1 Instalación de Gephi

Como se observa en la Figura 3-32, debemos introducir el comando en la consola para poder instalar Gephi. Se observa que el archivo descargado es un archivo comprimido, por lo que tendremos que proceder a extraerlo.

```
tfg@tfg:~$ wget -q --show-progress https://github.com/gephi/gephi/releases/download/v0.9.2/gephi-0.9.2-linux.tar.gz
```

Figura 3-32 Comando de descarga de Gephi

Una vez descarguemos el programa, mediante el comando *tar*, procederemos a descomprimirlo y posteriormente lo ejecutaremos con el comando *gephi* (Figura 3-33).

```
tfg@tfg:~$ tar xzf gephi-0.9.2-linux.tar.gz
tfg@tfg:~$ cd Escritorio/gephi-0.9.2/
tfg@tfg:~/Escritorio/gephi-0.9.2$ cd bin
tfg@tfg:~/Escritorio/gephi-0.9.2/bin$ ./gephi
```

Figura 3-33 Comandos de descompresión y ejecución

### 3.3.2 Procesado de datos

Una vez hemos realizado una búsqueda de tuits con cualquiera de los mecanismos que ofrece T-warder, empezaremos a procesar los tuits que hemos obtenido (Figura 3-34). Seleccionaremos la opción 6 en el menú de T-warder e indicaremos el tipo de relación que queremos analizar. Podemos elegir entre tres tipos de relaciones: según las menciones que haya tenido el tuit, cuantos *retuits* ha tenido el tuit y las respuestas que tiene el tuit. En este caso, hemos elegido la opción RT (*retuit*), como se ve en la Figura 3-34.

```
--> Enter option: ^C--> Enter option: 6
Enter input file name with the tweets (got from a query or in real time): pp-twe
ets-query.txt
Enter the relationship type (RT | reply | mention): RT
Introduce top size (100-50000): 20000
file name pp-tweets-query.txt
```

Figura 3-34 Selección del procesado de datos en la herramienta

T-warder generará con esto un archivo con extensión *gdf* que podremos abrir con Gephi. Una vez finalizado el proceso, abriremos Gephi y seleccionaremos el archivo que nos ha generado la herramienta (Figura 3-35).

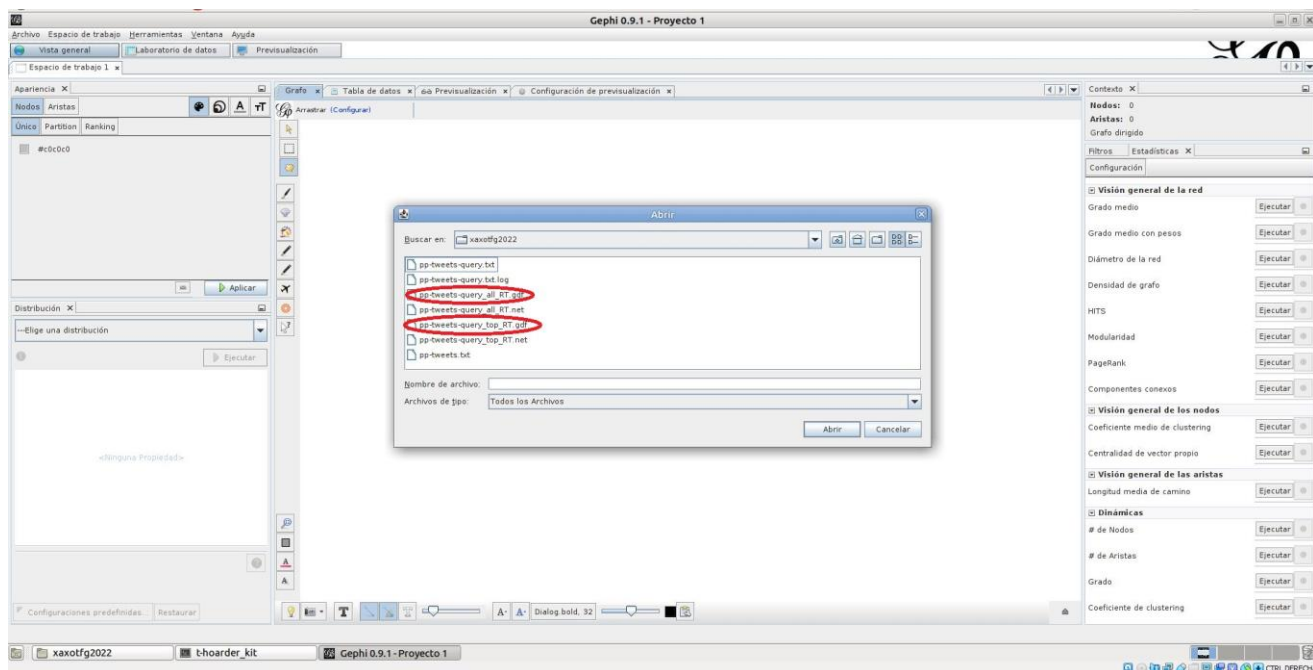


Figura 3-35 Interfaz gráfica de Gephi

Una vez cargados los datos del grafo en Gephi, podemos aplicar alguna de las opciones de análisis que proporciona la herramienta. En este ejemplo, aplicaremos el análisis de modularidad, que según [61] es una medida de la estructura de las redes o grafos que fue diseñado para medir la fuerza de la división de una red en módulos. Posteriormente de aplicar el análisis de la modularidad, vamos a dar tamaño a los nodos y cambiarlos de color para poder diferenciarlos de una forma más clara, tal y como observamos en la Figura 3-36.

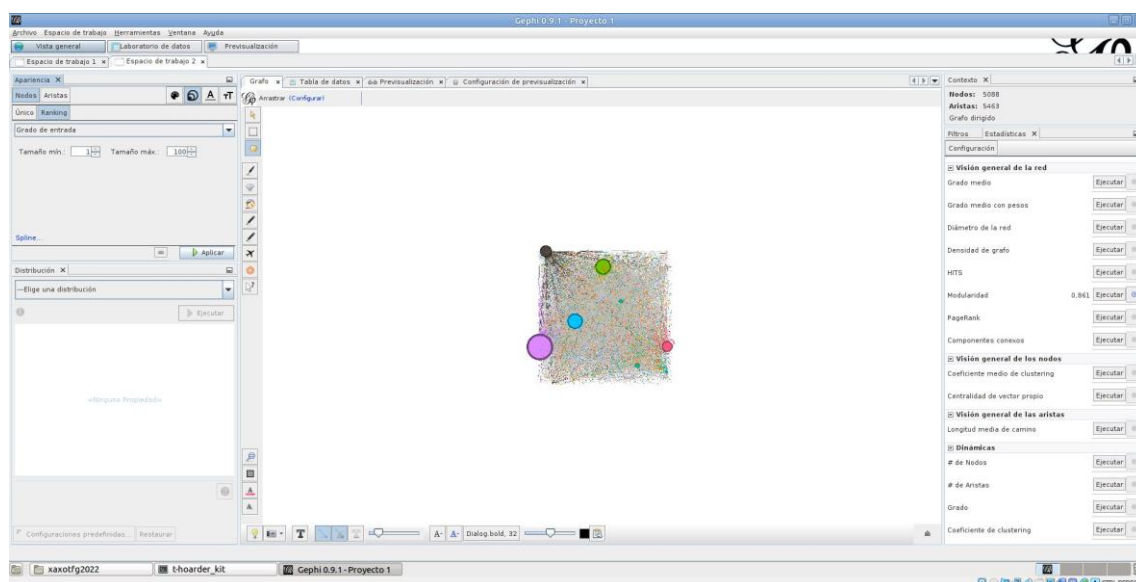
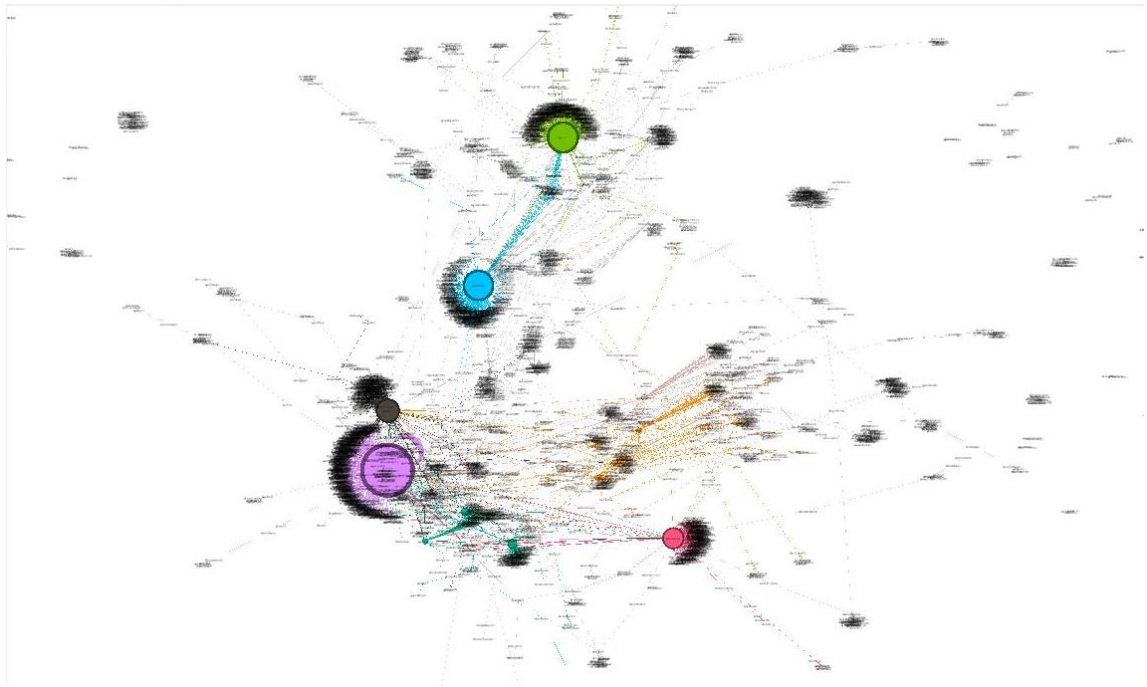


Figura 3-36 Nodos de los datos diferenciados

Para la visualización de los grafos, aplicaremos un algoritmo denominado *ForceAtlas2*, haciendo así que la representación del grafo sea más sencilla para el que lo visualiza (véase Figura 3-37).



**Figura 3-37 Grafo final**

Como podemos observar en la figura anterior, vemos que en el grafo final se ve de manera más sencilla la relación entre los usuarios que han interactuado mediante el mecanismo de *retuit*. Vemos cómo se agrupan en una serie de bloques donde observamos que los tuits nacen de 5 principales fuentes y que el resto de usuarios interactúan con los tuits de estos usuarios.

## 4 CASO DE USO

### 4.1 Comparativa de datos obtenidos mediante API Streaming y API Search

#### 4.1.1 API Streaming

Mediante la opción 4 de la herramienta T-hoarder, hemos obtenido un archivo de texto de 1,2 GB en donde podemos encontrar tuits publicados que contengan las palabras “russia” o “ukraine” desde las 19.30h (hora española) del día 01/03/2022 hasta las 07.30h del día 02/03/2022, representando un total de 1.816.544 tuits. Una vez hemos obtenido esta información vamos a proceder a evaluarla.

