



Facultad de Ciencias Económicas y Empresariales, ICADE

**SEGUIMIENTO DE EVENTOS EN  
TWITTER:  
TEXT MINING Y ANÁLISIS DE REDES  
APLICADO AL LANZAMIENTO DE  
VIDEOJUEGOS.**

Trabajo de Fin de Grado de Business Analytics

Autor: Jaime Alonso López-Linares

Directora: Lucía Barcos Redín

## **Resumen**

En la actualidad, las redes sociales se han convertido en un aspecto fundamental para el éxito de multitud de negocios. Una razón de esto es que aportan oportunidades únicas de obtención de información valiosa para las empresas, especialmente cuando se trata de empresas cuyo producto tiene que ver con los usuarios típicos de estas plataformas. Un ejemplo claro de esto es la industria de los videojuegos, en la que el consumidor típico cumple las condiciones de ser de una demografía joven y con mucha experiencia usando tecnología enfocado al juego multijugador. Por esto, resulta fundamental que las empresas tengan en cuenta los datos que las redes sociales proveen para poder crear una estrategia de marketing efectiva y atractiva y actualizar su producto con éxito. Tras revisar el contexto histórico de la industria de los videojuegos para entender como ha llegado a este punto y realizando una extensa revisión de literatura pertinente para el caso se van a desarrollar una serie de técnicas que tienen que ver con el análisis de texto, análisis de redes y el análisis de sentimiento. El objetivo de este proyecto será demostrar que estos métodos de análisis de datos pueden permitir a las empresas de esta industria obtener información veraz y accionable, de forma que puedan tomar decisiones basadas en datos. Para esto se van a ejemplificar estos procesos con una serie de casos reales enfocados alrededor de un momento clave en la venta de un videojuego: su lanzamiento.

**Palabras clave:** análisis de redes, análisis de sentimiento, análisis de texto, data science, influencers, lanzamiento, marketing analítico, redes sociales, Twitter, videojuegos.

## **Abstract**

Currently, social media has become a key aspect for the success of many businesses. One reason for this is that it provides unique opportunities for obtaining valuable information for companies, especially when it comes to companies whose products are related to the typical users of these platforms. A clear example of this is the video game industry, where the typical consumer meets the conditions of being young and having a lot of experience using technology focused on multiplayer gaming. Therefore, it is essential for companies to consider the data that social media provides in order to create an effective and attractive marketing strategy and update their products successfully. After reviewing the historical context of the video game industry to understand how it has come to this point and conducting an extensive review of relevant literature for the case, a series of techniques related to text mining, network analysis, and sentiment analysis will be carried out. The objective of this project is to demonstrate how these methods can allow companies in this industry to obtain accurate and actionable information so that they can make data-driven decisions. To do this, these processes will be exemplified with a series of real cases on the Twitter social network focused on a key moment in the sale of a video game: its launch.

**Keywords:** data science, influencers, marketing analytics, network analysis, product launch, sentiment analysis, social media, text mining, Twitter, videogames.

*A mi tutora, Lucia Barcos Redín, profesora de la Universidad Pontificia de Comillas, por su consideración y confianza depositada en mi para llevar a cabo este proyecto, además del consejo recibido.*

*A mi familia y a Celia por darme fuerzas y siempre haber marcado un ejemplo a seguir.*

*Al personal médico del Hospital General Universitario Gregorio Marañón, especialmente a los doctores Dr. Ignacio Navarro Cuéllar, Dr. Santiago Ochandiano Caicoya, Dra. Marta Arregui Valles, Dra. Carolina Agra Pujol, Dr. Carlos Navarro Cuéllar y a la enfermera Loreto Fernández Bermejo. Por la constante atención recibida durante estos más de 5 meses y por su gran humanidad.*

# INDICE

<b>1. INTRODUCCIÓN.....</b>	<b>9</b>
1.1 Interés del tema y contexto histórico .....	9
1.2 Objetivos.....	11
1.3 Metodología.....	12
1.4 Estructura del proyecto .....	13
<b>2. MARCO CONCEPTUAL Y APLICACIONES ANTERIORES .....</b>	<b>14</b>
2.1 Aplicaciones previas de text mining y análisis de redes en el contexto de las redes sociales y el seguimiento de eventos .....	14
2.1.1 Análisis de los tuits de instituciones de inmigración australiana en 2013 .....	14
2.1.2 Seguimiento de la ‘Paris Fashion Week’ en 2019.....	16
2.1.3 Análisis de sentimiento de reseñas en Twitter sobre resorts en Las Vegas ...	17
2.1 Text Mining y Sentiment Analysis .....	18
2.2.1 Análisis de texto.....	18
2.2.2 Análisis de sentimiento .....	21
2.3 Análisis de Redes.....	24
<b>3. CASO DE APLICACIÓN: ANÁLISIS DE LANZAMIENTOS DE VIDEOJUEGOS EN TWITTER .....</b>	<b>26</b>
3.1 Descripción del caso de estudio.....	26
3.2 Recolección de datos, tabla de variables y explicación .....	30
3.3 Limpieza y preprocesamiento de los datos .....	32
3.3.1 Proceso de limpieza de datos .....	32
3.3.2 Otros ajustes recomendados.....	33
3.3.3 Análisis de frecuencias.....	33
3.4 Sentiment Analysis .....	40
3.4.1 Análisis de polaridades.....	40
3.4.2 Análisis de emociones.....	42
3.5 Redes de términos y temas.....	48
3.6 Identificación de influencers.....	56

<b>4. DISCUSIÓN DE LOS RESULTADOS Y APLICACIÓN PRÁCTICA .....</b>	<b>67</b>
4.1 Resultados del caso .....	67
4.2 Aplicación práctica a la industria: uso empresarial .....	69
<b>5. CONCLUSIÓN .....</b>	<b>71</b>
<b>6. BIBLIOGRAFÍA .....</b>	<b>72</b>
<b>7. ANEXOS .....</b>	<b>75</b>
7.1 Código utilizado.....	75
7.1.1 Extracción de datos de Twitter.....	75
7.1.2 Unión de estructuras de datos.....	75
7.1.3 Limpieza de datos.....	76
7.1.4 Análisis de frecuencias.....	77
7.1.5 Análisis de sentimiento .....	78
7.1.6 Elaboración de redes de términos y clustering.....	81
7.1.7 Detección de influencers .....	83
7.2 Procesos de interés para trabajos futuros con código .....	85
7.2.1 Elaboración de redes de usuarios .....	85
7.3 Polaridades medias por plataforma durante el mes de lanzamiento .....	90

## ILUSTRACIONES Y GRÁFICOS

Ilustración 1: Estructura del caso de estudio.....	13
Ilustración 2: Mapa de procedencia de seguidores, estudio de inmigración .....	15
Ilustración 3: Grafo del Caso 3 #Chanelhaute couture durante y después PFW.....	16
Ilustración 4: Métodos de Sentiment Analysis.....	22
Ilustración 5: Rueda de emociones de Robert Plutchik.....	23
Ilustración 6: Ejemplo de tipos de centralidad.....	25
Ilustración 7: Nube de palabras lanzamiento de FIFA 23 para tuits en inglés.....	34
Ilustración 8: Nube de palabras lanzamiento de FIFA 23 para tuits en español .....	34
Ilustración 9: Nube de palabras lanzamiento de Call Of Duty: MW2 para tuits en inglés.....	35
Ilustración 10: Nube de palabras lanzamiento de Call Of Duty: MW2 para tuits en español.....	35
Ilustración 11: Nube de palabras lanzamiento de God Of War: Ragnarök para tuits en inglés .....	36
Ilustración 12: Nube de palabras lanzamiento de God Of War: Ragnarök para tuits en español.....	36
Ilustración 13: Nube de palabras lanzamiento de Pokémon Escarlata & Púrpura para tuits en inglés .....	37
Ilustración 14: Nube de palabras lanzamiento de Pokémon Escarlata y Púrpura para tuits en español.....	37
Ilustración 15: Evolución de Polaridad NRC durante el lanzamiento para todos los juegos.....	41
Ilustración 16 y 17: Análisis de sentimiento en tuits sobre FIFA 23 en inglés y en español.....	42
Ilustración 18 y 19: Emociones por canal de tuits para FIFA 23 en inglés y en español .....	43
Ilustración 20 y 21: Análisis de sentimiento en tuits sobre Call Of Duty: MW2 en inglés y en español.....	44
Ilustración 22 y 23: Emociones por canal de tuits para Call Of Duty: MW2 en inglés y en español .....	44
Ilustración 24 y 25: Análisis de sentimiento en tuits sobre God Of War: Ragnarök en inglés y en español.....	45
Ilustración 26 y 27: Emociones por canal de tuits para God Of War: Ragnarök en inglés y en español .....	45
Ilustración 28 y 29: Análisis de sentimiento en tuits sobre Pokémon Escarlata y Púrpura en inglés y en español .....	46
Ilustración 30 y 31: Emociones por canal de tuits para Pokémon Escarlata y Púrpura en inglés y en español.....	46
Ilustración 32: Redes de términos y comunidades para FIFA 23 en inglés.....	49
Ilustración 33: Redes de términos y comunidades para FIFA 23 en español.....	49
Ilustración 34: Redes de términos y comunidades para Call Of Duty: Modern Warfare 2 en inglés .....	50
Ilustración 35: Redes de términos y comunidades para Call Of Duty: Modern Warfare 2 en español.....	50
Ilustración 36: Redes de términos y comunidades para God Of War: Ragnarök en inglés.....	51
Ilustración 37: Redes de términos y comunidades para God Of War: Ragnarök en español.....	51
Ilustración 38: Redes de términos y comunidades para Pokémon Escarlata & Púrpura en inglés .....	52
Ilustración 39: Redes de términos y comunidades para Pokémon Escarlata & Púrpura en español.....	52
Ilustración 40 y 41: Top 20 usuarios verificados y no verificados con más retuits para FIFA 23 en inglés.....	57
Ilustración 42: Influencers verificados para tuits de FIFA 23 en inglés .....	57
Ilustración 43: Influencers no verificados para tuits de FIFA 23 en inglés.....	58
Ilustración 44 y 45: Top 20 usuarios verificados y no verificados con más retuits para FIFA 23 en español .....	58
Ilustración 46: Influencers verificados para tuits de FIFA 23 en español.....	58
Ilustración 47: Influencers no verificados para tuits de FIFA 23 en español .....	59
Ilustración 48 y 49: Top 20 usuarios verificados y no verificados con más retuits para Call Of Duty: MW2 en inglés .....	59
Ilustración 50: Influencers verificados para tuits de Call Of Duty: MW2 en inglés .....	59
Ilustración 51: Influencers no verificados para tuits de Call Of Duty: MW2 en inglés.....	60
Ilustración 52 y 53: Top 20 usuarios verificados y no verificados con más retuits para Call Of Duty: MW2 en español .....	60

Ilustración 54: Influencers verificados para tuits de Call Of Duty: MW2 en español.....	60
Ilustración 55: Influencers no verificados para tuits de Call Of Duty: MW2 en español.....	61
Ilustración 56 y 57: Top 20 usuarios verificados y no verificados con más retuits para God Of War: Ragnarök en inglés .....	61
Ilustración 58: Influencers verificados para tuits de God Of War: Ragnarök en inglés.....	61
Ilustración 59: Influencers no verificados para tuits de God Of War: Ragnarök en inglés .....	62
Ilustración 60 y 61: Top 20 usuarios verificados y no verificados con más retuits para God Of War: Ragnarök en español.....	62
Ilustración 62: Influencers verificados para tuits de God Of War: Ragnarök en español .....	62
Ilustración 63: Influencers no verificados para tuits de God Of War: Ragnarök en español.....	63
Ilustración 64 y 65: Top 20 usuarios verificados y no verificados con más retuits para Pokémon Escarlata & Púrpura en inglés.....	63
Ilustración 66: Influencers verificados para tuits de Pokémon Escarlata & Púrpura en inglés.....	63
Ilustración 67: Influencers no verificados para tuits de Pokémon Escarlata & Púrpura en inglés .....	64
Ilustración 68 y 69: Top 20 usuarios verificados y no verificados con más retuits para Pokémon Escarlata & Púrpura en español .....	64
Ilustración 70: Influencers verificados para tuits de Pokémon Escarlata & Púrpura en español .....	64
Ilustración 71: Influencers no verificados para tuits de Pokémon Escarlata & Púrpura en español.....	65

## TABLAS

Tabla 1: Ejemplo de Term-Document Matrix.....	20
Tabla 2: Ranking de ventas de videojuegos en España enero 2019-julio 2022 en total vendido (€).....	28
Tabla 3: Lanzamientos analizados en el caso de estudio.....	29
Tabla 4: Términos utilizados para la extracción de tuits .....	30
Tabla 5: Tabla de variables relevantes resumida.....	31
Tabla 6: Categorías de emociones en sentiment analysis .....	42
Tabla 7: Resumen de influencers más relevantes para tuits en inglés .....	65
Tabla 8: Resumen de influencers más relevantes para tuits en español.....	66

# **1. INTRODUCCIÓN**

## **1.1 Interés del tema y contexto histórico**

El estado del mercado de los videojuegos se ha mantenido relativamente estable desde principios de los 2000. En la industria de las consolas hay 3 grandes dominadores, para los cuales la competencia es extrema entre dos de ellos Xbox y PlayStation. Mientras tanto, el tercero, Nintendo, siempre enfocado hacia jugadores o jóvenes o muy conocedores de la industria, se mantiene distante gracias a la enorme calidad y originalidad de su propiedad intelectual (Belli & Raventós, 2008). El poder vender consolas no parece basarse únicamente en la calidad técnica del dispositivo, sino que parece incluso más importante la calidad de los juegos producidos exclusivamente para esa consola, y complementariamente la comunicación de esta calidad.

La competencia oligopolista con respecto a las consolas hace fundamental para una editorial de videojuegos, o “publisher” como se conoce en el mercado, obtener los mejores juegos exclusivos para sus plataformas y comunicar la calidad de estos. Esto se ve amplificado por el hecho de que, durante estos años, las ventas de videojuegos se acumulan en unos pocos títulos: en 2001 los top 5 juegos más vendidos acumularon el 50% de las ventas (Clements & Ohashi, 2005).

Mientras tanto, también hay otras empresas que, al ser incapaces de competir en el mercado de las consolas, se dedican únicamente al desarrollo de software (de juegos) y a la publicidad y distribución de estos. Un ejemplo de este tipo de empresa, que tiene rol de desarrollador y de editorial o “publisher” es Activision Blizzard con títulos como “Starcraft”, “World of Warcraft” o “Diablo”, que logró gran relevancia diseñando títulos para ordenador. De esta forma multitud de fabricantes de videojuegos pretenden lograr cierta independencia del oligopolio formado en el mundo de las consolas, al centrarse en juegos para PC (Belli & Raventós, 2008).

En la industria del videojuego el rol de editorial suele recaer sobre el propio fabricante de la consola, como por ejemplo Nintendo, cuyos juegos en su actual consola, la Nintendo Switch, suelen ser exclusivos para ésta. Sin embargo, también existen otras editoriales de enorme tamaño que no fabrican consolas, como son Electronic Arts, la anteriormente mencionada Activision Blizzard o desde China, Tencent Games (Buijsman, 2022). Así se ve que por un lado se tiene la competencia de las consolas y, por otro lado, aunque extremadamente relacionado, la competencia por la venta de videojuegos. Debido al rol

fundamental que tienen los títulos exclusivos (los videojuegos que solo se venden para una plataforma) y que los 3 fabricantes principales de consolas son algunos de los “publishers” más grandes del mercado, estos dos lados de la industria, aunque diferenciados, están extremadamente conectados.

Entender estas mecánicas es necesario para poder comprender como funciona la industria de los videojuegos. Sin embargo, a estas se suman cambios en el modelo de negocio recientes, como: el dominio de los juegos multijugador, la llegada de la monetización a través de micro transacciones o que cada vez se vendan más juegos en formato digital en vez de en físico. Estos fenómenos solo se han visto en aumento debido al cambio del comportamiento de los jugadores de videojuegos durante la Pandemia del COVID-19 y los confinamientos que supuso (Ortiz, Tillerias, H., C., & Toaza, 2020). Todos estos efectos y el rol creciente de Internet en cuanto a la información que tienen los consumidores y su capacidad para comunicarse entre sí y formar comunidades de jugadores con gustos parecidos (Ortiz, Tillerias, H., C., & Toaza, 2020) han provocado un antes y un después en lo que significa ser un Publisher de video juegos. Mientras que antes su rol principal era el de distribuidor y dotante de información para los consumidores, ahora han pasado a ser anunciantes y gestores de comunidades, que aseguran a los desarrolladores que, a través de ellos, su juego será visto por su comprador potencial.

Es aquí donde está el interés fundamental de este trabajo: la industria ha evolucionado hasta un punto en el que los consumidores no solo están interesados en compartir sus experiencias en los juegos multijugador, sino que buscan crear comunidades dedicadas a dichos títulos en plataformas como foros o las redes sociales. De esta manera, para un Publisher de videojuegos ya no solo es importante la información que proveen los profesionales, sino que es vital monitorizar, entender y tener en cuenta de forma extensa aquello que su propia comunidad expresa en redes para mantener su interés en el juego y así maximizar el rendimiento y las ventas de éste. Por esto, en el proyecto se va a realizar un caso de estudio dedicado al lanzamiento de una serie de títulos, todos ellos con comunidades masivas, para asegurar una gran cantidad de datos que analizar. El objeto del estudio serán comentarios realizados en redes sociales acerca de dichos lanzamientos por usuarios acerca de los nuevos videojuegos.

La razón por la que se ha centrado el proyecto en el lanzamiento se debe a que es el momento donde se realizan la mayoría de las ventas de un juego, especialmente en el caso

de los juegos triple A, que son equivalentes a los “blockbusters” cinematográficos, y se pueden definir como juegos con un gran presupuesto y gran expectación de cara a su lanzamiento al público. Este momento es clave a la hora de determinar el éxito de dichos juegos, por lo que de cara al caso de estudio es todavía más importante ser capaz de asimilar lo que las comunidades dicen en redes sociales sobre estos títulos.

Además de esto la industria de los videojuegos es idónea para el análisis de los comentarios de los usuarios en redes sociales, puesto que tiene una característica diferencial: los desarrolladores pueden ajustar su producto después de haberlo puesto en venta al público, en respuesta a la reacción de sus consumidores. Por esto, el análisis de estas reacciones y estudio de los grupos de consumidores puede ser clave a la hora de realizar actualizaciones en sus productos. Debido a las limitaciones de extensión de este proyecto, el centro de atención caerá sobre 2 tipos de técnicas que satisfacen estas necesidades de análisis de datos: el análisis de texto o text mining, y el análisis de redes a lo que se podría sumar clustering. De esta forma se pretende extraer información como el sentimiento en redes hacia cierto título, así como de forma más avanzada elaborar una red de términos y temas de conversación.

## **1.2 Objetivos**

El objetivo general de este trabajo es demostrar la importancia de analizar el contenido textual de los comentarios de usuarios en Twitter durante el lanzamiento de videojuegos para la toma de decisiones de las empresas de la industria. El análisis consiste en aplicar técnicas de análisis de texto y análisis de redes. Para llevar a cabo este objetivo general, se abordarán los siguientes subobjetivos específicos:

- Establecer un marco conceptual, desarrollando una bibliografía extensa sobre análisis de texto y análisis de redes. Se utilizarán una serie de papers para explicar cómo las técnicas de text mining y análisis de redes se han utilizado en contextos similares al caso de estudio y cómo aportan valor.
- Analizar un número de lanzamientos de videojuegos que se puedan comparar entre ellos, demostrando la existencia de particularidades en cada título. Los criterios de selección de los juegos serán: que aparentemente tengan comunidades diferentes, que sean franquicias exitosas (que mucha gente los juegue) y que se puedan obtener datos en inglés y en español. Se busca contestar: ¿De qué se está

hablando? ¿Quiénes están hablando? ¿Qué personas son más importantes? ¿Qué temas sobresalen y cómo se agrupan? ¿En qué difieren las comunidades, (por idioma, por juego, etc.)?

- En base a los resultados obtenidos en el estudio, se darán recomendaciones para los casos estudiados y se demostrará como estas técnicas pueden ayudar a la toma de decisiones. Los resultados y conclusiones estarán centrados en los títulos observados en este proyecto, pero es relevante que las técnicas y análisis aplicados sean replicables para otros casos en esta industria.

### **1.3 Metodología**

Se analizará el transcurso del lanzamiento de un número de videojuegos mediante técnicas de text mining y el análisis de redes y se desarrollará código para obtener los datos necesarios para el análisis. Para abordar los objetivos explicados anteriormente se seguirá la siguiente metodología:

Para alcanzar primer subobjetivo se han tomado como fuentes de información las bases de datos académicas: Scopus, Web of Science y Google Scholar. Usando keywords como Text Mining, Sentiment Analysis, Social Network Analysis, videogames, events, product launch, etc. Se ha construido una bibliografía básica y desde ahí se incluirán otros artículos, papers, manuales y libros para explicar ciertos procesos y cómo aportan valor

Para conseguir el segundo subobjetivo:

1. Descripción del caso: selección de juegos según el criterio explicado en el subobjetivo correspondiente: comunidades distintas y en inglés y español.
2. Desarrollo y tratamiento del dataset:
  - Extracción de datos a través de la API de Twitter.
  - Preprocesamiento de los datos: unificación de todos los datos en dos data sets para cada videojuego observado: uno en inglés y uno español.
  - Limpieza de estos datasets para que el análisis se centre en las palabras clave y funcionen correctamente técnicas como el sentiment analysis.
3. Análisis del texto obtenido:
  - Realización de técnicas simples de text mining, como puede ser por ejemplo una nube de palabras para cada etapa del lanzamiento.

- Elaboración de análisis de contenido: sentimiento, redes de términos, e identificación de a los usuarios más significativos (“influencers”).
- Para las técnicas pertinentes se podrá observar la evolución de ciertas métricas en el análisis (analizando cambios en las fases de “antes, durante y después”). Sin embargo, debido al enorme volumen de datos (se están analizando en todos los métodos 4 lanzamientos en 2 idiomas distintos), esto no será una prioridad: se priorizará el análisis global de cada lanzamiento.

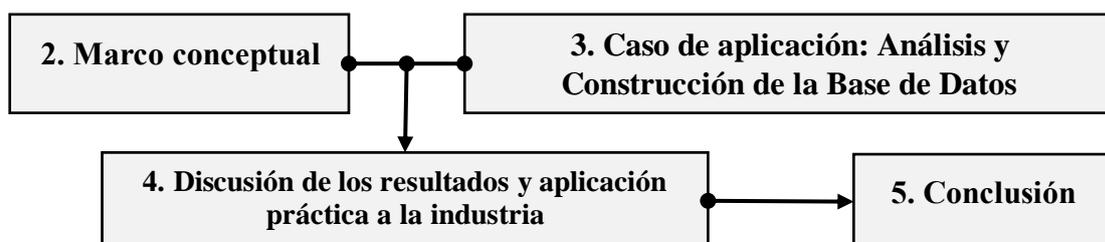
Para el tercer subobjetivo: a partir de los resultados obtenidos se extraerán recomendaciones que puedan ser relevantes para áreas como marketing, estrategia general de la compañía o para el desarrollo del propio videojuego. Se estudiarán las diferencias entre los títulos y entre la comunidad de habla hispana versus la angloparlante.