Primeramente, procederemos a importar los datos del archivo de texto en el software de Microsoft Excel mediante una opción que nos proporciona esta herramienta. Cuando se importen los datos del archivo de texto, por la misma configuración que tiene el archivo, quedarán separados en distintas celdas o columnas, como podemos observar en la Figura 4-1. Podemos destacar en la primera columna es el número identificador de tuit, la segunda la fecha de publicación, la tercera el usuario que publica el tuit y la cuarta el cuerpo del tuit (véase Figura 4-1), además del número de seguidores, el cuerpo del tuit o las imágenes que contenga.

	id tweet	date	author	text	app	id user	followers
1	1,49874E+18	01/03/2022 19:14:48	@ghostofidiligent	RT @ConceptualJames: When it's Taiwan instead of Ukraine, we won't...	Twitter for Android	1,49304E+18	
2	1,49874E+18	01/03/2022 19:14:48	@sciubidu	RT @markuspreiss: Noch haben Russlands Generale vielleicht eine letz...	Twitter for Android	34318759	
3	1,49874E+18	01/03/2022 19:14:48	@DoaSiddiqui	RT @ZaidZamanHamid: These rascals knew exactly what was coming t...	Twitter for Android	1,1638E+18	
4	1,49874E+18	01/03/2022 19:14:48	@jmh6225	RT @TheRaisedVoice: Ukrainians are valiantly fighting for their countr...	Twitter for Android	1,20529E+18	
5	1,49874E+18	01/03/2022 19:14:48	@SylvieClause	RT @sarrasan: @KARIMAMARKHODJ1 https://t.co/YpKsrsEWj	Twitter for Android	1,47071E+18	
6	1,49874E+18	01/03/2022 19:14:48	@TraderMom2	RT @AJentleson: As US banks face imminent Russian cyberattacks, ...	Twitter for iPhone	1,21787E+18	
7	1,49874E+18	01/03/2022 19:14:48	@AngelikaHTCG	RT @NetZeroWatch: "[Ukraine is] a wake up call. Everything has chang...	Twitter for iPad	798338324	
8	1,49874E+18	01/03/2022 19:14:48	@pryyasndkh	Please help support Ukraine, add a #Twibbon now! https://t.co/RTMJUY...	Twibbon	1,35272E+18	
9	1,49874E+18	01/03/2022 19:14:48	@kathyspence777	Trump has been a whipping boy since he left office. Say what you will ...	Twitter Web App	1,49777E+18	
10	1,49874E+18	01/03/2022 19:14:48	@mariangelesmelc	RT @mikelayestaran: Entrando a #Kiev después de un largo viaje en tr...	Twitter for Android	2184755390	
11	1,49874E+18	01/03/2022 19:14:48	@MiraCat777	RT @kylegriffin1: Pool report: President Biden held a call with Preside...	Twitter Web App	8,23042E+17	
12	1,49874E+18	01/03/2022 19:14:48	@Ashme13	RT @Kyvindependent: CNN: Russia fired more than 400 missiles on ...	Twitter for Android	51317298	
13	1,49874E+18	01/03/2022 19:14:48	@UlrichMathurin	RT @CalitoActivism: BREAKING: Zelensky says before Ukraine sits dow...	Twitter for Android	2486176181	
14	1,49874E+18	01/03/2022 19:14:48	@rinne_tensei1	RT @DazzaOssamu: @shelbiagota Russia sendo mais demonizada em...	Twitter Web App	1,20443E+18	
15	1,49874E+18	01/03/2022 19:14:48	@farismatisa	All pressure on Russia now. I personally do not think Putin will ever giv...	Twitter for iPhone	589606591	
16	1,49874E+18	01/03/2022 19:14:48	@ukfathers	INTERVIEW: Former UK Ambassador Peter Ford Warns About NATO Es...	Twitter Web App	1331464566	
17	1,49874E+18	01/03/2022 19:14:48	@rajikumar128	I did argue in previous tweets that Transsexuals should be allowed to f...	Twitter for iPhone	51122403	
18	1,49874E+18	01/03/2022 19:14:48	@Bear_Gardner	RT @sbattawden: Myself and a team of NHS doctors are raising mone...	Twitter for Android	2883184660	
19	1,49874E+18	01/03/2022 19:14:48	@charlesmead	RT @FPWellman: If you believe for a second that so many GOP Congre...	Twitter for iPhone	14467295	
20	1,49874E+18	01/03/2022 19:14:48	@DrAlakbarov	At the Ukraine-Poland border, MSF teams are currently trying to get e...	Twitter Web App	2318119938	
21	1,49874E+18	01/03/2022 19:14:48	@GW_LoRider	RT @ananavarro: How stupid. @marcorubio's one of highest ranking ...	Twitter for Android	3169683286	
22	1,49874E+18	01/03/2022 19:14:48	@ruff_1	@JulieBanderas Could've been Belgrade in 1998, it didn't appear that ...	Echofon	19340122	
23	1,49874E+18	01/03/2022 19:14:48	@IT_not_real	RT @DavidCornDC: @RepMcGovern also says he wishes @GOPLeader...	Twitter Web App	9,74492E+17	
24	1,49874E+18	01/03/2022 19:14:48	@HallJE9	RT @SenTedCruz: Putin funds his war machine and his invasion of Ukr...	Twitter for Android	7,16435E+17	
25	1,49874E+18	01/03/2022 19:14:48	@astroUrsaSpace	RT @gbazov: 3—#MARIUPOL—City has been completely encircled. Ca...	Twitter for Android	1,27293E+18	
26	1,49874E+18	01/03/2022 19:14:48	@MoTurn17	RT @peckham65Grant: I'm very surprised all are top journalists,Laura ,...	Twitter for Android	41170964	
27	1,49874E+18	01/03/2022 19:14:48	@licerila	RT @suetens2: Les occidentaux veulent lancer une procédure contre l...	Twitter for Android	135997560	
28	1,49874E+18	01/03/2022 19:14:48	@PaulxChuks	Chee. See Holy Ghost Inspired analysis on Geopolitics & political...	Twitter for Android	1,24505E+18	

Figura 4-1 Archivo de datos importado en Excel

Cuando ya hayamos terminado de importar los datos a *Excel*, procederemos a crear una gráfica ordenada indicando cuantos tuits ha podido escribir un usuario en función de las palabras clave que hemos establecido anteriormente. Una vez creada esta gráfica procederemos a ponerla en orden de forma que aparezcan primero los usuarios que más tuits han publicado, obteniendo así la gráfica que se puede observar en la Figura 4-2.

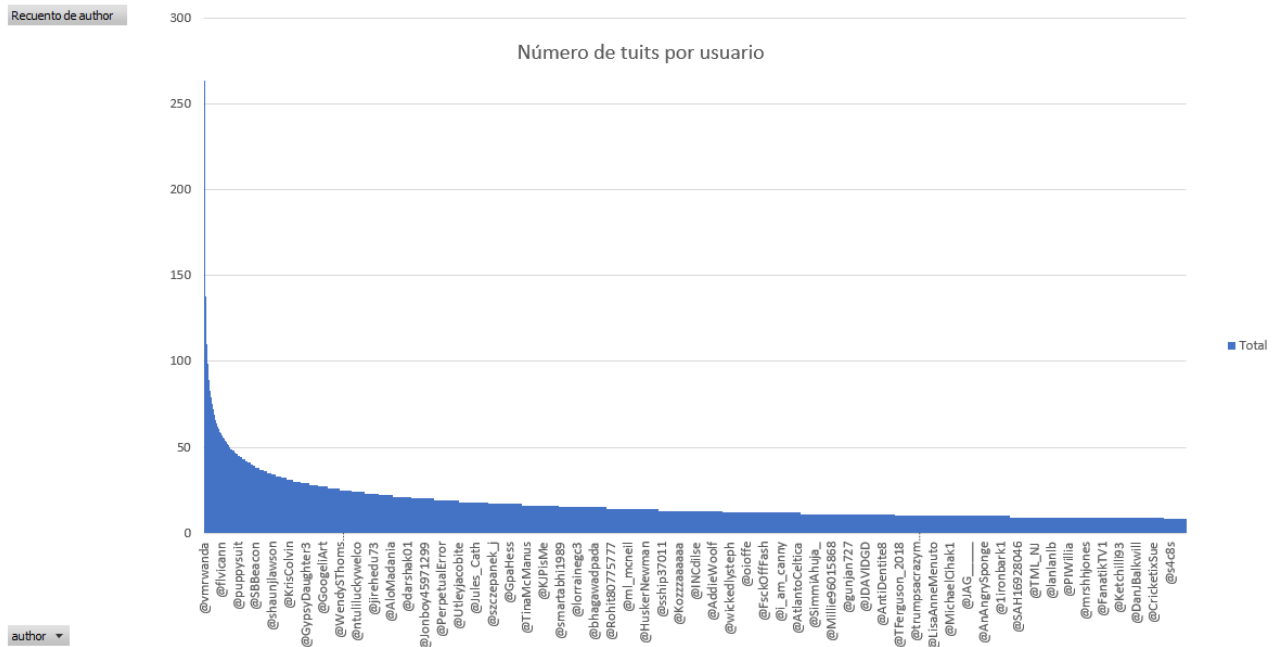


Figura 4-2 Gráfica de número de tuits por usuario (API streaming)

Como observamos en la Figura 4-2, el resultado es una gráfica descendente, por lo que nos interesa focalizarnos en los usuarios con más tuits publicados. Nuestro cometido será hacer otra gráfica, pero seleccionando solo los 20 primeros usuarios, tal y como se puede observar en la Figura 4-3.

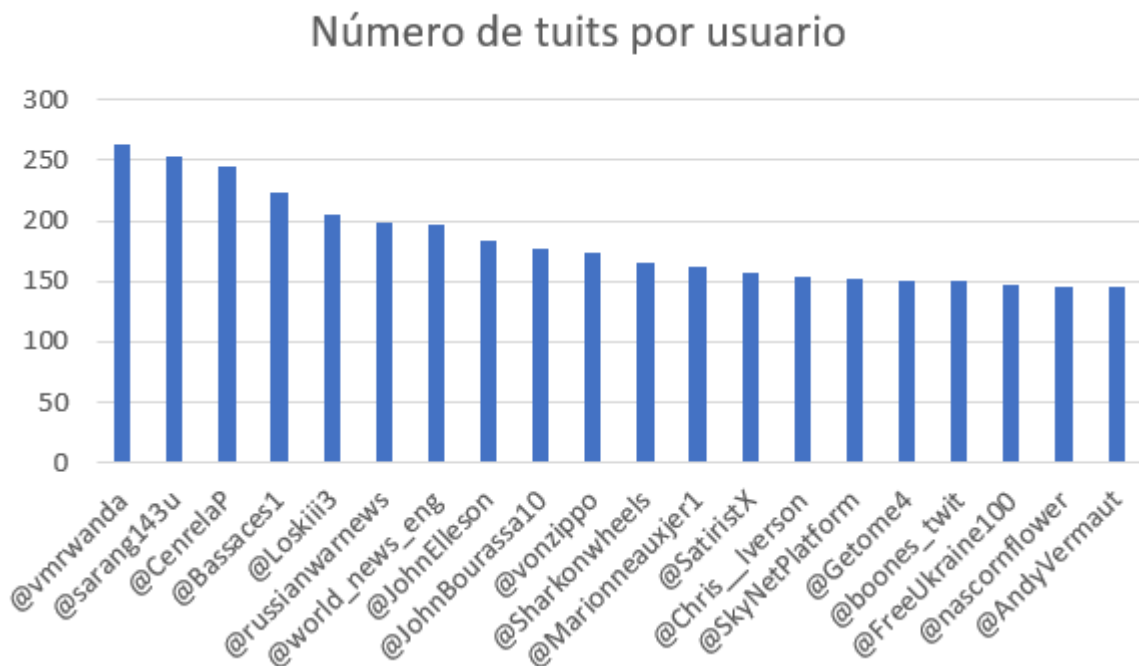


Figura 4-3 Gráfica con los 20 usuarios que más han tuiteado (API streaming)



Los valores obtenidos en estas gráficas no tienen nada que ver con que sean los tuits más vistos o más compartidos, solo que estos usuarios han sido los que más tuits han publicado con las palabras clave que habíamos mencionado anteriormente.

El siguiente paso que vamos a realizar será escribir un código en Python en donde nos diga qué usuarios han sido los que más *retuits*, menciones o respuestas han tenido. El código de Python se adjunta en el Anexo I. Una vez ejecutado el código Python obtendremos el resultado que se ve en la Figura 4-4.

```

=== Autores ===
@vmrwanda;264
@sarang143u;254
@CenrelaP;245
@Bassaces1;223
@Loskiii3;205
@world_news_eng;197
@russianwarnews;197
@JohnElleson;183
@JohnBourassa10;177
@vonzippo;171
@Sharkonwheels;164
@Marionneauxjer1;162
@SatiristX;156
@Chris_Iverson;153
@SkyNetPlatform;152
@boones_tweet;151
@Getome4;151
@FreeUkraine100;147
@AndyVermaut;146
@nascornflower;146
=== Retuiteados ===
KyivIndependent;4767
OccupayDemocrats;1685
POTUS;14449
nexta_tv;10624
sirotek;10183
Reuters;9301
spectatorindex;8354
ANI;6861
nytimes;6444
ZelenskyyUa;5978
CalltoActivism;5627
BijanCNN;5544
joncoopertweets;5268
TristanSnell;4933
ABC;4418
Kasparov63;4336
BNONews;4134
atrupar;4113
IAPONomarenko;3688
PalmerReport;3665
=== Mencionados ===
POTUS;3137
ua_parliament;1794
Ukraine;1090
RepStefanik;1052
KyivIndependent;1016
Reuters;999
FoxNews;968
ZelenskyyUa;833
marcorubio;670
nytimes;588
CNN;554
laurenboebert;543
JackPosobiec;511
RussianEmbassyC;501
ABC;487
Jim_Jordan;458
DonaldJTrumpJr;458
KremlinRussia_E;450
nexta_tv;421
RT_com;392
=== Influencias ===
@PTI_News;3680030.333333335
@TheOnion;2347698.2
@twentyonepilots;2041490.0
@BBCWorld;1944820.888888889
@SkyNewsBreak;1621128.666666667
@POTUS;1595991.694444445
@WhiteHouse;1326416.4
@ndtv;1120645.825
@FLOTUS;747555.2
@JoeBiden;685266.5
@kltraffcupdate;623741.0
@30SECONDDSTOMARS;604259.6666666666
@Leadlagreport;591414.333333334
@ajlnews;554129.2666666667
@republic;538448.5571428571
@PTTVOnlineNews;422943.2222222225
@Republic_Bharat;366416.2
@ndtvindia;349140.5555555556
@VEJA;337832.6538461539
@SkyNews;329848.7652173913
=== Seguidores ===
@CNN;56736926.375
@nytimes;51999501.52380952
@BBCWorld;35006776.0
@JoeBiden;32892792.0
@TheEconomist;26338732.90909091
@Reuters;24575345.44736842
@FoxNews;20810779.85714286
@WSJ;19531744.333333332
@POTUS;19151900.333333332
@washingtonpost;18845481.11764706
@TIME;18701763.57142857
@Forbes;17545884.64864865
@ABC;17232166.6
@ndtv;16809687.375
@aahtak;16385235.390243903
@AP;15600730.3125
@UN;15001356.5
@voguemagazine;14668757.0
@timesofindia;14286043.142857144
@BBCNews;13479477.666666666
    
```

Figura 4-4 Producto del código de Python

Copiando estos datos en un archivo de texto y después exportándolos a un Excel obtenemos las tablas de la Figura 4-5.

=== Autores ===	Columna1	=== Retuiteados ===	Columna1	=== Mencionados ===	Columna2	=== Influencias ===	Columna3	=== Seguidores ===	Columna4
@vmrwanda	264	KyivIndependent	47677	POTUS	3137	@PTI_News	3,68003E+16	@CNN	56736926375
@sarang143u	254	OccupayDemocrats	16856	ua_parliament	1794	@TheOnion	23476982	@nytimes	5,19995E+15
@CenrelaP	245	POTUS	14449	Ukraine	1090	@twentyonepilots	20414900	@BBCWorld	350067760
@Bassaces1	223	nexta_tv	10624	RepStefanik	1052	@BBCWorld	1,94482E+15	@JoeBiden	328927920
@Loskiii3	205	sirotek	10183	KyivIndependent	1016	@SkyNewsBreak	1,62113E+16	@TheEconomist	2,63387E+15
@world_news_eng	197	Reuters	9301	Reuters	999	@POTUS	1,59599E+16	@Reuters	2,45753E+15
@russianwarnews	197	spectatorindex	8354	FoxNews	968	@WhiteHouse	13264164	@FoxNews	2,08108E+15
@JohnElleson	183	ANI	6861	ZelenskyyUa	833	@ndtv	1120645825	@WSJ	1,95317E+16
@JohnBourassa10	177	nytimes	6444	marcorubio	670	@FLOTUS	7475552	@POTUS	1,91519E+16
@vonzippo	171	ZelenskyyUa	5978	nytimes	588	@JoeBiden	6852665	@washingtonpost	1,88455E+15
@Sharkonwheels	164	CalltoActivism	5627	CNN	554	@kltraffcupdate	6237410	@TIME	1,87018E+15
@Marionneauxjer1	162	BijanCNN	5544	laurenboebert	543	@30SECONDDSTOMARS	6,04259E+15	@Forbes	1,75459E+15
@SatiristX	156	joncoopertweets	5268	JackPosobiec	511	@leadlagreport	5,91414E+15	@ABC	172321666
@Chris_Iverson	153	TristanSnell	4933	RussianEmbassyC	501	@ajlnews	5,54129E+15	@ndtv	16809687375
@SkyNetPlatform	152	ABC	4418	ABC	487	@republic	5,38449E+15	@aahtak	1,63852E+16
@boones_tweet	151	Kasparov63	4336	Jim_Jordan	458	@PTTVOnlineNews	4,22943E+16	@AP	1,56007E+11
@Getome4	151	BNONews	4134	DonaldJTrumpJr	458	@Republic_Bharat	3664162	@UN	150013565
@FreeUkraine100	147	atrupar	4113	KremlinRussia_E	450	@ndtvindia	3,49141E+16	@voguemagazine	146687570
@AndyVermaut	146	IAPONomarenko	3688	nexta_tv	421	@VEJA	3,37833E+15	@timesofindia	1,4286E+16
@nascornflower	146	PalmerReport	3665	RT_com	392	@SkyNews	3,29849E+15	@BBCNews	1,34795E+16

Figura 4-5 Tabla de datos del código de Python

Aquí en estas tablas tenemos diferenciadas las cuentas con más *retuits*, las cuentas más mencionadas, aquellas que tienen más seguidores y además se ha calculado un parámetro denominado influencia. La influencia es una ratio derivada del cociente de los seguidores de un usuario entre los usuarios a los que sigue. Si el resultado de este cociente es mayor que 1, podemos decir que es una cuenta influyente, pero si este cociente es menor que 1 no es una cuenta influyente. Cuanto mayor sea esta ratio, más influyente sería este usuario [62].

A continuación, realizaremos una gráfica para cada parámetro. La primera es la gráfica de *retuits* (Figura 4-6).



Figura 4-6 Gráfica de usuarios con más *retuits* (API streaming)

La siguiente gráfica que se va a mostrar es la de los usuarios más mencionados (Figura 4-7).

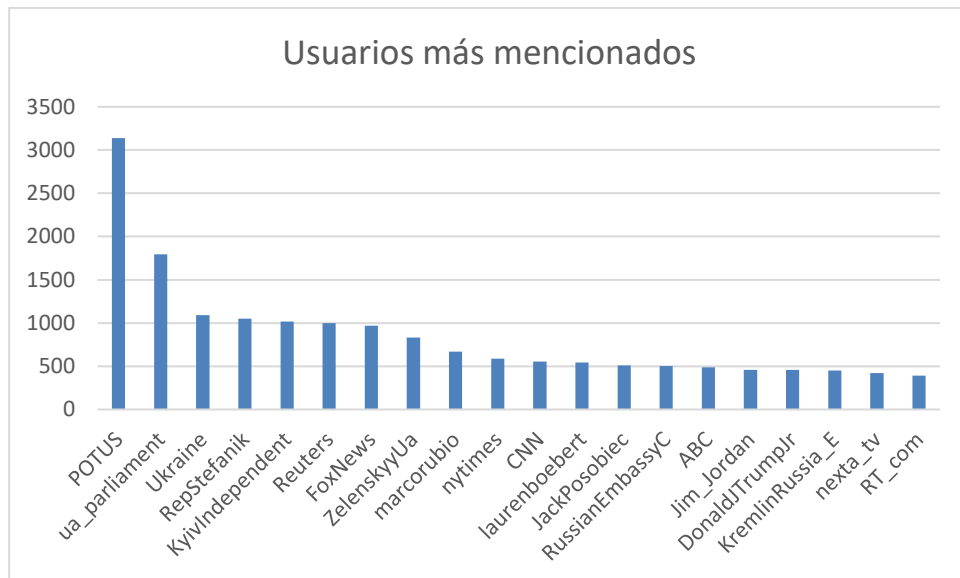


Figura 4-7 Gráfica de los usuarios más mencionados (API streaming)

Por último, pero no menos importante, será la gráfica de los usuarios más influyentes (Figura 4-8).

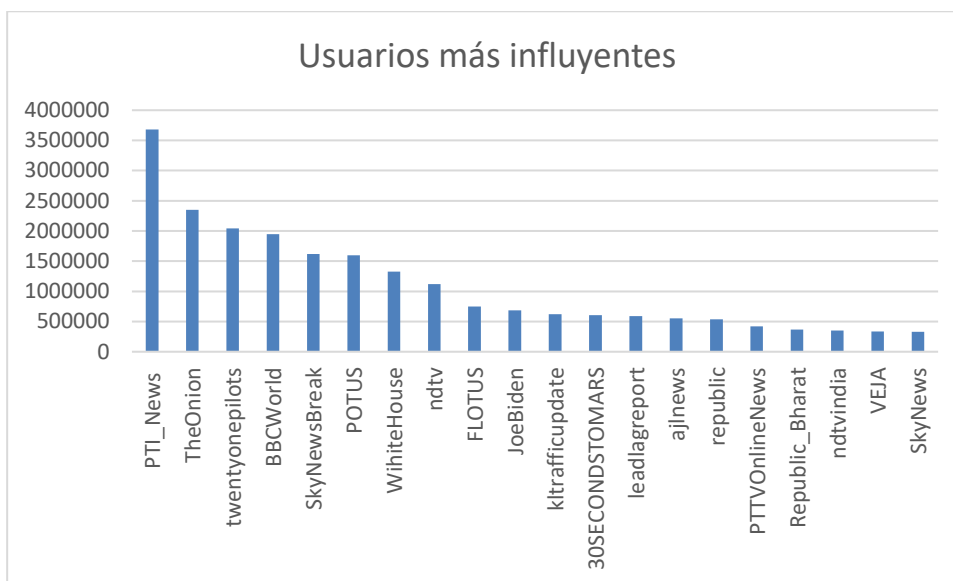


Figura 4-8 Gráfica de los usuarios más influyentes (API streaming)

Una vez realizadas las tres gráficas, observamos que varios usuarios de los que más *retuits* han tenido también se encuentran en los usuarios que más menciones han tenido y más influyentes, por lo que pueden ser usuarios para seguir.

Ahora bien, si lo que deseamos es ver estos resultados en forma de comunidades de ideas e interacciones entre los usuarios, deberemos de realizar un análisis mediante la opción 6 de la herramienta T-warder, que nos permite transformar el archivo de texto generado de la API streaming en un archivo compatible con el software Gephi. En este caso hemos elegido que la relación de los tuits sea de *retuit* (RT). Una vez haya terminado el proceso de conversión del archivo, abriremos el archivo en Gephi y empezaremos a trabajar con él.

Al abrir el archivo en Gephi, procederemos a realizar los pasos narrados en el punto 3.3.2, en el que se explica cómo procesamos el archivo extraído de T-warder. Cuando hayamos procesado el archivo obtendremos el resultado que se muestra en la Figura 4-9.

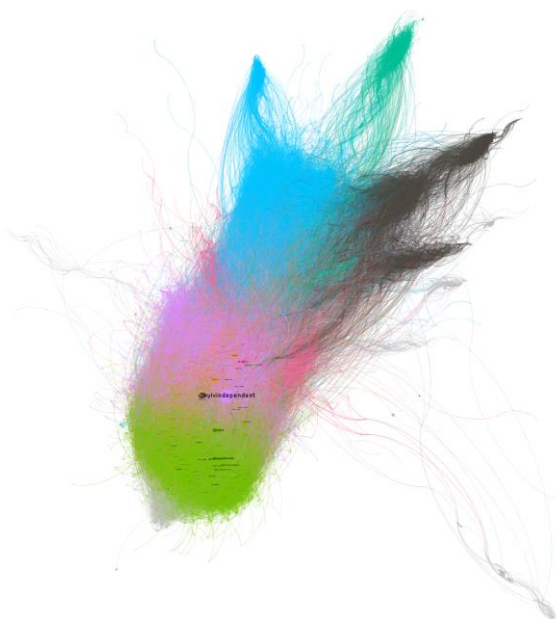
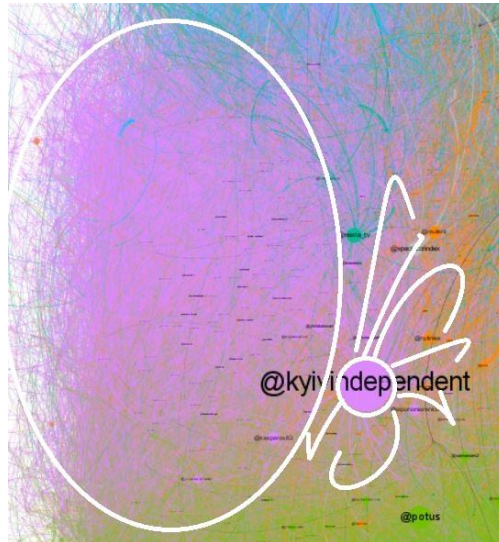


Figura 4-9 Grafo con los *retuits* de la búsqueda realizada

Se observan claramente en el grafo tres comunidades que alimentan a las demás cuentas. Procederemos a explicar cada comunidad.