Todo el análisis se va a realizar utilizando el lenguaje de programación R, con diferentes paquetes específicos como “rtweet”, “dplyr”, “tidyr”, “igraph”, “syuzhet”, “lubridate”, “stringr” y “ggplot2”.

#### **1.4 Estructura del proyecto**

En este primer capítulo se aborda la información general del proyecto, como son los objetivos o la estructura en sí. En el segundo capítulo se desarrolla el marco conceptual, estando relacionado al primer subobjetivo. En el tercer capítulo se trata el análisis mediante las técnicas mencionadas en el segundo subobjetivo, cuyos resultados y aplicaciones prácticas se discuten en el capítulo 4, haciendo referencia al tercer subobjetivo. Finalmente, las conclusiones se discuten en el capítulo 5, comprobando el cumplimiento de cada objetivo. El esquema del caso de estudio es el siguiente:

*Ilustración 1: Estructura del caso de estudio*



A parte de esta estructura simple, el proyecto contiene **Resumen, Abstract y palabras clave** antes del índice, esta **Introducción y Bibliografía** y **Anexos** al final.

## **2. MARCO CONCEPTUAL Y APLICACIONES ANTERIORES**

### **2.1 Aplicaciones previas de text mining y análisis de redes en el contexto de las redes sociales y el seguimiento de eventos**

Este apartado del proyecto va a servir como un repaso de algunos de los usos más notables que han tenido las técnicas de text mining y el análisis de redes, en cuanto al seguimiento de eventos, o en relación con el desarrollo de métodos similares a los que se plantea hacer en este proyecto.

Los casos históricos ayudan a entender las diversas utilidades que pueden tener estos procedimientos, así como facilitan la explicación del componente teórico de este trabajo, que se va a centrar en dos pilares fundamentales del análisis de datos: el análisis de texto (en el que va a tomar un rol protagonista el análisis de sentimiento) y el análisis de redes.

En cuanto a las aplicaciones previas que se van a estudiar, el foco va a estar sobre casos realizados previamente sobre datos obtenidos de Twitter, de forma que los trabajos que se observen tengan relación con este proyecto y se puedan usar técnicas similares.

#### **2.1.1 Análisis de los tuits de instituciones de inmigración australiana en 2013**

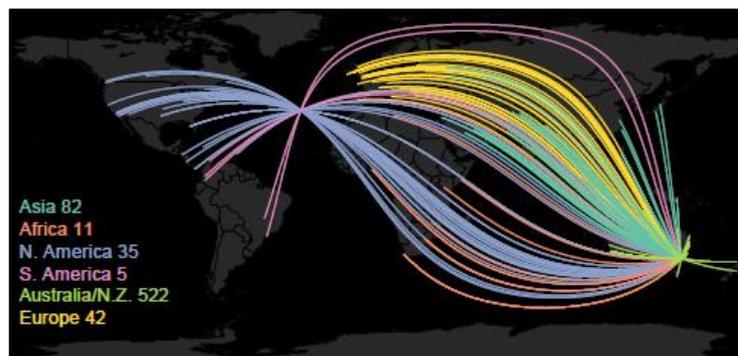
El artículo está publicado por Zhao (2013) en la onceava ‘Australian Data Mining Convention’. Desarrolla un estudio de las cuentas oficiales en Twitter del ‘Australian Department of Immigration and Citizenship’, (el departamento de inmigración australiano) con el objetivo de analizar si el desarrollo de técnicas de text mining aplicadas a su actividad en Twitter podrían mejorar el desarrollo de sus actividades.

El volumen de datos utilizados para el análisis es limitado (menos de 2000 tuits), la razón de esto es que se estaba estudiando una única cuenta, así como las respuestas e interacciones que esta generaba, además de tratarse de un tema informativo y poco comercial como es el de las migraciones. Sobre estos datos se realizó un proceso de limpieza que consistió en eliminar las partes que no ofrecían valor para el análisis de texto, como pueden ser los signos de puntuación, los links a imágenes u otras webs, o las ‘stopwords’ que en español se traduce a palabras vacías (artículos, preposiciones, adverbios...). Tras realizar este paso se reducen las palabras a lexemas o ‘stems’ (raíces) en un proceso denominado ‘stemming’.

A continuación, se realizaron una serie de técnicas como el análisis de frecuencia de términos, una red de términos y posteriormente ‘topic modelling’ (modelado por temas)

usando el modelo LDA (Latent Dirichlet Allocation) para establecer los temas. Sin embargo, el análisis no fue únicamente sobre el contenido de los tuits alrededor de esas cuentas, sino que también trató de estudiar a los seguidores de estas, lo cual se hizo mediante el desarrollo de un mapa geográfico de la procedencia de los seguidores de la cuenta (ilustración 2, en la que las líneas se refieren a la relación entre la cuenta observada y sus seguidores), así como un agrupamiento de estos en diversas categorías, según el contenido de la descripción de su perfil. También se ordenó a estos seguidores en orden de importancia, teniendo en cuenta tanto número de tuits realizado por la cuenta seguidora como su ratio entre seguidores y seguidos de la misma, el resultado de estos dos ejes se plasmó en un gráfico de dispersión donde se pretendía identificar a los seguidores más relevantes en cuanto a involucración con las autoridades australianas e influencia en la red de Twitter.

*Ilustración 2: Mapa de procedencia de seguidores, estudio de inmigración*



*Fuente:* (Zhao Y. , 2013)

Finalmente se analizaron las veces en las que la cuenta había sido retuiteada para encontrar los días en los que más se había producido esto, y con qué contenido. Los tuits más populares fueron, en primer lugar, uno sobre la baja tasa de desempleo para inmigrantes que venían a Australia, y el segundo fue sobre un anuncio para la gestión de la estancia de trabajo de estudiantes internacionales en Australia que deseen quedarse en el país trabajando después de sus estudios (Zhao Y. , 2013).

Las conclusiones generales que se extraen son que, usando técnicas de software libre para que la replicabilidad del trabajo sea alta, se ha logrado obtener ciertos conocimientos nuevos sobre el funcionamiento de estas cuentas en Twitter. En cuanto a áreas de mejora, la propuesta principal era la del estudio de hashtags en vez de cuentas, lo que aumentaría

el volumen de datos y la sofisticación del proyecto. Precisamente esto se va a realizar en el presente trabajo.

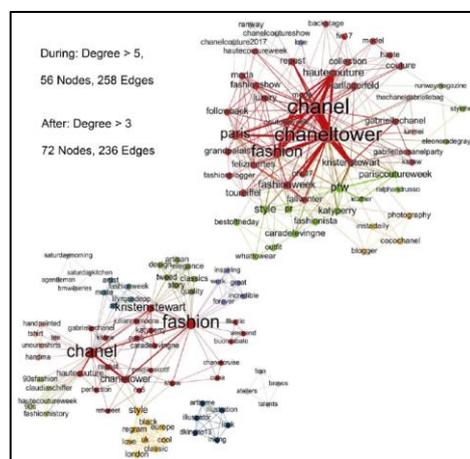
### 2.1.2 Seguimiento de la ‘Paris Fashion Week’ en 2019

Para este estudio: Zhao & Min (2019), el evento analizado es el ‘Paris Haute Couture Fashion Week’ de 2017. La razón de la elaboración del artículo está muy relacionada con lo discutido previamente en este trabajo. Estos investigadores detectan que la industria de la moda y la alta costura está cambiando. Previamente era una industria asociada tradicionalmente a élites y cuyas decisiones únicamente eran tomadas por los más altos diseñadores y ejecutivos, pero ahora ha llegado a las masas a través de las redes sociales, cuya voz tiene un valor altísimo para lograr la adaptabilidad de estas empresas de moda de lujo.

El ámbito fundamental en el que se desarrolla este proyecto es el de análisis de redes; tras identificar un alto número de hashtags relevantes, desarrolla redes de hashtags antes, durante y después PFW analizando su importancia en cuanto al valor que aportan mediante técnicas de análisis de texto. Desarrolla este estudio para:

- Caso 1: la ‘Paris Fashion Week’ en sí, estudiando los hashtags: #pfw #ParisFashionWeek.
- Caso 2: categorías de producto de la alta costura, estudiando el hashtag #HauteCouture.
- Caso 3: una marca en concreto, para estudiar como las marcas pueden usar estas técnicas en detalle, escoge Chanel bajo el hashtag: #Chanelhaute couture.

Ilustración 3: Grafo del Caso 3 #Chanelhaute couture durante y después PFW



Fuente: (Zhao & Min, 2019)

El artículo concluye enfatizando la importancia que pueden tener estas técnicas para el estudio de las relaciones que tienen los temas observados, permitiendo identificar personas de interés (como en el caso de, en la ilustración 3, la actriz Kristen Stewart para Chanel) y temas que ganan mayor atracción que otros permitiendo, en el caso de la industria textil observar el comienzo de modas de manera objetiva antes de que lleguen como tal (Zhao & Min, 2019).

### **2.1.3 Análisis de sentimiento de reseñas en Twitter sobre resorts en Las Vegas**

En este estudio de Philander y Zhong (2016), el objetivo es demostrar que mediante el análisis de sentimiento se pueden identificar una serie de medidas de bajo coste y accionables que puedan mejorar el rendimiento de los resorts observados.

Para comprobar la validez de sus resultados se incorporó la puntuación de los alojamientos en el sitio web TripAdvisor y con él se elaboró un ranking (en cuanto a los resorts en el análisis). De esta forma, tras una detallada limpieza de datos y revisión teórica del ‘Sentiment Analysis’ se compararon las puntuaciones obtenidas de sentimiento (puntuación de polaridad) en Twitter de cada resort versus las puntuaciones que obtenían en TripAdvisor. A continuación, se compararon ambos rankings, para estudiar la validez del proceso de análisis de sentimiento realizado. Por ejemplo, el hotel ‘Aria Las Vegas’ fue el mejor en cuanto a sentimiento en Twitter, pero en el ranking de la muestra de TripAdvisor era el quinto, y el número uno en cuanto a TripAdvisor era el décimo quinto en sentimiento en redes sociales (Philander & Zhong, 2016). Pese a un coeficiente de correlación positivamente moderado del 0,44, hay ciertas diferencias evidentes, las cuales se dedica a intentar explicar mediante el estudio en profundidad de los eventos positivos y negativos más notables.

Es en el estudio de estos eventos particularmente positivos y negativos donde este proyecto afirma aportar valor a la industria hostelera, puesto que esta identificación clara de momentos clave permite a los profesionales de este área actuar de manera rápida y acotada en tiempo real.

## **2.1 Text Mining y Sentiment Analysis**

### **2.2.1 Análisis de texto**

Se puede definir text mining como el proceso de descubrir o desarrollar información a partir de texto (Feldman & Dagan, 1995), en el que el análisis se realiza a través de o asistido con máquinas. Esta es la primera definición que se le da a este conjunto de técnicas, sin embargo, hoy en día se han realizado una serie de actualizaciones que especifican ciertos pasos necesarios, como el procesamiento del lenguaje natural ('natural language processing') y la necesidad de incorporar algoritmos y técnicas del ya asentado 'Knowledge Discovery in Databases' (KDD), como son ciertos procesos típicos de machine learning o estadística. De esta manera el término text mining engloba la extracción, el minado (o el análisis de patrones dentro del texto extraído) y finalmente el proceso de análisis u obtención de información (Hotho, Nürnberger, & Paaß, 2005).

Para explicar los conceptos que se van a ver a continuación es necesario entender con precisión cómo se organiza la información, para esto son clave los términos de documento, token y corpus.

- Documento: es el texto, puede ser una frase o hasta puede ser un libro, en este caso de estudio los documentos serán los contenidos de los comentarios de Twitter extraídos.
- Token: Son los términos individuales del documento, un conjunto de tokens compone un documento, una visión reduccionista sería la de afirmar que un token corresponde a una palabra.
- Corpus: es un conjunto de documentos.