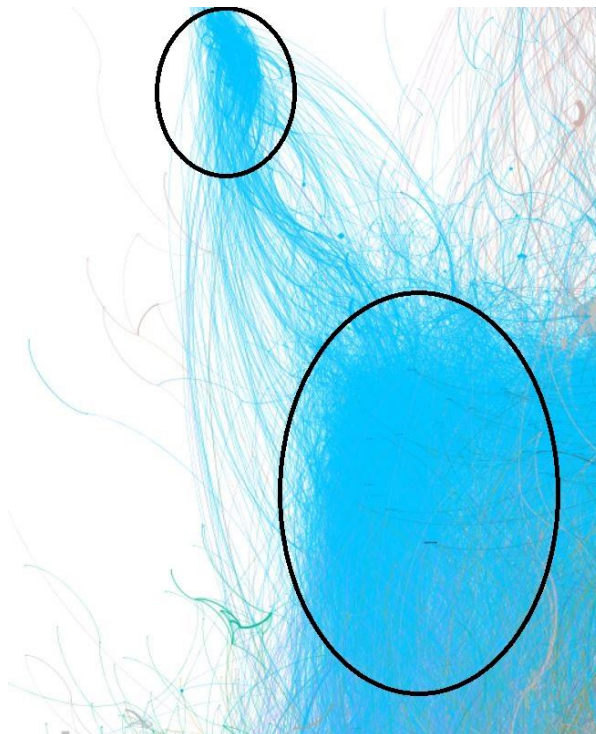
La primera comunidad que observamos es la comunidad de color morado, que es claramente la predominante (Figura 4-10).



**Figura 4-10 Parte morada del grafo**

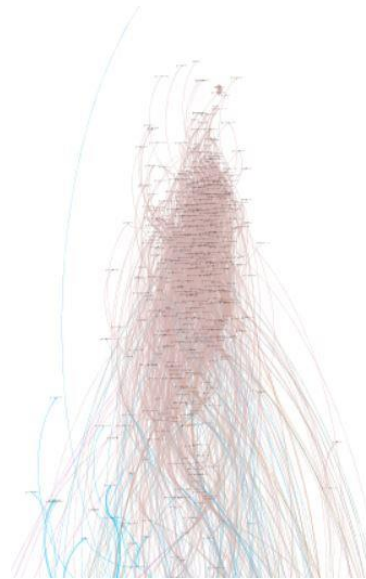
Se puede ver claramente que la influencia principal la tiene el usuario “@kyivindpendent”, es decir, los tuits publicados por este usuario han sido los más *retuiteados* y que tiene un gran alcance. Se puede apreciar también en la Figura 4-10 que además influye en otras comunidades como puede ser la azul. Podemos deducir a raíz del nombre del usuario y de la cantidad de repercusión que tienen sus tuits que este usuario se decanta por el bando ucraniano, es decir, publica tuits a favor de Ucrania.

Si seguimos analizando el grafo general, podemos observar otra gran comunidad que es la mostrada en color azul, tal y como se observa en la Figura 4-11.



**Figura 4-11 Parte azul del grafo**





**Figura 4-14 Parte rosada del grafo**

Por otra parte, mezclado dentro de las grandes comunidades, encontramos otras de color naranja o de color celeste que nos quiere dar a entender que se encuentran mezcladas y diseminadas por dentro de estas comunidades, tal y como se puede observar en la Figura 4-15.



**Figura 4-15 Comunidades celeste y naranja**

Habiendo hecho una evaluación de las principales comunidades del grafo, procederemos a buscar información sobre los usuarios influyentes en estas comunidades. Primero realizaremos en Linux listados de usuarios relevantes para cada comunidad, dejando cada uno de los listados en un archivo, es decir, en un archivo estarán los usuarios de una comunidad, en otro archivo estarán usuarios de otra comunidad y así con el resto de las comunidades. Estos usuarios han sido elegidos debido a que han sido los usuarios con más *retuits* dentro de su comunidad.

Tras la creación de los archivos con los usuarios, utilizaremos la opción de T-warder para sacar la información básica de estas cuentas como puede ser su biografía, su nombre en Twitter o incluso su localización si la tienen seleccionada en Twitter.

Analizando los resultados obtenidos podemos deducir que la comunidad naranja son usuarios de medios de comunicación tradicionales como pueden ser ABC, CNN o la BBC, tal y como observamos en la Figura 4-16.

screen_name	net	relation	followers	following	statuses	lists	since	name
Reuters	reuters	root	24612428	1204	824589	132754	20/03/2007 17:46	Reuters
nytimes	nytimes	root	52037833	872	464829	214266	02/03/2007 20:41	The New York Times
ABC	abc	root	17245278	488	378009	64938	04/04/2009 12:40	ABC News
NBCNews	nbcnews	root	8934586	1859	332203	50077	18/03/2008 23:19	NBC News
BBCWorld	bbcworld	root	35067245	18	342724	129392	01/02/2007 7:44	BBC News (World)
CNN	cnn	root	56791069	1103	367142	153212	09/02/2007 0:35	CNN

Figura 4-16 Lista de usuarios de la comunidad naranja

Siguiendo con la comunidad morada, observamos que solo uno de los usuarios es un medio de comunicación, el resto son celebridades o periodistas, pero podemos relacionarlo con el ámbito de la divulgación de información (Figura 4-17).

screen_name	net	relation	followers	following	statuses	lists	since	name
KyivIndependent	kyivindependent	root	1617860	19	1225	10400	21/11/2021 22:30	The Kyiv Independent
DmytroKuleba	dmytrokuleba	root	549550	205	6800	4473	20/11/2014 11:38	Dmytro Kuleba
Kasparov63	kasparov63	root	811051	593	35317	8220	29/12/2011 8:17	Garry Kasparov
olgatokariuk	olgatokariuk	root	248596	2023	15207	4498	17/04/2013 14:11	Olga Tokariuk
JuliaDavisNews	juliadavisnews	root	206571	1571	98192	2840	16/01/2010 1:34	Julia Davis

Figura 4-17 Lista de usuarios de la comunidad morada

Hablando de la comunidad azul, vemos que son usuarios que se encuentran a favor de Ucrania pero que no están de acuerdo con las políticas que está haciendo el presidente de Estados Unidos, Joe Biden, y son seguidores del anterior presidente Donald Trump. Podemos deducir que estos usuarios son seguidores del Partido Republicano de Estado Unidos. Vemos en la Figura 4-18 que los usuarios de esta comunidad son usuarios como el hijo de Donald Trump o anteriores consejeros del expresidente.

screen_name	net	relation	followers	following	statuses	lists	since	name
JackPosobiec	jackposobiec	root	1663581	944	122830	7426	28/05/2012 13:31	Jack Posobiec US
DonaldJTrumpJr	donaldjtrumpjr	root	7363811	1603	44588	17307	11/05/2009 21:18	Donald Trump Jr.
kylenabecker	kylenabecker	root	225213	4086	45648	1436	22/06/2011 12:06	Kyle Becker
MichaelPSenger	michaelpsender	root	110652	22229	25783	769	17/08/2013 23:24	Michael P Senger
realArmandKlein	realarmandklein	root	30204	30045	217728	32	12/01/2021 15:11	Armand Klein
theIOPod	theiopod	root	7382	6418	25201	6	17/06/2020 2:47	The Independent Opinion Podcast

Figura 4-18 Usuarios de la comunidad azul

En cambio, si evaluamos los usuarios de la comunidad verde observamos que también son usuarios que apoyan a Ucrania, pero que al contrario que los usuarios de la comunidad azul, son usuarios partidarios de las políticas del actual presidente de los Estados Unidos y rechazan las políticas del anterior presidente. Se puede deducir que son seguidores del partido demócrata de Estados Unidos. Podemos ver en la lista de la Figura 4-19 que usuarios como Joe Biden se encuentran dentro de esta lista de usuarios.

screen_name	net	relation	followers	following	statuses	lists	since	name
POTUS	potus	root	19296898	12	2602	16377	13/01/2021 0:37	President Biden
kaitlancollins	kaitlancollins	root	1229366	997	19590	7469	18/08/2010 21:21	Kaitlan Collins
OccupyDemocrats	occupydemocrats	root	288130	13412	88823	1599	13/10/2012 17:12	Occupy Democrats
joncoopertweets	joncoopertweets	root	910605	63706	316013	3036	29/03/2009 20:12	Jon Cooper
CalltoActivism	calltoactivism	root	228010	1191	9988	462	13/01/2017 19:49	CALL TO ACTIVISM
briantylercohen	briantylercohen	root	472348	995	16181	1971	05/01/2012 11:43	Brian Tyler Cohen

Figura 4-19 Lista de usuarios de la comunidad verde

Por otro, la comunidad negra se puede ver que son usuarios indios y medios de comunicación rusos que se encuentran en el otro lado de la balanza, es decir, a favor de Rusia. Usuarios como la prensa del Kremlin o usuarios indios que tuitean a favor de Rusia debido a que Rusia tradicionalmente ha proporcionado a la India recursos económicos y militares, por lo que son un aliado en el conflicto que estamos viviendo hoy en día.

### 4.1.2 API Search

Mediante la opción 3 de la herramienta T-hoarder hemos obtenido un archivo de 100MB que ha ido captando información filtrando con tuits que contengan el hashtag “#UkraineRussia” cada 4 horas desde las 11.30h (hora española) del día 25 de febrero de 2022 hasta las 08.43h del día 26 de febrero de 2022. Como en el apartado anterior, vamos a proceder a importar los datos desde el archivo de texto a un Excel en donde podremos evaluar los datos de una manera más sencilla (Figura 4-20).

id tweet	date	author	text	app	id user	followers	following	statuses	loc
1,49716E+18	25/02/2022 10:55	@12numaorg	Ekatrina Gamova: "Şimdi tüm dünya Rusya'ya karşı. Bize yaptırımlar uyguluyorlar ve kendi ülkelerin Twitter Web App		322951176	1898952	908	193546	Kad
1,4973E+18	25/02/2022 19:51	@3807Limon	RT @EnsembleMacron: Lecture du message d'@EmmanuelMacron : "Cette crise majeure aura des Twitter for iPhone		8,67852E+17	9711	661	7407	Isèr
1,4973E+18	25/02/2022 20:05	@93wibc	RT @hammerandnigel: TODAY: 3-7 on @93wibc: @guyefford is in for @nigelnow as Nigel recovers Twitter for iPhone		12673222	32189	1764	17356	Indi
1,49716E+18	25/02/2022 10:38	@a_chauprade	RT @FranceNews24: DERNIERE MINUTE - #Ukraine : Le président #Zelenskiy : "L'Ukraine est prè Twitter for iPhone		365765843	19765	1413	7644	
1,49748E+18	26/02/2022 8:10	@aachinjain	RT @RaviSinghKA: #Ukraine: Guru Ka Langar on a train. These guys were fortunate to get on this tra Twitter for iPhone		862777776	676	340	6406	Mur
1,49723E+18	25/02/2022 15:25	@AaronDAndrea21	The latest: Ukrainian military vehicles are in #Kyiv to defend it against approaching Russian troops, Twitter Web App		1478485422	946	695	11752	Tor
1,4973E+18	25/02/2022 19:44	@aaronhuertas	RT @CAugustenborg: Tensions high as @EmbassyofRussia tries to enter residence during #Ukraine Twitter Web App		19026735	13044	3818	240	Was
1,49742E+18	26/02/2022 4:03	@aartic02	Heart goes out to the Indian & other Students stuck in #Ukraine , their Families & to Every Twitter for Android		1016786527	226412	5647	319412	
1,49741E+18	26/02/2022 3:23	@abc4utah	Russia continues its relentless attack on Ukraine. In Orem, one family is staying in touch with loved Twitter Web App		15771003	70436	4118	162093	Uta
1,49716E+18	25/02/2022 10:33	@abdibaktur	Ukrayna Hava Kuvvetleri'ne ait bir MiG-29, Rusya Hava Kuvvetleri'ne ait bir Su-35 savaş uçağını d Twitter Web App		117347410	1000853	810573	6306	Kon
1,49729E+18	25/02/2022 19:31	@abdibaktur	Ukrayna'dan ayrılacak Türk vatandaşları için kara sınır kapılarındaki uygulamalar: #ukrainerussia #Uk Twitter for iPhone		117347410	1000853	810551	6332	Kon
1,49716E+18	25/02/2022 10:33	@abdibaktur	#Zelenskiy, ABD'nin kendisini Kiev'den tahliye etme teklifini reddetti. #UkraineRussia #RussiaUkr Twitter for iPhone		117347410	1000761	810500	6320	Kon
1,49716E+18	25/02/2022 10:33	@abpanandavt	জেলেনস্কির মন্তব্য, ইউক্রেনের নেতারা বলেছিলেন যে, তারা আমাদের সঙ্গে রয়েছেন। কিন্তু তারা Twitter Web App		594676291	624474	11	189371	Kolb
1,49723E+18	25/02/2022 15:24	@abpanandavt	সরকারকে উৎসাহিত করুন আপনারা, ইউক্রেনের সেনাকে বার্তা পুঁজির #UkraineRussia #RussiaUk Twitter Web App		594676291	624603	11	189395	Kolb
1,49717E+18	25/02/2022 11:09	@abpmajhatv	Russia-Ukraine Conflict : भारत सरकारचं मिशन अवरलियेट, AIR INDIA चें दोन विषय विमान तयार ठेवण Twitter Web App		52121743	1190340	387	322892	Mur
1,49716E+18	25/02/2022 10:44	@abpmajhatv	Russia Vs Ukraine : रशिया आणि युकेन युद्धात सुखवात, पाहा कोण आहे जास्त ताकददार #RussiaUkrai Twitter Web App		52121743	1190344	387	322893	Mur
1,49716E+18	25/02/2022 10:37	@abpmajhatv	Russia Ukraine Crisis : प्रवास करताना गाडीवर भातवात झडा लावा, युकेनमधील भारतीय दूतावासाकडून Twitter Web App		52121743	1190345	387	322893	Mur
1,49748E+18	26/02/2022 8:00	@abpmajhatv	युक्रेनमधून परतणाऱ्या विद्यार्थ्यांच्या स्वागतासाठी छत्रपती शिवाजी महाराज अंतर्राष्ट्रीय विमानतळ सज्ज #M Twitter Web App		52121743	1191127	387	323030	Mur
1,49717E+18	25/02/2022 11:24	@abpnadu	உ.கரைன் ராணுவம் சண்டையை நிறுத்தினால் பேச்சுவார்த்தைக்கு தயார் - Twitter Web App		1,37968E+18	13427	16	31781	Che
1,49716E+18	25/02/2022 10:37	@abpnadu	#JUSTIN   அமெரிக்கா, பிரிட்டன், ஜெர்மனி உள்ளிட்ட நாடுகள் மீது பொருளாத Twitter Web App		1,37968E+18	13428	16	31782	Che
1,49724E+18	25/02/2022 15:43	@abpnadu	Gold Price: ரஷ்யா - உ.கரைன் போரால் தங்கம் விலை இத்தனை ஆயிரம் வரை Twitter Web App		1,37968E+18	13431	16	31816	Che
1,49749E+18	26/02/2022 8:37	@abpnadu	#BREAKING   ரஷ்யாவை தடைப்படுத்தும் - உலக நாடுகளுக்கு உ.கரைன் Twitter Web App		1,37968E+18	13434	16	31861	Che
1,49716E+18	25/02/2022 10:57	@abpsanja	Ukraine-Russia War: দুবনে 'স' সস্ত্রীক সাধে 'ব'ই ডমে 3 বিংশ শতাব্দী, মাথিআ ন স্ত্রীকিআ' গুল https:// Twitter Web App		1037022253	23826	22	74818	Pun
1,49743E+18	26/02/2022 4:42	@abpsanja	Ukraine-Russia Conflict: দুবনে 'সি'ই ডমে 3 বিংশ শতাব্দী, মাথিআ ন স্ত্রীকিআ' গুল https:// Twitter Web App		1037022253	23845	22	74849	Pun
1,49735E+18	25/02/2022 23:36	@Abu_Orwa91	RT @Abood_Dwehe: في لبنان واليمن والاندلس فصار الأندلس وفتحهم في لبنان واليمن Twitter for Android		1958718764	7529	984	4326	Ista
1,49717E+18	25/02/2022 11:19	@Adnamstoon1	Friday's @EveningStandard #UkraineRussia #UkraineInvasion #UkraineUnderAttack #NATO #cartoo Twitter for iPhone		20530425	9912	955	16680	Lon
1,49717E+18	25/02/2022 11:13	@AdnanSamiLive	This sums it up! #UkraineRussia https://t.co/ExxzP59om Twitter for iPhone		236826818	682769	961	30277	Mur
1,49724E+18	25/02/2022 15:49	@Arnkronos	#UkraineRussia. Putin a esercito ucraino: "Non permettete che usino civili come scudi umani" #Ucr Twitter for iPhone		75676606	556892	143	336200	Ital