A continuación, se deben conocer una serie de pasos clave para el funcionamiento de este. Como en cualquier proceso de análisis de datos, la limpieza de la base de datos (en este caso el texto), es fundamental para asegurar un análisis valioso, interpretable e informativo. En el caso del análisis de texto, este paso viene caracterizado por una serie de procesos de limpieza de texto, que pretenden asegurar que se estudian los términos de valor, y que dichos términos están en un formato que se presta más fácilmente al análisis. Las tres técnicas más importantes de preprocesamiento de texto son 'filtering', 'lemmatization' y 'stemming'.

**Filtering:** Consiste en la selección de las palabras de valor para el análisis dentro del texto (o la exclusión de aquellas palabras que no tengan valor). Esto suele consistir en la eliminación de artículos, preposiciones, etc. También se suelen excluir palabras de poco valor informativo, como en el caso de nuestro caso de estudio, se eliminarían las palabras ‘videojuego’ o ‘juego’ puesto que son palabras que, debido a su alta frecuencia, eclipsarían otros términos de valor y que no aportan valor semántico. Los métodos utilizados para realizar estos procesos están basados en la definición de entropía de Lochbaum & Streeter (1989), que dependerá de cómo de útil es una palabra a la hora de diferenciar documentos, si una palabra aparece en múltiples documentos, tendrá baja entropía porque no será diferencial. El proceso que recoge la selección de términos con mayor entropía se denomina ‘Index Term Selection’ (Hotho, Nürnberger, & Paaß, 2005). En programación a estas palabras que se desea eliminar se les suele llamar ‘stopwords’ y suelen venir en listas llamadas diccionarios que pueden ser utilizadas para eliminarlas del conjunto del texto de forma rápida y sencilla.

**Lemmatization:** La lematización conlleva encontrar el lema de una forma flexionada (Castillo Fadić, 2020), que significa obtener el lema de una palabra que puede estar en plural, femenino, etc. El formato dependerá del tipo de palabra:

- Sustantivos: singular
- Adjetivos: masculino singular
- Verbos: infinitivo

Por esto es clave conocer la categoría gramatical de la palabra (en inglés ‘Part of Speech’). Un ejemplo de esta transformación sería para el término comiste, el lema sería comer.

**Stemming:** Consiste en transformar (o en este caso más bien recortar) las palabras en sus raíces, es decir, se elimina el componente de género, número o conjugación, dejando la palabra en su forma más simple; su lexema. Esto puede desembocar en una serie de errores, como son el overstemming y el understemming. El overstemming se refiere a cuando en términos de diferente origen semántico se llega a la misma raíz, perdiendo información (se tienen las palabras ‘university’ y ‘universe y se obtiene para ambas univers) y el understemming se refiere al caso contrario, cuando dos palabras que deberían tener la misma raíz no llegan a esta (Jivani, 2011) (por ejemplo: ‘absorption’ y ‘absorb’ se reducen a absorpt y absorb).

Existen diferentes formas de organizar la información recogida en una matriz (Document Term Matrix) para un proceso de análisis de texto. En este repaso teórico se va a explicar el método Bag of Words, que es especialmente pertinente para las técnicas que se van a desarrollar en este estudio en relación con el análisis de redes y clustering dentro de estas (Bloehdorn & Hotho, 2004).

En el método Bag of Words, cada documento supone un conjunto de palabras, que a su vez serán puestas en una matriz como variables potenciales. El valor de la variable será 0 o 1 dependiendo de si aparece o no en el documento: de esta forma cada fila representará un documento mientras que cada columna representa una palabra como variable binaria. También se puede realizar un ajuste en el que las variables dejan de ser binarias y pasan a ser medidores de frecuencia; es decir, el valor de la variable es el número de veces que la palabra aparece en el documento, esto se conoce como Term Frequency Representation.

*Ejemplo:* Para las siguientes frases se desarrolla la siguiente tabla Term-Frequency

- **D1:** jazz music has a swing rhythm.
- **D2:** swing is hard to explain.
- **D3:** swing rhythm is a natural rhythm.

Como se puede observar en la tabla, como la palabra “rhythm” aparece 2 veces en D3, tiene un valor de 3 en la tabla.

*Tabla 1: Ejemplo de Term-Document Matrix*

	a	explain	hard	has	is	jazz	music	natural	rhythm	swing	to
<b>D1</b>	1	0	0	1	0	1	1	0	1	1	0
<b>D2</b>	0	1	1	0	1	0	0	0	0	1	1
<b>D3</b>	1	0	0	0	1	0	0	1	2	1	0

*Fuente:* (Provost & Fawcett, 2013)

Otra forma de representar la matriz Document-Term bajo el enfoque Bag of Words es la del Inverse Document Frequency que supone una alteración al anterior método explicado. Este consiste en, con un documento llamado ‘d’ y un término llamado ‘t’, calcular:

$$TFIDF(d, t) = TF(d, t) \times IDF(t)$$

El TF se refiere a las veces que aparece un término en un documento. El IDF o Inverse Document Frequency se incorpora para tener en cuenta cómo de común es un término del corpus, de esta forma si hay un token o término que aparece en todos o casi todos los documentos en el corpus, seguramente la información que proporciona no sea excesiva. Su fórmula es la siguiente:

$$IDF(t) = \log_2 \left( \frac{\text{Número total de documentos en el corpus}}{\text{Número de documentos que contienen el término } t} \right)$$

El uso general de estas fórmulas es el estudiar cómo de característico es un término para un documento dentro del corpus. Por ejemplo, si una palabra como “jugar” aparece mucho dentro de un documento, pero también es muy frecuente en el corpus, no tendrá un TFIDF alto, mientras que, si es poco frecuente en el corpus, sí que tendrá un valor alto de TFIDF. Un uso concreto de este cálculo es, que, con ello se pueden otorgar pesos de importancia a los términos y así estudiar como de parecida es una ‘query’ o frase buscada a los diferentes documentos del corpus. Esto se realiza mediante similitud coseno (Provost & Fawcett, 2013), y así es como funcionan algunos buscadores.

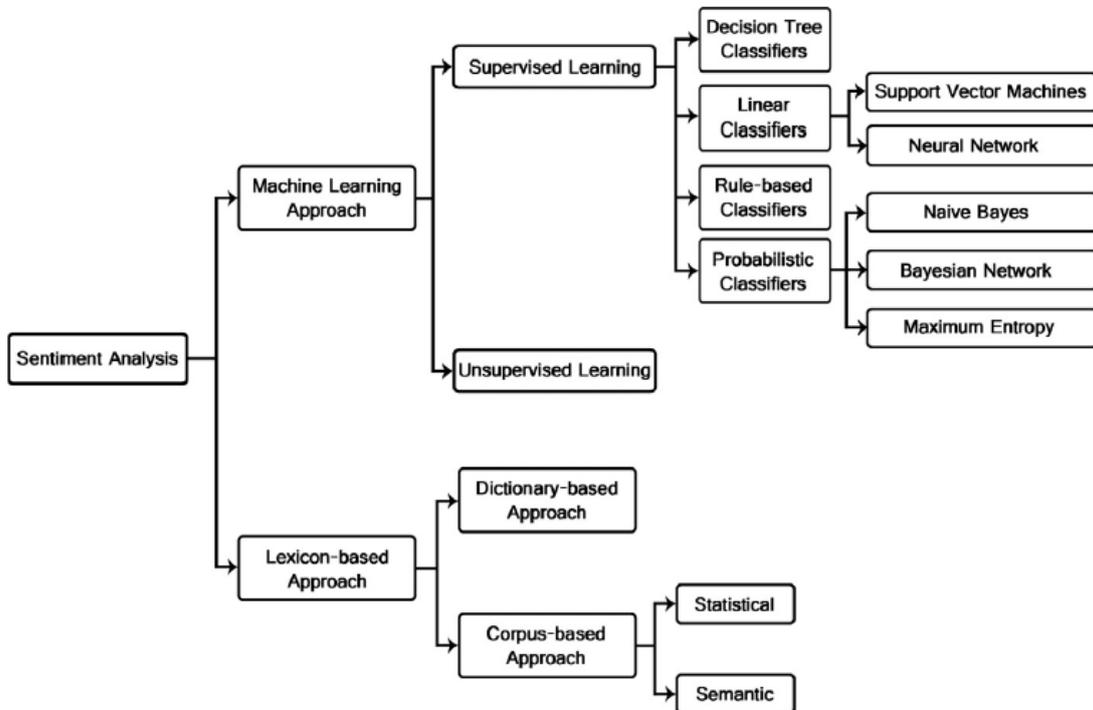
Finalmente, cabe destacar que cuando se observa el análisis de texto básico también es relevante que para realizar la limpieza de texto no solo se deben reducir y seleccionar las palabras, sino que se deben eliminar todos los elementos que puedan perjudicar al análisis del propio texto. Esto puede ser, eliminar signos de puntuación, exclamación, números o en el caso de comentarios de Twitter, hashtags, arrobas (@) o links.

### **2.2.2 Análisis de sentimiento**

Una vez se tiene el texto en un formato óptimo para analizarlo, se pueden realizar procesos más avanzados. Uno de estos procesos es el análisis de sentimiento. El análisis de sentimiento consiste en, como su nombre indica, procesar el componente emocional de un texto de forma que se obtenga información sobre el mismo. De esta forma se obtiene qué sentimiento destaca en un texto, y cuál puede ser la emoción que se expresa en este.

Esta capacidad que tienen los seres humanos por naturaleza es algo más difícil de implementar para una herramienta, y se puede hacer de diferentes maneras:

Ilustración 4: Métodos de Sentiment Analysis



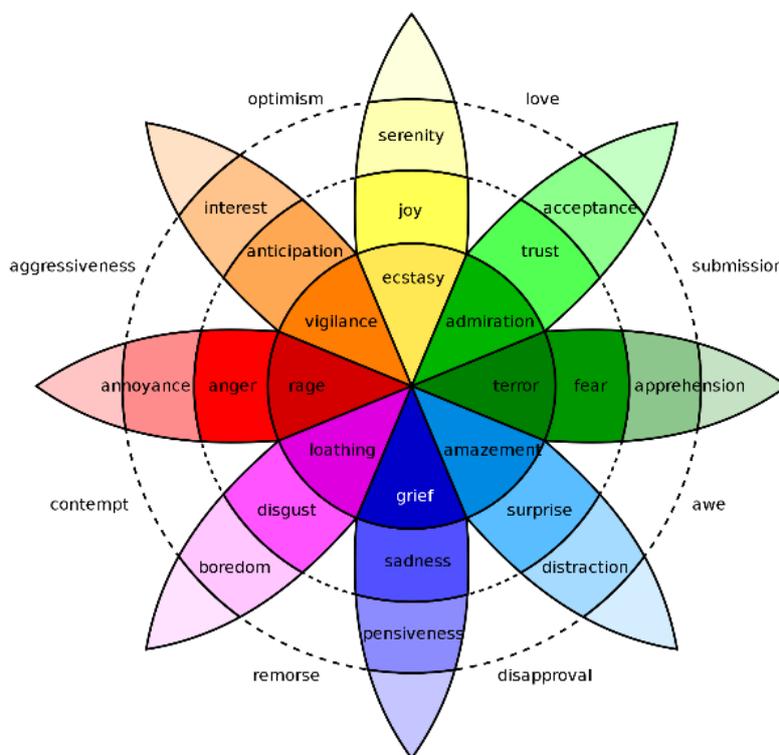
Fuente: (Medhat, Hassan, & Korashy, 2014)

En la ilustración 4 se muestra que hay dos tipos de métodos: los que utilizan machine learning para clasificar el sentimiento de los documentos, y los que utilizan lexicones. Los métodos que hacen uso de lexicones consisten en la aplicación de un diccionario donde las palabras ya tienen asignado un sentimiento (Medhat, Hassan, & Korashy, 2014) y con esto se calculan puntuaciones de sentimientos para los textos. Es el método más sencillo y con interpretación más simple. El otro tipo de sentiment analysis, que utiliza técnicas de machine learning, consiste en la utilización de algoritmos de aprendizaje supervisado, requiriendo un conjunto de documentos previamente etiquetados con el documento o no supervisado en el que el modelo aprende a clasificar las palabras y su sentimiento. Por esto si está bien realizado y cuenta con un gran volumen de datos, suele ser más preciso y sofisticado. Entendiendo estas diferencias y teniendo en cuenta las otras técnicas que se van a llevar a cabo, el análisis de sentimiento realizado en este proyecto va a ser el basado en el uso de un lexicón.

Este puede estar orientado a dos tipos de análisis

- Análisis de polaridad: si el sentimiento conjunto del documento (de todos los tokens de un documento) es positivo o negativo.
- Análisis de emociones: extrae las diferentes emociones que puede contener un documento, esto se suele realizar sobre la rueda de emociones de Plutchik (Chafale & Pimpalkar, 2014), que se puede observar en la ilustración 5.

Ilustración 5: Rueda de emociones de Robert Plutchik



Fuente: (Chafale & Pimpalkar, 2014)

Un ejemplo de cómo funciona este algoritmo para el siguiente tuit procesado:

“Llevo toda mi vida viendo pokemon a veces con más o menos entusiasmo pero siempre he estado al pendiente agradecido de ver a uno de mis héroes de la infancia cumplir sus metas gracias ash mi niño interior está feliz.”

Palabras como: “entusiasmo”, “héroes”, “gracias”, “agradecido”, “metas”, “cumplir” hacen que se le asigne un valor de la emoción joy (alegría) de 6.

### **2.3 Análisis de Redes**

La teoría del Análisis de Redes Sociales (en cuanto a la elaboración de una red de nodos) proviene de la teoría de grafos (Scott, 2011) y sigue fuertemente vinculada a esta. Por esto, la concepción clásica del análisis de redes es la de un conjunto de nodos que representan a los usuarios de la red, conectados por las interacciones que tienen entre sí.

De esta forma el análisis de redes consiste en el estudio de los nodos y las relaciones entre estos. En el caso de las redes sociales esos nodos, según el análisis pueden ser: actores, valores o temas, sentimientos, ubicaciones, etc. La segunda parte del análisis de redes es la del estudio de los vínculos que unen dichos nodos, que pueden ser asociaciones, interacciones o evaluaciones entre otras (Cioffi-Revilla, 2014). En otras palabras, el análisis de redes define una red social como un grupo de personas o entidades que están conectadas por diversas relaciones o vínculos, y que comparten intereses, valores, sentimientos, ideas, ubicaciones y atributos en común. Las relaciones pueden ser de diferentes tipos, como amistad, familia, trabajo, intereses compartidos, entre otros. Estas redes sociales se han vuelto muy populares en la era digital, ya que permiten una comunicación y conexión más fácil entre las personas (Wasserman & Faust, 1994). En relación con la industria de los videojuegos estos nodos casi siempre serán; empresas de videojuegos ('publishers' o editoriales, por ejemplo: Nintendo), jugadores cotidianos, jugadores con especial relevancia ('influencers', por ejemplo, en España: TheGrefg) y medios informativos de videojuegos o páginas de críticas (por ejemplo: IGN o Metacritic). Las relaciones en este contexto serían interacciones entre los nodos, por ejemplo, en Twitter, retweets.

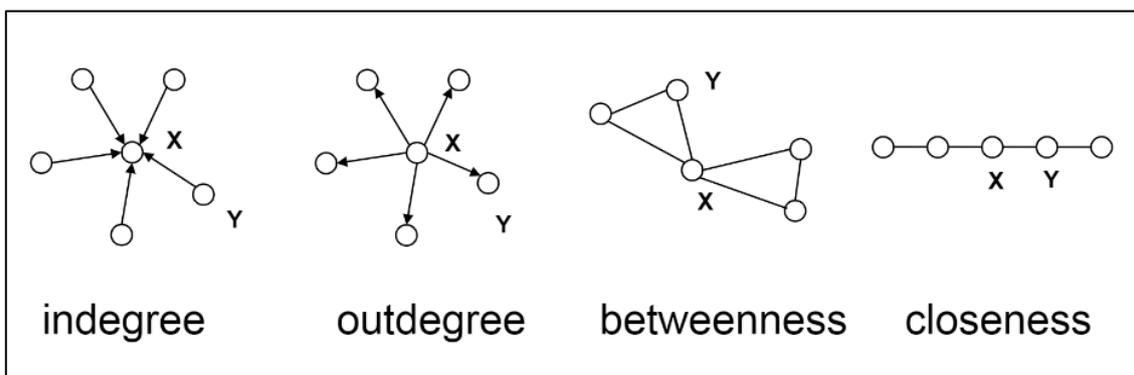
Sin embargo, también pueden elaborarse redes en las que los nodos representan ciertos términos o hashtags, y las relaciones corresponden a coocurrencias de dichos términos dentro del mismo texto, tal y como se muestra en el caso de Zhao y Min (2019).

Para estudiar la importancia de los nodos dentro de una red se utilizan una serie de métricas, pero la más básica es la centralidad del grado (degree centrality): es la medida más básica para medir centralidad y se refiere al número de conexiones que tiene un nodo. Pueden distinguirse dos medidas de grado para un nodo: el indegree y el outdegree. El indegree se refiere al número de conexiones desde otros nodos al observado y el outdegree se refiere al número de conexiones desde el nodo estudiado a otros nodos. Otra medida relevante es la de betweenness o intermediación, que mide la capacidad de un nodo o

vínculo (edge) de conectar a dos colectivos diferentes. Los nodos con mayor betweenness suelen ser aquellos en el extremo de una red de nodos que conectan con otra red. Finalmente, también cabe destacar la métrica de centralidad de closeness o cercanía de un nodo, se calcula observando cuantos de los caminos más cortos para llegar a cada nodo tiene un nodo en particular, si tiene muchos, significará que está en una posición central (Freeman, 1978). Es decir, para dos nodos la cercanía o closeness será la distancia del camino mínimo por el que estén unidos, y, así, la cercanía para un nodo significa la media de las cercanías para cada nodo con el que conecte (Needham & Hodler, 2018).

En la siguiente imagen, la ilustración 6, X siempre tiene mayor centralidad que Y según los diferentes criterios mencionados:

*Ilustración 6: Ejemplo de tipos de centralidad*



*Fuente: adaptado de (Adamic, 2014)*

### **3. CASO DE APLICACIÓN: ANÁLISIS DE LANZAMIENTOS DE VIDEOJUEGOS EN TWITTER**

#### **3.1 Descripción del caso de estudio**

Como se desarrollaba en la sección de **objetivos**, la misión fundamental de este proyecto es la de demostrar la manera en que las técnicas de text mining, análisis de contenido, sentiment analysis y análisis de redes, pueden ser aplicadas a la industria de los videojuegos, y cómo se pueden utilizar para la toma de decisiones. Por esto, a la hora de diseñar el caso de estudio, se va a hacer el seguimiento del lanzamiento de una serie de videojuegos en Twitter. Esto consistirá en el análisis del contenido de los tuits y el sentimiento o emociones asociados a los mismos, así como en realizar un estudio de la estructura de los temas de la conversación y de los usuarios más importantes. Se buscará encontrar diferencias entre los casos estudiados, para los datos en inglés y en español, y entre los propios juegos.

Como se ha visto en el repaso del marco teórico el análisis de texto aplicado a las redes sociales ofrece una perspectiva única y detallada de los consumidores, que permite a las empresas adaptar sus productos, entender temas de interés, e identificar personas clave. Existen proyectos que han estudiado eventos para industrias o áreas distintas aplicando técnicas de análisis de texto, como el artículo sobre la ‘Paris Fashion Week’, lo cual permite que el desarrollo del caso se base en técnicas ya comprobadas.

A esta base de procesos ya establecidos se suman, como se explicaba brevemente en la sección de **interés del tema**, particularidades de la industria de los videojuegos. Estas características hacen que este tema sea extremadamente adecuado para este proyecto, y son:

1. **Gran relación entre personas que usan redes sociales y personas que juegan a videojuegos:** muchos estudios afirman que cada vez hay menos diferencias entre las redes sociales y los videojuegos, con muchos juegos online convirtiéndose en cuasi redes sociales (Leonhardt & Overå, 2021). Esto encaja cuando al ver que jugar a videojuegos se ha convertido en el hobby más grande del planeta, con 3,24 mil millones de jugadores en todo el mundo, con una media de edad de 35 a nivel global y 31,4 en Europa. Poniendo el ejemplo de Reino Unido, el 25% del mercado corresponde directamente a juegos para ordenador

(Jovanovic, 2023). Por otro lado, las mujeres han tomado un rol fundamental en el desarrollo del ‘mobile gaming’ (juegos de teléfono), logrando ser un 49% del mercado, y apunta a que cada vez va a ganar más relevancia (McDonald, 2017). En resumen, el perfil que juega a videojuegos es joven, está fuertemente familiarizado con la tecnología (ya sea PC o móvil, ambos dispositivos con los que se pueden usar redes sociales), y frecuentemente tiene la necesidad de interactuar en línea con otras personas.

2. **Existencia de gran cantidad de ‘influencers’ o personas influyentes en el mercado:** esta industria se presta fácilmente a la creación de contenido, que es subido a otras redes sociales o plataformas de entretenimiento como Twitch o YouTube. Google afirma que 2 de los 5 canales más grandes de YouTube realizan predominantemente contenido sobre videojuegos (Petrova & Gross, 2017). Con audiencias masivas y con alta movilidad entre redes -muchos creadores de contenido usan Twitter como ‘tablón de anuncios’ oficial de su marca, para informar a sus seguidores de sus opiniones y actividades- estas figuras atraen gran cantidad de la participación de las comunidades. Por esto, el identificar focos de atención, como se describía hablando de la teoría de grafos, puede ser vital para entender el funcionamiento de las comunidades de forma más simplificada.
3. **Carácter actualizable de los videojuegos:** Por último, en este repaso de las razones de ser del caso de estudio, cabe destacar que, a diferencia de otras industrias como la cinematográfica, los videojuegos se pueden ajustar en tiempo real mediante actualizaciones, según lo que sea necesario para que triunfe con su comunidad. Estos cambios pueden ser complejos de ejecutar a tiempo, puesto si el desarrollador del juego no es capaz de percibir correctamente lo que quiere la comunidad, los cambios pueden ser erróneos o llegar demasiado tarde. Un ejemplo de esto muy conocido en el sector es el juego “No Man’s Sky”, que tras una extensa campaña de crowd funding (fue financiado por donaciones públicas), su anticipado lanzamiento fue motivo de innumerables críticas por parte de fans decepcionados (Lu, Li, Nummenmaa, Zhang, & Peltonen, 2020). Sin embargo, a través de un compromiso constante de actualizar el juego, sus desarrolladores lograron un éxito tardío con muchas más ventas y percepción muy positiva del público.

A continuación, se debe escoger los lanzamientos que se van a estudiar. Como se ha descrito ya con anterioridad, el análisis va a ser realizado sobre texto extraído de una red social: Twitter. Pese a que, como se ha mencionado antes, los videojuegos gozan de una presencia relevante en las redes, esta presencia se espera que, como las ventas, esté concentrada en los mayores lanzamientos. Por esto, los videojuegos que se van a observar y las comunidades que se pretende analizar van a tener que ser aquellas de juegos Triple A, es decir, los juegos más grandes en cuanto a propiedad intelectual, marca, inversión, ventas, etc. Esta condición que asegura un gran volumen de interacciones entre usuarios durante el lanzamiento limita la cantidad y diversidad de juegos observables.