Figura 4-20 Usuarios obtenidos de la búsqueda (API search)

Una vez hayamos terminado de hacer el anterior proceso, realizaremos una gráfica con los usuarios que más tuits han publicado durante el rango de tiempo que hemos estado realizando la búsqueda. Hay que la gráfica está ordenada de mayor a menor número de tuits. Vemos en la Figura 4-21 que tiene una forma parecida a la gráfica realizada en el apartado anterior de la API Streaming.

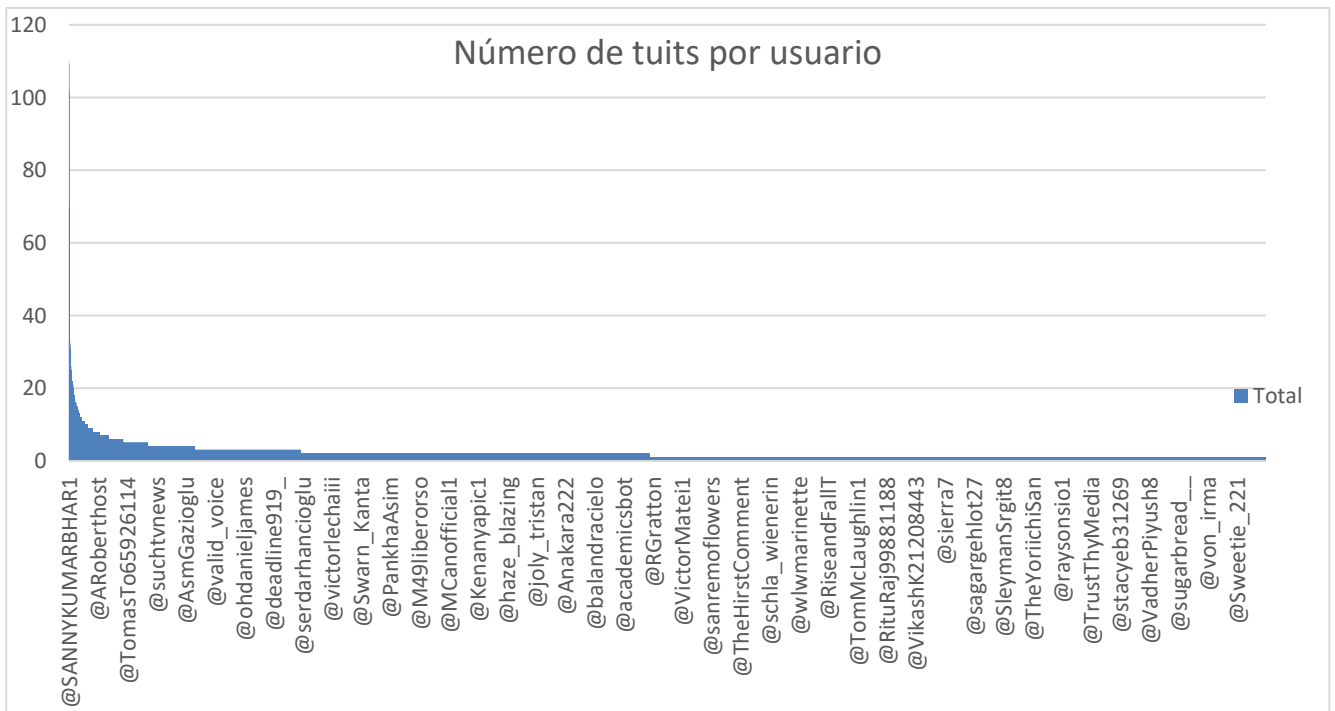


Figura 4-21 Gráfica con los usuarios que más han tuiteado (API search)



Para centrar más la gráfica, vamos a fijarnos en los 20 usuarios con más tuits publicados (Figura 4-22).

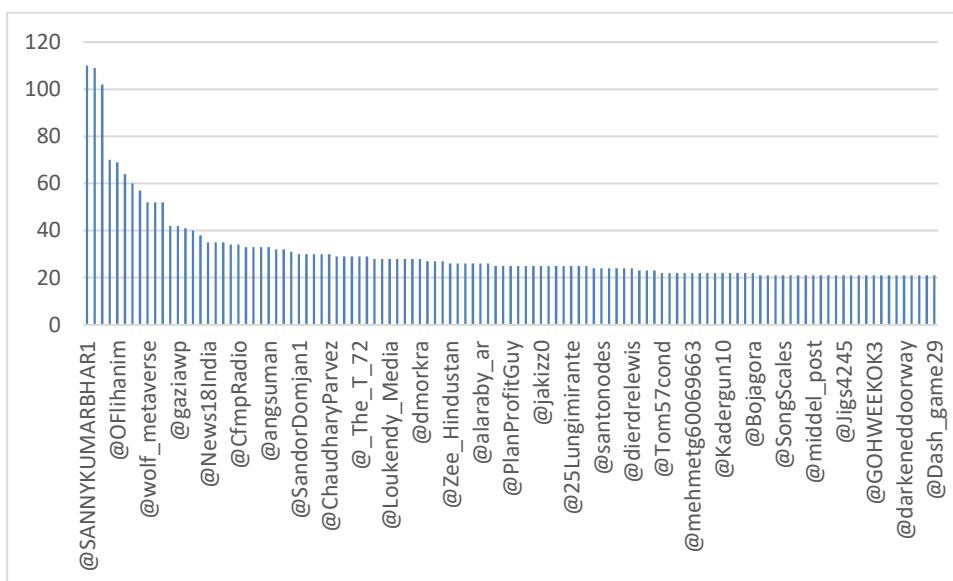


Figura 4-22 Los 20 usuarios con más tuits (API search)

Para seguir evaluando los datos obtenidos, vamos a ejecutar el código de Python que se encuentra en el Anexo I para visualizar de una manera más clara cuáles han sido los usuarios con más *retuits*, los más mencionados y los que más influencia tienen. Una vez obtenido el producto, lo copiaremos en un archivo de texto y posteriormente lo importaremos en un Excel donde quedará dispuesto como en la Figura 4-23.

Autores	Column1	Retuiteados	Columna1	Mencionados	Columna2	Influencias	Columna3	Seguidores	Columna4
@SANNYKUMARBHAR1	110	@Mediavenir	6650	@ZelenskyyUa	44	@ShehabAgency	581644625	@ABPNews	1,23005E+16
@HungoaClub07	109	@root_nomad	2985	@FoxNews	39	@Mediavenir	45929475	@ElBaradei	63547750
@Ichbinsotrendy	102	@CerfiaFR	2187	@ether_city	19	@ABPNews	2,9287E+15	@IndiaToday	59726382
@OFlihanim	69	@KyivPost	1403	@nexta_tv	18	@ThanthiTV	187823375	@ZeeNewsEnglish	5,57022E+15
@_SirDonBradman_	69	@CAugustenborg	1324	@KhalilovaGulsum	18	@CharismaNfts	1868070	@CNNnews18	467587325
@anujgsj	64	@nnnnnevae	1272	@KremlinRussia_E	18	@bstvlive	1,81283E+16	@Haberturk	45991250
@Abhishe35522389	60	@ether_city	1197	@nytimes	17	@ZeeNewsEnglish	1,79684E+16	@ugurdandarsozcu	43101040
@TimelsFinishing	57	@thestandardth	816	@Reekedson_TR	17	@YoITV	1,71283E+16	@virsanghvi	42504960
@hsyncvk23	52	@medyaadami	792	@MeeWitty	16	@thestandardth	1,54802E+16	@EconomicTimes	41362220
@DianaLanciotti	52	@AMISHDEVGAN	679	@CBSNews	15	@AlBawabaNews	1485940	@IndianExpress	41002420
@wolf_metaverse	51	@dnchoudhry	622	@Reuters	13	@abplivenews	1,39476E+16	@HergunYeniBilg	38942880
@gaziawp	42	@zoo_bear	548	@POTUS	13	@LatestinKw	1346250	@Le_Figaro	35257970
@ZVdc14aknlPkV75	42	@arkadasteroru	532	@goliammar	13	@CerfiaFR	1,33031E+16	@SirPareshRawal	33356060
@SBR2020060606	41	@techz_dave	513	@medyaadami	12	@Rudawkurdish	122599125	@eNCA	32507690
@joeyyyyyyyyyy1	40	@berojag59060636	507	@Ukraine	12	@ihacomtr	116556625	@airnewsalerts	29531440
@bulgurcrax	38	@HananyaNaftali	442	@yugan__	11	@DailyUrduPoint	1136520	@DunyaNews	29520050
@News18India	35	@RaviSinghKA	439	@ANI	11	@arynewsud	1,0313E+16	@le_Parisien	29484800
@HerkesBunu	35	@Ukrainelive5	430	@PaliwalGlobal	10	@dna	9,8449E+15	@News18India	2,80941E+15
@kelebeknebula	35	@ihansenocak	419	@cnni	10	@EconomicTimes	9,1916E+15	@dna	22643260
@ChainsChecked	34	@yenisafak	410	@Mediavenir	9	@virsanghvi	8500992	@DainikBhaskar	21753865

Figura 4-23 Tabla de datos del código de Python

Siguiendo con la dinámica anterior, vamos a proceder a realizar las gráficas de los datos obtenido del producto de Python. La primera gráfica que observamos en la Figura 4-24 es la gráfica de los usuarios con más tuits *retuiteados*.

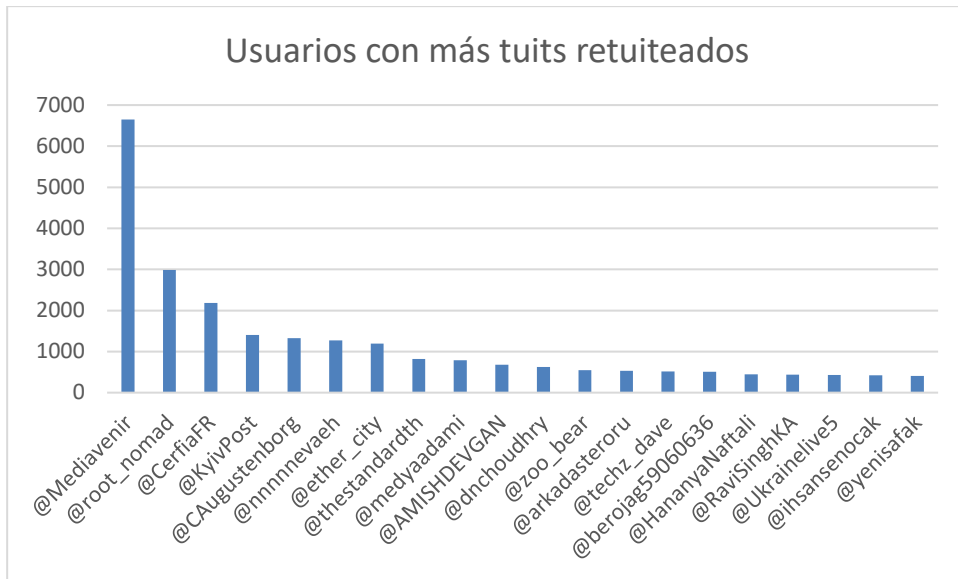


Figura 4-24 Gráfica con los usuarios que más *retuits* tienen (API search)

A continuación, se muestra la gráfica con los usuarios con más menciones (Figura 4-25).



Figura 4-25 Gráfica de los usuarios con más menciones (API search)

Finalizando con las gráficas, a continuación, en la Figura 4-26 se muestra la gráfica con los usuarios más influyentes.

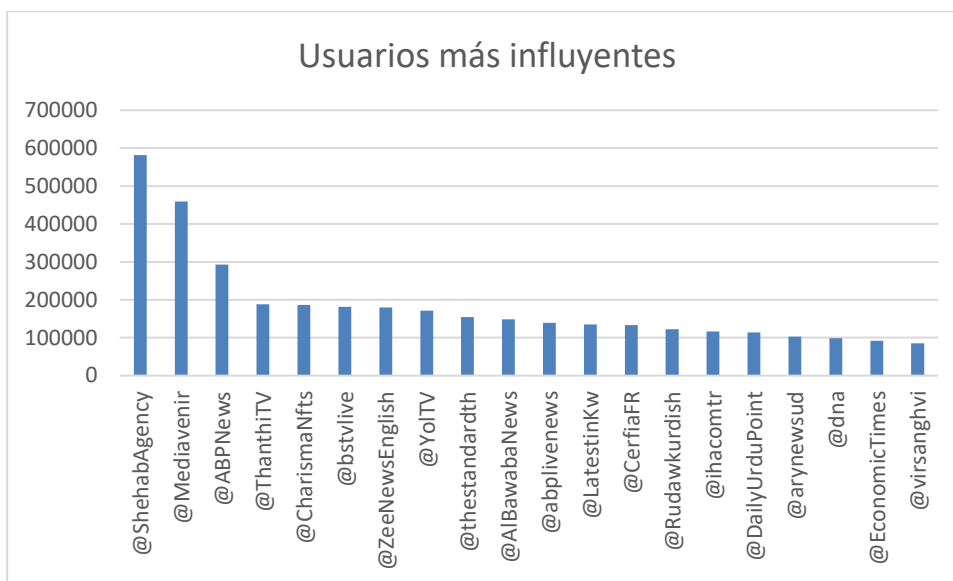


Figura 4-26 Gráfica de los usuarios más influyentes (API search)

De manera parecida, observamos que se encuentran usuarios que son muy influyentes y además tienen muchas menciones y sus tuits son muy *retuiteados*, lo que puede ser un indicativo de que pueden ser usuarios a considerar para obtener información del actual conflicto.

Siguiendo con el esquema del apartado anterior, continuaremos evaluando las comunidades de tuits mediante la opción 6 de la herramienta T-warder, que nos permite exportar el archivo de texto con los tuits a un archivo compatible con el software Gephi que nos hace ver de una manera más gráfica las comunidades existentes (Figura 4-27). Seleccionaremos la relación de *retuits* (RT) para que sea igual que el apartado anterior.

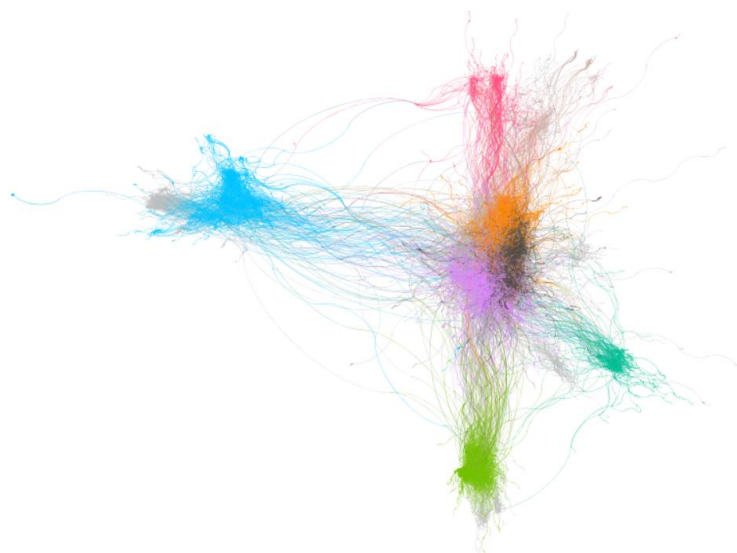


Figura 4-27 Representación del grafo de *retuits* (API search)

Se puede observar en el grafo general que existen diversas comunidades de opinión que pasaremos a continuación a evaluarlas.

Como observamos en la Figura 4-28, la primera comunidad que se observa es la comunidad verde.



**Figura 4-28 Comunidad verde del grafo**

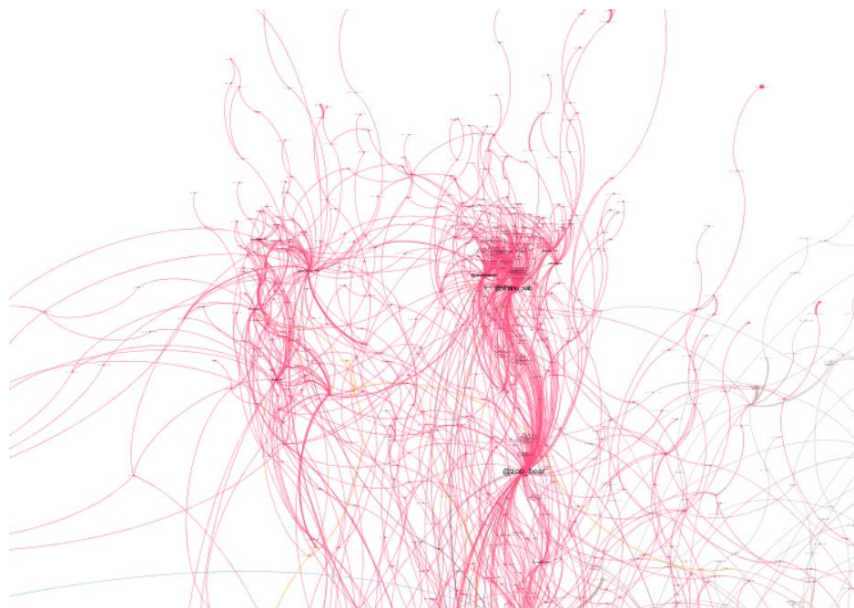
Viendo el grafo y analizando las cuentas que dentro se encuentran, podemos suponer que el software las separa debido a que los tuits no están escritos en el mismo idioma. Hay que tener en cuenta que se ha realizado una búsqueda por *hashtag*, al realizar este tipo de búsqueda nos saldrán tuits de múltiples idiomas y lo normal es que las comunidades se agrupen principalmente por el idioma. Como principales cuentas de esta comunidad nos encontramos con “@mediavenir” y “@cerfiafr”.

La siguiente comunidad que se puede observar es la comunidad azul (Figura 4-29). Esta comunidad se encuentra en otro extremo como la comunidad verde, indicativo de que los usuarios pueden escribir en otro idioma o que tener una opinión diferente de las otras comunidades. Como mayores cuentas nos podemos encontrar a “@berkayyakdenizz” o “@mehmetardic\_”.



**Figura 4-29 Comunidad azul del grafo**

Observando las comunidades que se encuentran en los exteriores del grafo, vemos la comunidad rosa (Figura 4-30). Vemos que esta comunidad es de menos repercusión, pero tiene uno de los usuarios con más *retuits* de toda la búsqueda, por lo que es una comunidad tener en cuenta. Este usuario es “@zoo\_bear”. También hay cuentas como “@shanu\_sab” o “@raimahasan” que pueden ser de interés.



**Figura 4-30 Comunidad rosa del grafo**

La última comunidad externa que observamos es la comunidad verde celeste (Figura 4-31), que se encuentra en la parte inferior derecha del grafo general. Se puede observar a simple vista que es una comunidad que tiene poca repercusión debido a la cantidad de tuits y menciones que tienen.



**Figura 4-31 Comunidad verde celeste del grafo**

Centrándonos en la estructura principal, como se puede ver en la Figura 4-32, consta de tres comunidades bien diferenciadas. La primera comunidad que se puede observar es la comunidad naranja. Observando los usuarios y el idioma, suponemos que se tratan de cuentas indias, tradicionalmente apoyando a Rusia. Por otro lado, vemos la comunidad negra, que también se tratan de usuarios indios,

pero también nos encontramos con medios de comunicación internacionales. Por último, nos encontramos con la comunidad rosa, en la que nos encontramos un amplio sector de usuarios que se manifiestan a favor de Ucrania.

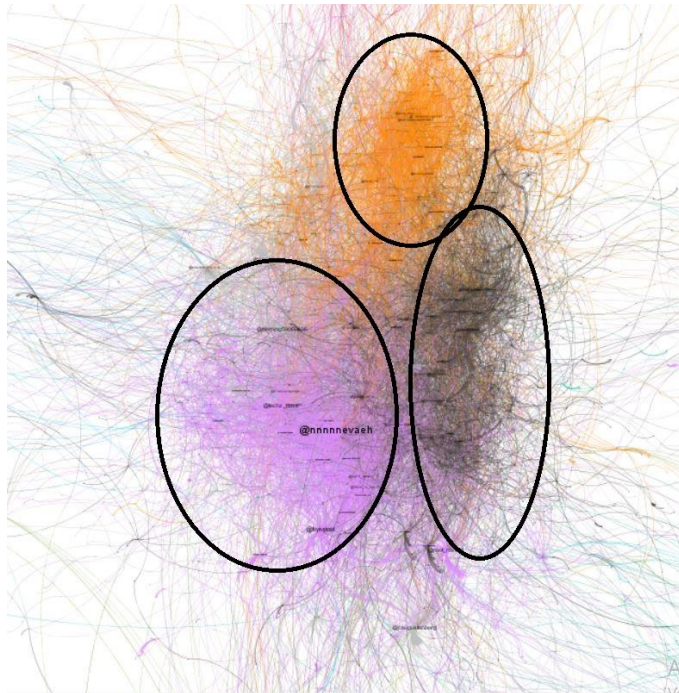


Figura 4-32 Estructura central del grafo

Viendo los grafos exteriores observamos que están vinculados al uso diferente de idiomas por parte de los usuarios. Lo siguiente es evaluar los usuarios de la estructura principal (Figura 4-32) para diferenciar las comunidades. En la Figura 4-33 se muestran los usuarios más relevantes para la comunidad naranja.

screen_name	net	relation	followers	following	statuses	lists	since	name	time zone	location
Sajidalam0033	sajidalam0033	root	28	46	128	4	21/03/2020 13:58	Sajid	None	
Supriya404	supriya404	root	3186	3555	1604	5	03/03/2021 11:10	Supriya	None	Kolkata, India
neeraj_kushwah_	neeraj_kushwah_	root	82	674	539	1	17/10/2012 7:32	Neeraj kushwah	None	New delhi
greenaugustus12	greenaugustus12	root	158562	2222	149321	40	18/07/2019 6:07	Green 🇺🇦	None	world
dkio_16	dkio_16	root	24	7	12	0	08/02/2022 9:57	Darish Kio	None	

Figura 4-33 Usuarios de la comunidad naranja

Podemos observar que se trata en algunos casos de usuarios indios, debido a la ubicación y también al nombre de usuario. Observando a los usuarios mediante sus tuits publicados y compartidos, vemos que se trata de cuentas indias que apoyan a Ucrania en el conflicto. Usualmente India es partidaria de Rusia, pero observamos que hay usuarios que apoyan al otro bando.