Para escoger entre los juegos más populares se puede recurrir a los datos de ventas de videojuegos en España de los últimos tres años, que han sido provistos por la Asociación Española de videojuegos:

*Tabla 2: Ranking de ventas de videojuegos en España enero 2019-julio 2022 en total vendido (€)*

	<b>TÍTULO</b>
<b>1º</b>	FIFA 20
<b>2º</b>	FIFA 22
<b>3º</b>	ANIMAL CROSSING: NEW HORIZONS
<b>4º</b>	FIFA 21
<b>5º</b>	MARIO KART 8 DELUXE
<b>6º</b>	GRAND THEFT AUTO V
<b>7º</b>	RING FIT ADVENTURE
<b>8º</b>	THE LAST OF US PART II
<b>9º</b>	POKEMON SWORD
<b>10º</b>	SUPER MARIO ODYSSEY

*Fuente: AEVI 2022*

Viendo la lista de los juegos que más dinero han generado se puede extraer cuales son algunas de las franquicias más grandes en nuestro país, así como internacionalmente, puesto que el comportamiento de los consumidores españoles no será tan diferente en cuanto a los juegos más populares. El primer título que sobresale es el de la franquicia FIFA, puesto que ocupa 3 de los top 4 juegos que más han vendido durante los últimos 3 años. Otros juegos que se podrían observar son juegos de la franquicia de Mario, Pokémon, o exclusivos de PlayStation como ‘The Last Of Us Part II’. Con este conocimiento previo y estableciendo que el periodo de recolección de datos de Twitter será de julio a diciembre de 2022, se han definido los siguientes lanzamientos a analizar:

Tabla 3: Lanzamientos analizados en el caso de estudio

TÍTULO	FECHA	RAZÓN
‘FIFA 23’	27-09-2022	Es la franquicia más vendida en España de la actualidad; une comunidades normalmente no asociadas a los videojuegos, con otras muy relacionadas. Se juega en todas las plataformas
‘Call Of Duty: Modern Warfare 2’	28-10-2022	Históricamente siempre ha sido una de las franquicias más jugadas y vendidas de la industria, sin embargo, es ligeramente menos de masas que FIFA. Se juega en todas las plataformas, pero a mucho menor nivel en Nintendo
‘God Of War: Ragnarök’	09-11-2022	Se consideraba de especial relevancia tratar un juego exclusivo de PlayStation, y este es probablemente el lanzamiento más grande que ha lanzado el estudio en los últimos 5 años.
‘Pokémon Escarlata y Púrpura’	18-11-2022	La franquicia de Pokémon es la que más dinero ha generado en la historia del entretenimiento, y su base sigue siendo la de los videojuegos. Juego exclusivo de Nintendo

De esta forma se ha logrado identificar 4 lanzamientos extremadamente relevantes para la industria, que también sirven para explicar comunidades diferentes dentro de esta: el público de masas con el FIFA o el Call Of Duty (aunque algo menos), y luego 2 comunidades extremadamente comprometidas con los juegos como son la de PlayStation con ‘God Of War: Ragnarök’ y la de jugadores de Pokémon y Nintendo Switch (se descarta que sean niños puesto que están comentando en Twitter).

El estudio de estos juegos servirá para poder observar cómo difieren los temas de conversación, los sentimientos o la comunidad en general, entre títulos y plataformas que son consideradas muy distintas. De los juegos observados: las franquicias más populares son predominantemente de PlayStation y Xbox (FIFA y Call of Duty), hay un exclusivo de PlayStation (God of War) y finalmente un exclusivo de Nintendo Switch (Pokémon).

Una vez se tengan estos datos, se limpiarán y se prepararán para el análisis, realizando un análisis de frecuencias para entender mejor la información extraída. Cuando estén listos los datos se desarrollará el análisis de sentimiento, las redes de términos y la detección de influencers. El análisis se realizará en el mismo capítulo que este apartado (**capítulo 3**), pero los resultados de estos análisis, aunque mencionados brevemente para cada apartado, serán explicados en profundidad, así como sus implicaciones prácticas, en el **capítulo 4**, para que la extensión de esta parte del trabajo no sea demasiado extensa.

### 3.2 Recolección de datos, tabla de variables y explicación

Se extraerán comentarios de Twitter que contengan los hashtags más representativos para cada juego a lo largo del mes de lanzamiento. Para poder comparar el comportamiento de la comunidad española versus la internacional, los tuits se recolectarán tanto en español como en inglés. Esto se realizará a través de la Api de Twitter con su versión de hasta diciembre de 2022, para esto es necesario el uso de una cuenta de desarrollador de Twitter.

El desarrollo de este proyecto depende de la librería de R ‘rtweet’. Con esto se obtendrán los datos mediante la extracción en Twitter a lo largo del lanzamiento del videojuego observado. El código de extracción de datos obtiene tuits de hasta 9 días atrás, por lo que se harán 4 extracciones por título (en inglés y en español), dejando la fecha de lanzamiento del videojuego centrada en la extensión de los 30 días (o 4 semanas) monitorizadas.

El código utilizado extrae los tuits que contengan un texto específico, por lo que es necesario realizar una serie de pruebas previas para optimizar la recolección de datos para el estudio. Durante la semana previa al lanzamiento de cada videojuego se probaron diferentes combinaciones de términos que tenían una relación directa con el título estudiado, y se escogieron aquellos términos que conseguían extraer más datos. Esto se puede hacer de diferentes maneras, pero este proceso fue limitado por el hecho de que una cuenta de Desarrollador de Twitter básica solo permite la extracción de 18.000 tuits y con 9 días de antigüedad. Los términos que se utilizaron para extraer tuits fueron los siguientes:

*Tabla 4: Términos utilizados para la extracción de tuits*

<b>TÍTULO</b>	<b>TÉRMINOS</b>
‘FIFA 23’	“#fifa23”, “fifa23”
‘Call Of Duty: Modern Warfare 2’	“#modernwarfare2”, “call of duty”
‘God Of War: Ragnarök’	“GodOfWarRagnarok”, “god of war”
‘Pokémon Escarlata y Púrpura’	“#PokemonEscarlataPurpura” (español) “#PokemonScarletViolet” (inglés) “pokemon”

Como se puede observar en la tabla, la mejor combinación de los intentos realizados fue el de incluir el hashtag oficial (el hashtag que utilizaba el Publisher para anunciar el juego)

y título del juego o franquicia, según la ocasión. En el caso de Pokémon, como el hashtag era diferente para español e inglés, se usaron los correspondientes para cada uno.

En definitiva, tras el proceso de extracción desde la Api de Twitter se obtuvieron 8 data sets para cada videojuego (4 títulos observados), la mitad en español y la otra mitad en inglés, sumando un total de 32 archivos. Tras esto se unificó cada data set según el juego y según el idioma obteniendo 8 data sets finales, eliminando los posibles tuits duplicados que se encontraron.

Los datos generados van a tener una estructura de 91 variables (columnas) con cada fila siendo una nueva tuit, y todos los datasets completos cuentan con un volumen alrededor de 30 mil tuits recogidos. Dentro de las 91 variables son de especial interés para el análisis:

*Tabla 5: Tabla de variables relevantes resumida*

<b>VARIABLE</b>	<b>EXPLICACIÓN</b>
user_id	El identificador único del usuario que ha escrito el tuit.
status_id	El identificador único del propio tuit recolectado.
created_at	La fecha y hora en la que se publicó el tuit.
screen_name	El nombre de usuario de Twitter de la persona que publicó el tuit.
text	El texto del tuit (lo que se ha publicado), va a ser el contenido fundamental del análisis.
favorite_count	Número de veces que ha sido seleccionado como favorito el tuit observado.
retweet_count	Número de veces que ha sido retuiteado el tuit observado.
source	Plataforma desde la que se ha emitido el tuit.

La tabla anterior explica las variables más relevantes para este análisis, pero cabe destacar que el data set también cuenta con otras variables de valor informativo. Algunas de estas hacen referencia a los datos del usuario que ha tuiteado: amigos, seguidores, nombre completo, biografía, ubicación, etc. Por esto, aunque solo se hayan destacado estas variables para ilustrar cómo funciona el conjunto de datos, no significa que sean las únicas que se vayan a usar en el análisis (por ejemplo, en el caso de la detección de influencers se van a usar datos de seguidores o si el usuario está verificado).

### **3.3 Limpieza y preprocesamiento de los datos**

En este apartado se va a describir el proceso de preparación del texto para el desarrollo de técnicas más avanzadas, así como un breve análisis previo que ayude a entender mejor los datos de los que se parten, estudiando principalmente los términos y sus frecuencias.

#### **3.3.1 Proceso de limpieza de datos**

En primer lugar, para poder aplicar técnicas de procesado de texto se debe de contar con datos limpios, para ello se utilizarán funciones en R que eliminen del texto algunos elementos y se apliquen ciertos cambios, para lo que habrá que tener en cuenta cómo funciona Twitter. Estos elementos y pasos que seguir van a ser los siguientes:

1. Eliminar menciones: en el texto de los tuits puede haber referencias a otros usuarios, esto para este análisis de contenido no es de interés. Sin embargo, si se tuviese el objetivo de analizar menciones, el carácter “@” utilizado para dichas menciones sería clave.
2. Pasar el texto a minúsculas: de esta forma las palabras están todas en el mismo formato y pueden agruparse. Así los términos “FIFA”, “fifa” o “Fifa” serán términos equivalentes.
3. Eliminar URLs: de la misma forma que se han eliminado las menciones, se deben eliminar links a otras páginas puesto que son caracteres que no aplican valor información y confunden en el análisis.
4. Eliminar Ampersand: en datos descargados de Twitter es muy común encontrar el término “&”, que se da cuando se lee texto en R que incluye el signo “&”.
5. Eliminar puntuación, puesto que únicamente se está estudiando el contenido del texto en sí. Esto incluye punto, punto y coma, interrogación, exclamación, entre otros.
6. Eliminar emojis o emoticonos: al tratarse de una red social es muy común el uso de caracteres especiales, estos interrumpen directamente con las técnicas de procesamiento puesto que están en formatos que estos algoritmos no pueden tener en cuenta. Hay métodos que sí que tienen en cuenta estos términos, cuya capacidad para expresar emociones es muy valiosa para el sentiment analysis, pero para el análisis en este proyecto no se tendrán en cuenta.
7. Eliminar dígitos. Por no ser de interés para el análisis.

8. Eliminar tuits duplicados: en Twitter y en otras redes sociales es muy común encontrar cuentas que se dedican a emitir los mismos mensajes repetidamente a modo de herramienta publicitaria, especialmente en casos comerciales como puede ser el caso de lanzamientos de videojuegos.

### **3.3.2 Otros ajustes recomendados**

Una vez realizado todo este proceso también cabe destacar que existen otros ajustes que pueden ser relevantes para la elaboración de ciertas técnicas. Se debe mencionar:

1. Asegurarse que el texto esté en una codificación que pueda entender el algoritmo, por regla general para texto se suele usar el UTF-8.
2. Quitar stopwords: quitar palabras que se usan en el lenguaje cotidiano pero que no tienen valor informativo, únicamente actúan como conectores necesarios para el habla.
3. Quitar palabras que no tengan valor para el análisis: en el caso de estudio estas palabras serían los títulos de los videojuegos (en el caso de FIFA 23, la palabra más mencionada en los tuits es FIFA con mucha diferencia, esto es redundante). Esto tiene relación al concepto de filtering descrito con anterioridad.
4. Para ciertos análisis que requerirán un tiempo largo de procesamiento, es recomendable aumentar la memoria disponible para R, lo cual se puede hacer con una línea simple de código.

### **3.3.3 Análisis de frecuencias**

Este apartado pretende elaborar una primera toma de contacto con los datos de forma que se empiecen a observar ciertos comportamientos dentro de estos. De esta forma se obtendrá información pertinente tanto para las conclusiones como los análisis más complejos. Se van a observar los términos más frecuentes tanto en los datos en español como en inglés para los 4 juegos observados.

Cabe destacar que en este proceso no se ha realizado la técnica de stemming puesto que, tanto para los tuits en español como los ingleses, al ser algunos de los términos nombres propios, marcas y abreviaciones al realizarse esto se eliminaba información y los resultados obtenidos eran menos interpretables. A la hora de realizar las siguientes técnicas se volverá a valorar si es apropiado realizar stemming o lematización para el caso de estudio.









difíciles de encontrar y capturar, una comparativa en la vida real sería pensar en un animal habitual y un animal albino. La captura de estos es una actividad que se desarrolla sobre todo al final del juego, una vez transcurrida la historia, lo cual indica que la mayoría de la comunidad de Pokémon en Twitter son jugadores experimentados, que están tan pendientes de lo que van a hacer una vez completen el juego como del juego en sí.

La otra información que sobresale es la más prevalente a nivel de habla hispana: el fuerte interés de la comunidad sobre el desenlace del anime de Pokémon (la serie televisiva de la saga). Se ven reflejados tanto el nombre de su personaje principal “ash” como ciertos comentarios sobre la conclusión de esta serie con las palabras “campeón” o “mundial”. Este tipo de interacciones entre videojuegos y series puede aumentar el interés una comunidad en un lanzamiento. De entre los 4 lanzamientos observados, Pokémon fue el que más tuits obtuvo tanto en español como inglés, con los mismos o menos días de recolección.

Para Pokémon también es relevante mencionar que sale en un formato de dos juegos de propiedades prácticamente idénticas, la edición Escarlata y la Edición Púrpura. Realizando este breve análisis, se puede conocer de forma informada cual es el juego que más conversación está generando en redes, y por lo tanto predecir cual es el que va a vender más. Esto es un dato clave a la hora de realizar y actualizar inventario tanto de los propios juegos, como de ‘merchandising’, ya que una franquicia tan grande como esta cuenta con innumerables versiones de figuras, peluches, ropa... y una parte significativa estos productos suelen ser específicos de una versión del juego o de otra. En este caso se puede ver cómo tanto para los tuits en inglés como para los que están en español, el juego más popular es el Pokémon edición Púrpura, que fue la que más se vendió, dando validez a esta conclusión.

El último concepto que se debe tener en cuenta en este breve análisis exploratorio es el de la presencia de los ‘publishers’ o editoriales en la conversación en redes. En todos los casos estudiados el nombre del ‘publisher’, ya sea EA con FIFA, Activision con Call Of Duty, PlayStation con God Of War o Nintendo/ Game Freak con Pokémon, las comunidades de fans opinan abiertamente sobre las editoriales, además de compararlas con otras y con juegos similares. Por ejemplo, en el caso de Call Of Duty se pueden encontrar tuits como:

“@MFreakmennn @VGC\_News I mean when Jim Ryan is crying about call of duty to the EU why can't playstation make (or reboot) an ip, they clearly have the talent (especialmente cuando tienen a Bungie ahora) no tiene sentido para mí.” – (emitido por el usuario cbaoth\_landon).

“@Separer Call of duty out here making skin packs faster than Bungie can make sandbox changes” - (emitido por el usuario ethan\_lechich).

“@xTosai @Bungie @DestinyTheGame booted to orbit with no error codes in the middle of every other game. I can play call of duty for 8 hours with no disconnects but can't even get past the character screen on Destiny. Idk what this game has going on underneath the hood but it's terrible.”- (emitido por el usuario sammypls\_).

En estos tweets podemos ver ciertas comparaciones entre Activision y Bungie, dos desarrolladores y editoriales líderes en la industria en el género de FPS (‘first person shooter’), el primero con la saga Call Of Duty, y el segundo con las sagas Destiny y Halo. Se puede apreciar que se discuten temas estratégicos, como la batalla legal entre PlayStation (haciendo alusión a su CEO Jim Ryan) y Xbox por la adquisición del segundo de Activision, temas comerciales, como la venta de artículos cosméticos dentro de los juegos y la rapidez de un desarrollador frente a otro (Activision se percibe más ágil que Bungie) o temas de mantenimiento donde también Activision es mejor percibido.

Para cualquier empresa, y en específico en el mundo de los videojuegos, es fundamental que las empresas puedan entender a sus consumidores y su comportamiento para poder ofrecer un producto de calidad que cumpla con sus expectativas y necesidades. Por eso, contar con hallazgos rápidos y sencillos sobre la forma de pensar de los consumidores, como los que se han visto, es esencial para tener una visión amplia del mercado desde diferentes niveles. Por ejemplo, una empresa que se dedica a la venta de juegos no solo debe preocuparse por la venta de la plataforma o del juego en sí, sino también por aspectos estratégicos como el desarrollo de nuevas versiones, el lanzamiento de expansiones o la fidelización de los clientes. Además, es importante tener en cuenta aspectos comerciales, como la competencia en el mercado y los precios de venta, y de mantenimiento, como el soporte técnico y las actualizaciones necesarias para mantener el juego funcionando de manera óptima.

### **3.4 Sentiment Analysis**

Para este apartado del análisis se va a utilizar el diccionario de emociones NRC tanto en español como en inglés y se van a comparar polaridades NRC con polaridades Bing. El lexicón NRC, el principal que se va a utilizar clasifica las palabras en 8 emociones diferentes; anticipación, disgusto (asco), confianza, alegría, enfado, miedo, tristeza, y sorpresa, y 2 polaridades, positivo y negativo.

El funcionamiento de este algoritmo es el siguiente: las palabras contenidas en el diccionario están asociadas a un sentimiento, que ha sido identificada mediante crowdsourcing (Mohammad & Turney, 2013), por lo que, si se encuentra esa palabra en el texto analizado, se identifica ese sentimiento.

Este lexicón permite que con el algoritmo utilizado se obtienen datos tanto de polaridades: es decir, si un tuit en su conjunto se considera positivo o negativo, como de emociones: identifica la presencia de emociones concretas como la alegría o la anticipación. Por esto se va a dividir esta fase del análisis en estas dos etapas.

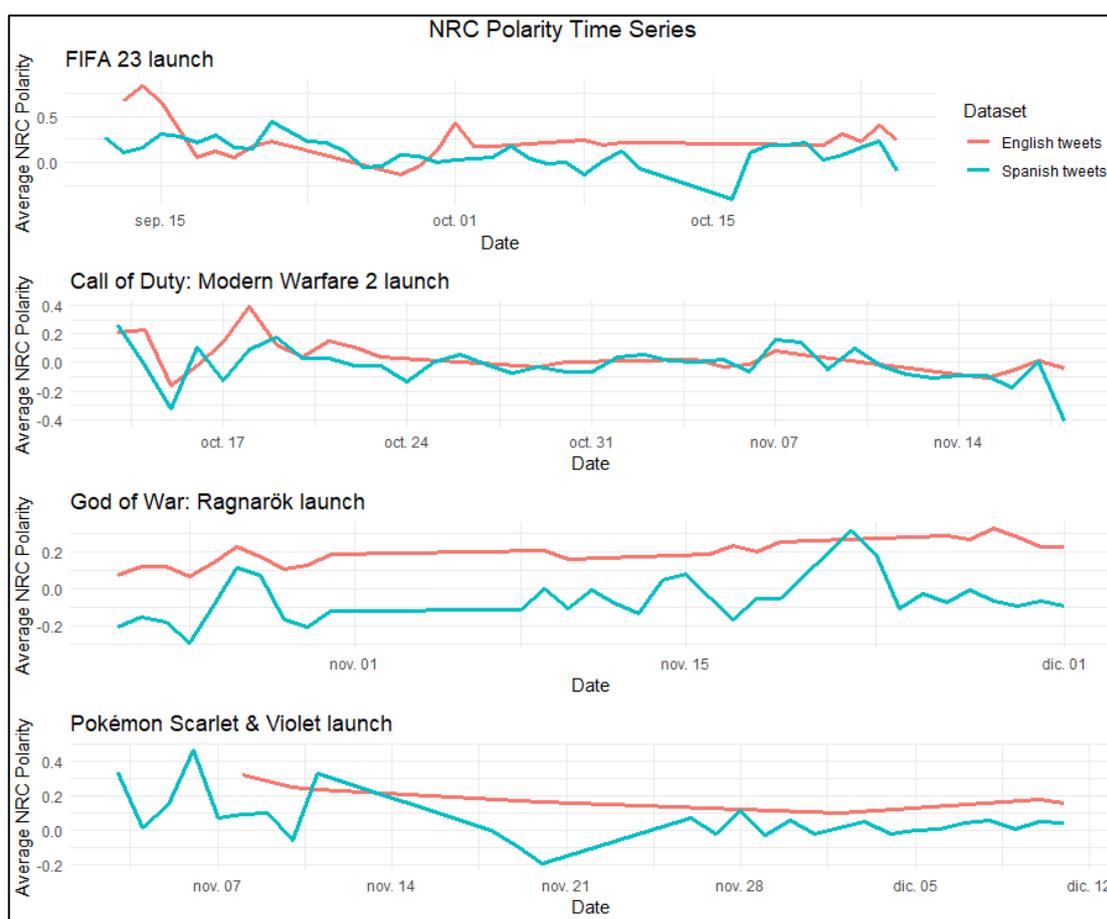
#### **3.4.1 Análisis de polaridades**

Se va a estudiar la polaridad NRC de los tuits a lo largo del mes de lanzamiento. Esta se calcula como la diferencia entre las emociones “positive” y negative” que detecta este algoritmo (cuanto mayor, más positivo el texto analizado).

De esta forma se puede observar cómo ha cambiado esta métrica a lo largo del mes, cumpliendo el objetivo de medir cómo ha evolucionado la opinión de la comunidad desde antes de que saliese a la venta el título hasta unas semanas después de esto. Debido al gran volumen de tuits, para todos los casos se tiene más de un tuit por fecha, por lo que se ha calculado una cifra de polaridad media por cada fecha en la que se recogieron datos.

Este tipo de análisis son especialmente útiles a la hora de realizar actualizaciones en el lanzamiento, para mantener que la conversación de la comunidad sea positiva. El conocer en qué momento sube o baja la opinión de la comunidad de un juego es fundamental para entender su estado y los hechos que han causado estos cambios. En los siguientes gráficos se representan estas polaridades medias por día tanto en inglés como en español para todos los juegos del caso de estudio:

Ilustración 15: Evolución de Polaridad NRC durante el lanzamiento para todos los juegos



En la ilustración 15 cabe destacar que para todos los casos la polaridad media en inglés suele ser más positiva que en español, aunque estas actúan de forma muy parecida (para cada videojuego). Esto es una demostración rápida de que la comunidad angloparlante es más positiva que la española. Además de esto para ambas series en todos los juegos se pueden apreciar cambios, con máximos y mínimos marcados, lo que puede permitir a los publishers investigar lo que pudo ocurrir en esas fechas

Es muy notable que se puede apreciar como en la serie de God of War la positividad sube según transcurre el mes. Por otro lado, entre el resto de los lanzamientos se observa un comportamiento similar, teniendo un máximo de positividad antes del lanzamiento y viendo una bajada sucesiva. Esto se puede explicar teniendo en cuenta que, tras llegar el juego al público, este puede decepcionar o superar las expectativas, en este análisis parece que el resto de los juegos no fueron tan bien recibidos como God of War: Ragnarök. También entrará el factor de que va bajando la expectativa o “hype” de las comunidades conforme a obtengan el juego.