Siguiendo con la comunidad rosa (Figura 4-34), son usuarios de distintas partes del mundo que informan sobre los acontecimientos del conflicto.

screen_name	net	relation	followers	following	statuses	lists	since	name	time zone	location
nnnnevaeh	nnnnevaeh	root	236	103	984	6	15/01/2022 10:15	Neavah 🇺🇦	None	Bay of Plenty Region, New Zeal
techz_dave	techz_dave	root	762	31	1253	13	02/09/2017 18:23	Dave Techz	None	Canada
saint_rebel	saint_rebel	root	1691	294	16118	27	30/04/2010 18:11	Rita ArmstrongUA	None	
mars_is_calling	mars_is_calling	root	217	26	1492	1	16/07/2011 20:03	кучерява ЮА	None	Івано-Франківськ/Київ, Україна

Figura 4-34 Usuarios de la comunidad rosa

Por último, nos encontramos con la comunidad representada en negro (Figura 4-35), que son usuarios ucranianos a favor de su país, según vemos a través de sus tuits y de su biografía.

screen_name	net	relation	followers	following	statuses	lists	since	name
HaveSomeFaith10	havesomefaith10	root	118	59	93	1	05/02/2022 16:03	Have Some Faith
momentonews	momentonews	root	1329	0	297	25	04/10/2021 20:18	MOMMENTOnews
whossiry	whossiry	root	79	12	552	2	28/07/2021 22:24	whossiry

**Figura 4-35 Usuarios de la comunidad representada en negro**





## 5 CONCLUSIONES Y LÍNEAS FUTURAS

### 5.1 Conclusiones

Revisando los objetivos que hemos establecido, podemos concluir que se han cumplido todos los objetivos ya que hemos realizado una búsqueda de herramientas adecuadas para la obtención de información en Twitter. Posteriormente seleccionamos una de ellas e investigamos las funcionalidades que tiene dicha herramienta y la hemos aplicado en un caso de estudio de interés en el ámbito de la defensa: el conflicto existente a día de hoy entre Rusia y Ucrania. Tras hacer un análisis de los datos hemos llegado a una serie de usuarios que pueden ser de interés en el contexto de dicho conflicto.

La principal conclusión alcanzada mediante la realización de este trabajo es la viabilidad (incluso podríamos decir que sencillez) de aplicar herramientas públicamente disponibles a la obtención y análisis de información en Twitter. Con solo un ordenador y las herramientas necesarias instaladas, cualquier persona con unos mínimos conocimientos de informática podría realizar tareas de OSINT.

### 5.2 Líneas futuras

Uno de los inconvenientes que tienen los métodos de inteligencia de fuentes abiertas es que deben lidiar con el problema de veracidad de la información. Actualmente existen agencias que se encargan de inundar Twitter con tuits cuyo objetivo es desinformar o manipular a la opinión pública en una determinada dirección. A esto se le denomina *astroturfing* [63]. Una línea de continuación viable de este trabajo podría ser detectar cuáles son los usuarios que se encargan de publicar tuits de este tipo, mediante la evaluación de sus contenidos, de sus relaciones con otros usuarios y de sus perfiles.

Otra de las limitaciones del presente trabajo es que analizamos una pequeña fracción de la información que maneja Twitter, por lo que estamos desechando mucha información que puede ser potencialmente relevante para la obtención de inteligencia. Aumentar el volumen de datos a analizar requeriría tener acceso a dichos datos, lo que puede suponer el desembolso económico, y la necesidad de herramientas software de procesado de grandes volúmenes de datos (*big data*) distintas a las empleadas en este trabajo.

Por último, hemos visto en el capítulo de estado del arte que existen diversas herramientas de potencial interés para las actividades OSINT. Aunque nos hemos centrado en este trabajo en una de ellas, realmente algunas tienen funcionalidades complementarias. Desarrollar una plataforma OSINT que integre las funcionalidades de las distintas herramientas disponibles es otra de las posibles vías para continuar este trabajo en el futuro.



## 6 BIBLIOGRAFÍA

- [1] S. Kemp, «Data Reportal,» Hoopsuite, 27 Enero 2021. [En línea]. Available: <https://datareportal.com/reports/digital-2021-global-overview-report>. [Último acceso: 14 Enero 2022].
- [2] M. Piñero, «Estadísticas de redes sociales en España 2021,» 5 Mayo 2021. [En línea]. Available: <https://www.concepto05.com/2021/05/estadisticas-de-redes-sociales-en-espana-2021/>. [Último acceso: 27 Enero 2022].
- [3] W. «Que es Twitter y como funciona,» 1 Marzo 2018. [En línea]. Available: <https://www.webempresa.com/blog/que-es-twitter-como-funciona-2.html>. [Último acceso: 27 Enero 2022].
- [4] Hotmart, «¿Qué es Twitter, cómo funciona y para qué sirve esta red social?,» 20 Marzo 2021. [En línea]. Available: <https://blog.hotmart.com/es/que-es-twitter/>. [Último acceso: 27 Enero 2022].
- [5] France24, «El trauma en Nagorno Karabaj, la herida abierta de los armenios,» 19 Abril 2021. [En línea]. Available: <https://www.france24.com/es/programas/reporteros/20210418-reporteros-nagorno-karabaj-guerra-azerbaiyan>. [Último acceso: 27 Enero 2022].
- [6] M. Osman, «Estadística tuits al día,» 3 Enero 2021. [En línea]. Available: <https://kinsta.com/es/blog/estadisticas-twitter/>. [Último acceso: 25 Enero 2022].
- [7] US Government, «govinfo.gov,» US Government, 6 Enero 2006. [En línea]. Available: <https://www.govinfo.gov/content/pkg/USCODE-2011-title50/html/USCODE-2011-title50-chap15-subchapI-sec403-5.htm>. [Último acceso: 25 Enero 2022].
- [8] G. Romero Sánchez, «cronicaseguridad,» 1 Noviembre 2019. [En línea]. Available: <https://cronicaseguridad.com/2019/11/01/introduccion-a-las-tecnicas-y-herramientas-osint-para-la-investigacion-en-internet/>. [Último acceso: 25 Enero 2022].
- [9] V. Alfaro Parodi, «La historia de Twitter: una red que inició desde el fracaso, pero que hoy goza del éxito mundial,» 6 Septiembre 2021. [En línea]. Available: <https://www.crehana.com/es/blog/marketing-digital/historia-twitter/>. [Último acceso: 25 Enero 2022].

- [10] WikiTechNews, «El nuevo odeo llega el proximo mes,» 4 Febrero 2005. [En línea]. Available: <https://wikitechnews.net/el-nuevo-odeo-llega-el-proximo-mes-fireant-con-una-cap-a-de-pintura/>. [Último acceso: 25 Enero 2022].
- [11] R. Iglesias, «movilnvasion,» 20 Septiembre 2011. [En línea]. Available: <http://www.movilnvasion.es/el-incicio-de-twitter-antes-era-twtr/>. [Último acceso: 25 Enero 2022].
- [12] Globbos.com Periodismo Digital Interesante, «Las redes sociales más populares en España,» 15 Noviembre 2009. [En línea]. Available: <https://globbos.com/2009/11/redes-sociales-mas-populares-en-espana/>. [Último acceso: 8 Febrero 2022].
- [13] Marketingdirecto, «2011: el año del boom de las redes sociales, los móviles y los consumidores multitarea,» 2 Enero 2012. [En línea]. Available: <https://www.marketingdirecto.com/especiales/recopilatorios-2011-tendencias-2012/2011-el-ano-del-boom-de-las-redes-sociales-los-moviles-y-los-consumidores-multitarea>. [Último acceso: 15 Febrero 2022].
- [14] Twitter, «Página Oficial de la Armada Española,» 28 Enero 2022. [En línea]. Available: [https://twitter.com/Armada\\_esp?ref\\_src=twsrc](https://twitter.com/Armada_esp?ref_src=twsrc). [Último acceso: 28 Enero 2022].
- [15] Importancia, «importancia de Twitter en la actualidad,» 25 Septiembre 2016. [En línea]. Available: <https://www.importancia.org/twitter.php>. [Último acceso: 25 Enero 2022].
- [16] S. Carlos, «Twitter,» 9 Enero 2017. [En línea]. Available: <https://twitter.com/cscolari/status/818592251433648128?lang=ga>. [Último acceso: 25 Enero 2022].
- [17] Twitter, «Paso de Fragatas rusas por el Mar Mediterraneo,» 27 Enero 2022. [En línea]. Available: <https://twitter.com/dparody/status/1486660268390240259?cxt=HHwWhsCy8eXo1qEpAAA>. [Último acceso: 28 Enero 2022].
- [18] Twitter Inc, «Información sobre las API de Twitter,» 2 Marzo 2022. [En línea]. Available: <https://help.twitter.com/es/rules-and-policies/twitter-api>. [Último acceso: 2 Marzo 2022].
- [19] Y. Rodríguez, «Inteligencia de fuentes abiertas,» Grupo de estudios en Seguridad Internacional, 6 Junio 2019. [En línea]. Available: <https://www.seguridadinternacional.es/?q=es/content/inteligencia-de-fuentes-abiertas-osint-caracter%C3%ADsticas-debilidades-y-enga%C3%B1o#:~:text=Se%20tiende%20a%20considerar%20que,consulados%2C%20escu-chando%20las%20emisiones%20de>. [Último acceso: 26 Enero 2022].
- [20] M. Arrizabalaga, «Relato de uns superviviente a la invasión de bahía cochinos,» 26 Mayo 2020. [En línea]. Available: [https://www.abc.es/archivo/abci-relato-superviviente-invasion-bahia-cochinos-hombres-castro-estaban-demasiado-ocupados-matando-gente-cuba-para-buscarnos-202005260408\\_noticia.html](https://www.abc.es/archivo/abci-relato-superviviente-invasion-bahia-cochinos-hombres-castro-estaban-demasiado-ocupados-matando-gente-cuba-para-buscarnos-202005260408_noticia.html). [Último acceso: 26 Enero 2022].
- [21] F. Avila, «GOOGLE DORKS, Una manera fácil de Hacking.,» 28 Octubre 2019. [En línea]. Available: <http://www.disoftin.com/2019/10/google-dorks-una-manera-facil-de-hacking.html>. [Último acceso: 29 Enero 2022].
- [22] T. Kopca, «Google Dorks Como Encontrar Datos Interesantes y Buscar Como Un Hacker,» 16 Abril 2021. [En línea]. Available: <https://www.ma-no.org/es/seguridad/google->

- dorks-como-encontrar-datos-interesantes-y-buscar-como-un-hacker. [Último acceso: 29 Enero 2022].
- [23] Linuxito, «Google Dorks,» 18 Diciembre 2013. [En línea]. Available: <https://www.linuxito.com/seguridad/294-google-dorks>. [Último acceso: 9 Febrero 2022].
- [24] L. Fernández, «¿Conocías Shodan? Todo sobre este motor de búsqueda orientado al hacking,» 7 Marzo 2020. [En línea]. Available: <https://www.redeszone.net/tutoriales/seguridad/shodan-busqueda-hacking/>. [Último acceso: 11 Marzo 2022].
- [25] S. Clements, «¿Es Shodan el buscador web más peligroso del mundo?,» 2 Mayo 2013. [En línea]. Available: <https://www.vice.com/es/article/7bagyd/shodan-el-buscador-web-mas-peligroso-del-mundo>. [Último acceso: 29 Enero 2022].
- [26] A. Fonte, «Recopilatorio de herramientas OSINT,» 30 Abril 2021. [En línea]. Available: <https://derechodelared.com/herramientas-osint-recopilatorio/>. [Último acceso: 26 Enero 2022].
- [27] L. Fernández Manzano, «La utilización de la herramienta Maltego,» 29 Octubre 2021. [En línea]. Available: <https://atalayar.com/content/la-utilizaci%C3%B3n-de-la-herramienta-maltego>. [Último acceso: 29 Enero 2022].
- [28] M. Technologies, «Sobre Maltego,» 29 Enero 2022. [En línea]. Available: <https://www.maltego.com/about-us/>. [Último acceso: 29 Enero 2022].
- [29] Internet paso a paso, «Maltego IA ¿Qué es, para qué sirve y cuáles son todas sus funciones?,» 4 Enero 2018. [En línea]. Available: <https://internetpasoapaso.com/maltego/>. [Último acceso: 9 Febrero 2022].
- [30] A. Azamar, «Maltego,» 14 Noviembre 2017. [En línea]. Available: <https://securityassessmentsblog.wordpress.com/2017/11/14/maltego/>. [Último acceso: 29 Enero 2022].
- [31] Noticias de Seguridad Informática, «The Harvester – Busque a los empleados que trabajan en una empresa,» 8 Mayo 2015. [En línea]. Available: <https://noticiasseguridad.com/tutoriales/the-harvester-busque-a-los-empleados-que-trabajan-en-una-empresa/>. [Último acceso: 15 Febrero 2022].
- [32] E. Security, «Utilizando The Harvester para analizar el riesgo de la información pública,» 8 Abril 2015. [En línea]. Available: <https://www.welivesecurity.com/la-es/2015/04/08/the-harvester-riesgo-nformacion-publica/>. [Último acceso: 29 Enero 2022].
- [33] J. Encinar, «Recolección de información con Recon-ng,» 23 Noviembre 2018. [En línea]. Available: <https://www.boomernix.com/2018/05/recoleccion-de-informacion-con-recon-ng.html>. [Último acceso: 1 Febrero 2022].
- [34] E. Bello, «¿Qué es pentesting? Cómo detectar tus debilidades ante un ciberataque,» 9 Marzo 2021. [En línea]. Available: <https://www.iebschool.com/blog/que-es-pentesting-tecnologia/>. [Último acceso: 9 Febrero 2022].
- [35] Ciberseguridad, «Instalando y probando Recon-ng,» 2 Agosto 2020. [En línea]. Available: <https://shepherd58.wordpress.com/2020/02/08/instalando-y-probando-recon-ng/>. [Último acceso: 1 Febrero 2022].

- [36] Tannhausser, «Creepy: Herramienta de hacking para geolocalizar usuarios de Twitter y Flickr,» 15 Diciembre 2012. [En línea]. Available: <http://lamiradadelreplicante.com/2012/12/15/creepy-herramienta-de-hacking-para-geolocalizar-usuarios-de-twitter-y-flickr/>. [Último acceso: 1 Febrero 2022].
- [37] Osintux, «Creepy,» 14 Abril 2017. [En línea]. Available: <https://www.osintux.org/documentacion/creepy>. [Último acceso: 10 Febrero 2022].
- [38] I. Santiago, «Qué Es foursquare, Cómo Funciona y Para Qué Sirve,» 1 Noviembre 2021. [En línea]. Available: <https://ignaciosantiago.com/foursquare-que-es-como-funciona/>. [Último acceso: 10 Febrero 2022].
- [39] DNSDumpster, «Pagina principal DNSDumpster,» 15 Febrero 2022. [En línea]. Available: <https://dnsdumpster.com/>. [Último acceso: 15 Febrero 2022].
- [40] R. Velasco, «Obtén toda la información de un dominio en segundos con DNSdumpster,» 19 Enero 2019. [En línea]. Available: <https://www.redeszone.net/2019/01/19/dnsdumpster-informacion-dominio/>. [Último acceso: 1 Febrero 2022].
- [41] J. Jugal, «DNSDumpster – Information Gathering with DNSDumpster,» 22 Noviembre 2020. [En línea]. Available: <https://thedarksource.com/information-gathering-with-dnsdumpster/>. [Último acceso: 10 Febrero 2022].
- [42] B. N. Hacking, «DNSdumpster debe estar entre vuestras webs de confianza,» 26 Agosto 2019. [En línea]. Available: <https://www.blacknetworkhacking.com/2019/08/dnsdumpster-debe-estar-entre-vuestras.html>. [Último acceso: 1 Febrero 2022].
- [43] V. Aguilera, «Tinfoleak v.2.3 - Vicente Aguilera,» 2 Enero 2022. [En línea]. Available: <https://www.isecauditors.com/herramientas-tinfoleak>. [Último acceso: 1 Febrero 2022].
- [44] Kali Linux, «What is Kali Linux, and what is a Penetration Testing Distribution?,» 4 Febrero 2017. [En línea]. Available: <https://www.kali.org/features/>. [Último acceso: 2 Marzo 2022].
- [45] CAINE, «About the Project and Feedbacks,» 22 Noviembre 2014. [En línea]. Available: <https://www.caine-live.net/page4/page4.html>. [Último acceso: 2 Marzo 2022].
- [46] BlackArch, «About BlackArch,» 15 Junio 2015. [En línea]. Available: <https://blackarch.org/>. [Último acceso: 2 Marzo 2022].
- [47] V. Aguilera, «Tinfoleak:«The most complete open-source tool for Twitter intelligence analysis»,» 2 Enero 2022. [En línea]. Available: <http://vicenteaguileradiaz.com/tools/>. [Último acceso: 1 Febrero 2022].
- [48] N. Ferrán Ferrer, «Periodismo de datos,» 23 Enero 2018. [En línea]. Available: <http://periodisme-dades.recursos.uoc.edu/es/6-1-4-preguntas-a-resolver/>. [Último acceso: 10 Febrero 2022].
- [49] Universidad Politécnica de Valencia, «Currículum de Mariluz Congosto,» 18 Mayo 2019. [En línea]. Available: <https://www.ctranspa.webs.upv.es/webinarcovid19/mariluz-congosto-2/>. [Último acceso: 15 Febrero 2022].
- [50] Gephi, «What is Gephi,» 2 Marzo 2022. [En línea]. Available: <https://gephi.org/>. [Último acceso: 2 Marzo 2022].

- [51] JustAnotherArchivist, «Wiki sncrape,» 28 Septiembre 2018. [En línea]. Available: <https://github.com/JustAnotherArchivist/sncrape/issues?page=2&q=is%3Aissue+is%3Aopen>. [Último acceso: 10 Marzo 2022].
- [52] P. Sharma, «Scraping tweets using sncrape,» 10 Marzo 2021. [En línea]. Available: <https://www.kaggle.com/getting-started/222277>. [Último acceso: 10 Marzo 2022].
- [53] Twitter, «Developers Page,» 2 Marzo 2022. [En línea]. Available: <https://developer.twitter.com/en>. [Último acceso: 2 Marzo 2022].
- [54] Twitter, «About the Twitter API,» 16 Febrero 2022. [En línea]. Available: <https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api#v2-access-level>. [Último acceso: 16 Febrero 2022].
- [55] Oracle, «Welcome to VirtualBox,» 2 Marzo 2022. [En línea]. Available: <https://www.virtualbox.org/>. [Último acceso: 2 Marzo 2022].
- [56] P. Cordero, «Como instalar Ubuntu en VirtualBox,» 29 Septiembre 2020. [En línea]. Available: <https://osl.ugr.es/2020/09/29/como-instalar-ubuntu-en-virtual-box/>. [Último acceso: 17 Febrero 2022].
- [57] J. Gomar, «Cómo instalar Ubuntu 16.04 LTS en VirtualBox,» 22 Abril 2016. [En línea]. Available: <https://www.profesionalreview.com/2016/04/22/como-instalar-ubuntu-16-04-lts-en-virtualbox/>. [Último acceso: 17 Febrero 2022].
- [58] S. Luz, «GNU LinuxSoftware,» 11 Noviembre 2017. [En línea]. Available: <https://www.redeszone.net/2017/11/11/descarga-maquinas-virtuales-sistema-linux-virtualbox-vmware-lista-funcionar/>. [Último acceso: 17 Febrero 2022].
- [59] M. Congosto, «t-hoarder\_kit,» 10 Octubre 2017. [En línea]. Available: [https://github.com/congosto/t-hoarder\\_kit](https://github.com/congosto/t-hoarder_kit). [Último acceso: 17 Febrero 2022].
- [60] Tweepy, «What is Tweepy,» 2 Marzo 2022. [En línea]. Available: <https://www.tweepy.org/>. [Último acceso: 2 Marzo 2022].
- [61] A. Rayón, «Análisis de las redes sociales: El poder de la teoría de grafos,» 13 Diciembre 2015. [En línea]. Available: <https://blogs.deusto.es/bigdata/analisis-de-redes-sociales-el-poder-de-la-teoria-de-grafos/>. [Último acceso: 2 Marzo 2022].
- [62] D. Gayo-Avello, «De retibus socialibus et legibus momenti,» 15 Febrero 2011. [En línea]. Available: [https://www.researchgate.net/publication/231150468\\_De\\_retibus\\_socialibus\\_et\\_legibus\\_moment](https://www.researchgate.net/publication/231150468_De_retibus_socialibus_et_legibus_moment). [Último acceso: 4 Marzo 2022].
- [63] J. Quaglia, «El astroturfing, una nueva forma de mentir,» 13 Enero 2014. [En línea]. Available: [https://www.huffingtonpost.es/juan-quaglia/el-astroturfing-una-nueva-forma-de-mentir\\_b\\_4560165.html](https://www.huffingtonpost.es/juan-quaglia/el-astroturfing-una-nueva-forma-de-mentir_b_4560165.html). [Último acceso: 9 Marzo 2022].





## ANEXO I: CÓDIGO PYTHON PARA EL ANÁLISIS DE LOS DATOS OBTENIDOS CON T-HOARDER

```
import csv
import sys
import re

import codecs
import statistics

# CONFIGURAR
# Número de elementos a imprimir de las tablas
N = 20
# Posición en el CSV de los campos de interés para el análisis
POS_AUTOR = 2
POS_FOLLOWERS = 6
POS_FOLLOWING = 7
POS_REPLICA = 19
POS_RETUIT = 21
POS_VERIFICADA = 27

# Mostramos las primeras N entradas de los diccionarios
def imprime_N_diccionario(N, diccionario):
    idx = N
    for k, v in diccionario.items():
        print(k+";"+str(v))
        idx = idx-1
    if idx <= 0:
        break

# Cargamos los datos de interés en diccionarios para sacar distribuciones más adelante
autores = {}
influencias = {}
retuiteados = {}
mencionados = {}
verificados = {}
seguidores = {}
```

with open(sys.argv[1]) as csv\_file:

```

csv_reader = csv.reader((line.replace("\0", "") for line in csv_file), delimiter='\t')
try:
    for row in csv_reader:

        if len(row) < POS_AUTOR:
            continue

        # Extraemos la información de autoría
        autor = row[POS_AUTOR]
        if not autor in autores:
            autores[autor] = 1
        else:
            autores[autor] = autores[autor] + 1

        # Extraemos la información de followers y following para calcular la influencia
        # Puede ir cambiando con el tiempo: las añadimos y calcularemos la media
        if row[POS_FOLLOWERS].isnumeric() and row[POS_FOLLOWING].isnumeric():
            followers = float(row[POS_FOLLOWERS])
            following = float(row[POS_FOLLOWING])
            if following > 0.0:
                influencia = followers/following
                if not autor in influencias:
                    influencias[autor] = [influencia]
                else:
                    influencias[autor].append(influencia)
            if not autor in seguidores:
                seguidores[autor] = [followers]
            else:
                seguidores[autor].append(followers)

        # Extraemos información de retuits
        if len(row) >= POS_RETUIT:
            retuiteado = row[POS_RETUIT]
            if retuiteado != "None":
                if not retuiteado in retuiteados:
                    retuiteados[retuiteado] = 1
                else:
                    retuiteados[retuiteado] = retuiteados[retuiteado] + 1

```

```
# Extraemos la información de menciones
if len(row) >= POS_REPLICA:
    mencionado = row[POS_REPLICA]
    if mencionado != "None":
        if not mencionado in mencionados:
            mencionados[mencionado] = 1
        else:
            mencionados[mencionado] = mencionados[mencionado] + 1

# # Comprobamos si la cuenta es verificada o no
# if len(row) >= POS_VERIFICADA:
#     verificado = row[POS_VERIFICADA]
#     if verificado == "True":
#         if not autor in verificados:
#             verificados[autor] = True
except csv.Error as error:
    print("ERROR: ", error, "\n", row)

# Reemplazamos las influencias por valores medios
influencias_medias = {}
for k, v in influencias.items():
    influencias_medias[k] = statistics.mean(v)

# Reemplazamos los seguidores por valores medios
seguidores_medios = {}
for k, v in seguidores.items():
    seguidores_medios[k] = statistics.mean(v)

# Ordenamos los diccionarios por el valor
autores_sorted = dict(sorted(autores.items(), key=lambda item: item[1], reverse=True))
retuiteados_sorted = dict(sorted(retuiteados.items(), key=lambda item: item[1], reverse=True))
influencias_medias_sorted = dict(sorted(influencias_medias.items(), key=lambda item: item[1], reverse=True))
seguidores_medios_sorted = dict(sorted(seguidores_medios.items(), key=lambda item: item[1], reverse=True))
mencionados_sorted = dict(sorted(mencionados.items(), key=lambda item: item[1], reverse=True))

print("=== Autores ===")
imprime_N_diccionario(N, autores_sorted)
print("=== Retuiteados ===")
imprime_N_diccionario(N, retuiteados_sorted)
print("=== Mencionados ===")
```

```
imprime_N_diccionario(N, mencionados_sorted)
print("=== Influencias ===")
imprime_N_diccionario(N, influencias_medias_sorted)
print("=== Seguidores ===")
imprime_N_diccionario(N, seguidores_medios_sorted)
```