### 3.4.2 Análisis de emociones

La siguiente fase de este análisis consiste en diferenciar las emociones que se pueden identificar dentro de cada tuit. Estas emociones corresponden a las descritas en el marco teórico y son: alegría, anticipación, miedo, asco, enfado, confianza, tristeza y sorpresa. Para la descripción de los gráficos van a ser categorizadas en positivas y negativas:

Tabla 6: Categorías de emociones en sentiment analysis

CATEGORÍAS	EMOCIONES
Positivas	Alegría (joy), anticipación (anticipation), confianza (trust), sorpresa (surprise).
Negativas	Miedo (fear), asco (disgust), tristeza (sadness), enfado (anger).

Por esto, para cada lanzamiento se va a estudiar tanto de forma global en todos los tuits, como por plataforma desde donde se ha emitido el tuit, los sentimientos más y menos importantes. En cuanto a las plataformas observadas, estas van a ser ‘PlayStation Network’, ‘Xbox One Social’, ‘Nintendo Switch Share’ o ‘Twitter Web App’ atendiendo a las 4 plataformas principales en la industria de los videojuegos; PlayStation, Xbox, Nintendo y PC o juego en ordenador. Por cada juego solo se incluirán las plataformas donde se puedan jugar. ‘Twitter Web App’ también sirve como categoría promedio, puesto que el escribir tuits desde el ordenador no significa que se juegue también en ordenador, pero entre las categorías posibles era la más aproximada a esto.

Para el caso de los videojuegos ‘Call Of Duty: Modern Warfare 2’ y ‘God Of War: Ragnarök’ se han eliminado palabras en relación con su título del corpus, como pueden ser “war” o “warfare” puesto que están identificadas en el lexicón NRC con fuerte significado emocional y por lo tanto alteraban los resultados del análisis.

#### Para el lanzamiento de FIFA 23:

Ilustración 16 y 17: Análisis de sentimiento en tuits sobre FIFA 23 en inglés y en español

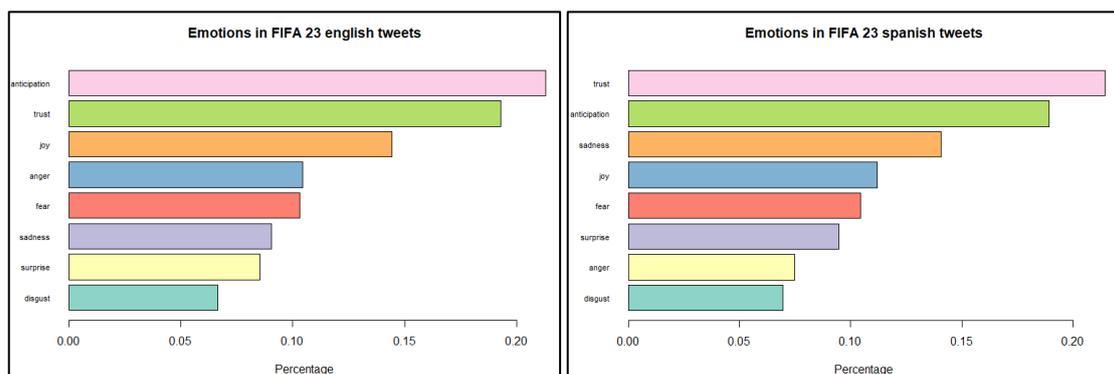
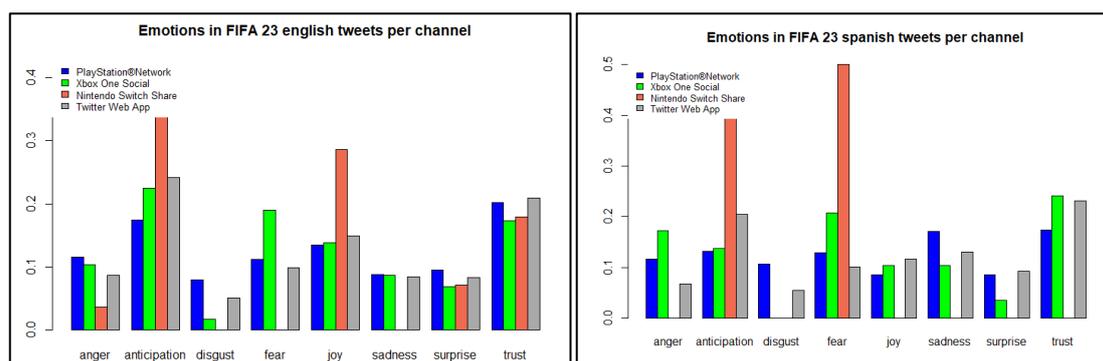


Ilustración 18 y 19: Emociones por canal de tuits para FIFA 23 en inglés y en español



En las ilustraciones 16 y 17 se aprecia que a nivel español e inglés en el caso del lanzamiento se han recibido resultados similares, con las emociones de anticipación y confianza formando el top 2 en ambos casos. Sin embargo, con el tercer sentimiento más identificado la situación cambia notablemente, ya que mientras que, para los tuits ingleses es alegría, para los españoles es tristeza. Si se extraen algunos tuits con el valor de tristeza alto (la media es de 0,2 y el máximo es de 12, por lo que se cogen tuits de tristeza 5):

- “pregunta sería el fifa el peor que hayan hecho el concierto de errores que se pueden cometer es terrible no dan pases de primera para todo se traba la pelota espantoso ojalá mejoren la experiencia porque sin duda es el peor”
- “la peor basura de la historia neta no compren fifa es un juego malísimo todo lo malo del en este fifa se multiplica por” \*

Observando el contenido de los textos en cuestión se puede ver como en ciertos casos sí que se produce cierta crítica constructiva, como en el primer tuit en el que se habla de los problemas de rendimiento que tiene el juego que “se traba” y de jugabilidad, como el problema que parece que tienen los jugadores al dar pases. Por otro lado, también hay críticas menos constructivas como la segunda que se dedica a desprestigiar el juego.

En cuanto a las ilustraciones 18 y 19, al ser la saga FIFA jugable en todas las plataformas de juego de sobremesa (consolas y ordenador), se han diferenciado las emociones para cada una de las plataformas pertinentes. En PlayStation y en Xbox las emociones registradas son similares en ambos datos, sobresaliendo ligeramente las emociones negativas en el caso español de Xbox. Si se comprueba la polaridad discutida en el apartado anterior, por fuente, se obtiene que mientras que para todas las plataformas la mayoría de los tuits son neutros, la fuente con mayor polaridad negativa es Twitter Web App con más de un 17% (se han incluido los datos de polaridad por plataforma para todos los juegos en una tabla en el apartado **7.3 de anexos**). También se debe mencionar el

comportamiento de Nintendo Switch, puesto que pese a ser una consola atípica para jugar a este tipo de juegos, no se observan casi sentimientos realmente negativos, únicamente destacando el miedo en la comunidad española, que puede estar relacionado al poco apoyo que recibe este soporte

### Para el lanzamiento de Call Of Duty: MW2:

Ilustración 20 y 21: Análisis de sentimiento en tuits sobre Call Of Duty: MW2 en inglés y en español

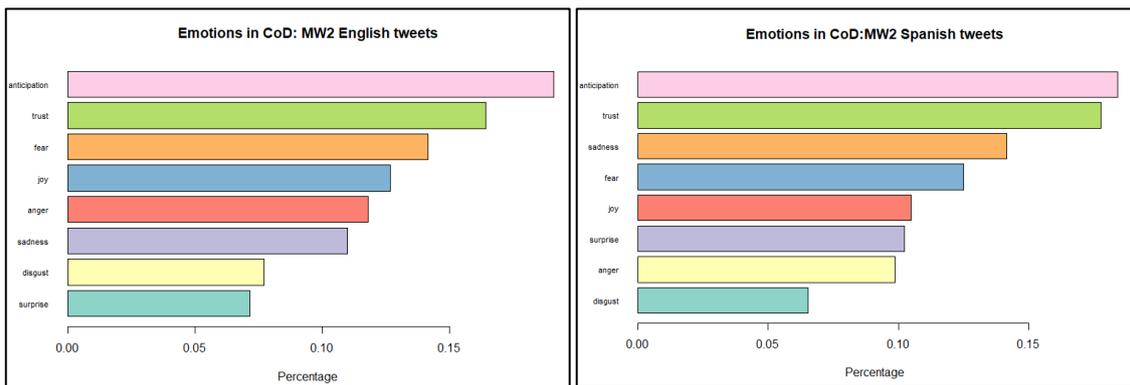
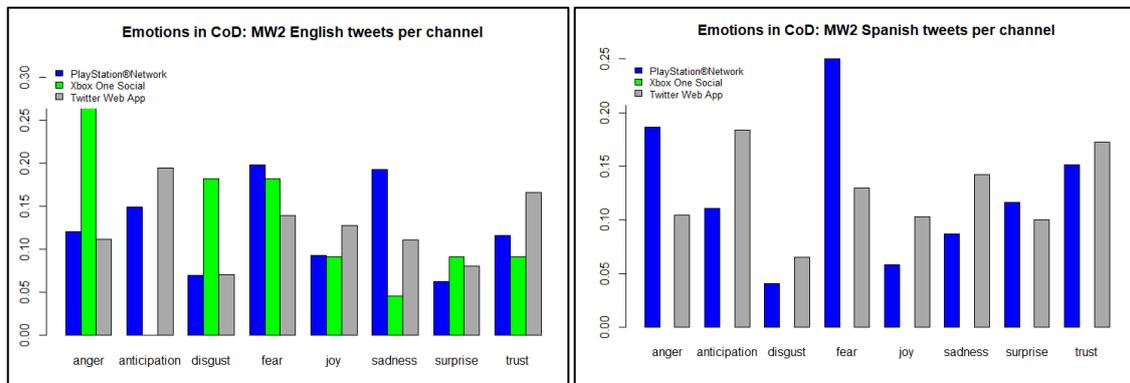


Ilustración 22 y 23: Emociones por canal de tuits para Call Of Duty: MW2 en inglés y en español



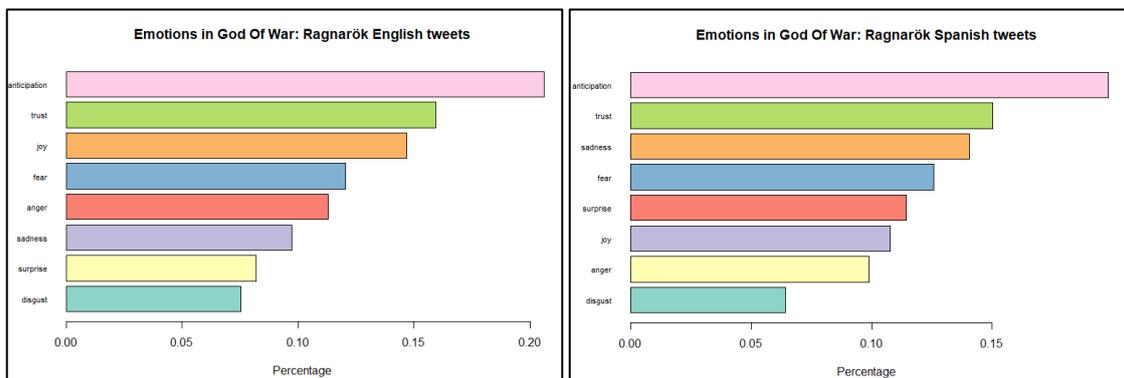
En las ilustraciones 20 y 21 se comprueba que, otra vez, los resultados obtenidos son muy similares entre los tuits en español y en inglés, con las emociones más relevantes siendo en ambos casos la confianza y la anticipación. Esto es lógico para el caso de un lanzamiento de producto de una saga tan establecida como esta en la industria. Sin embargo, en ambos casos el tercer sentimiento es negativo, siendo en el caso inglés el miedo y en el español la tristeza. En el caso inglés esto pierde importancia al ser el cuarto sentimiento más relevante la alegría (estando muy próxima a la tercera), mientras que en el caso en español gana relevancia puesto que es el miedo. Claramente se puede observar que hay una proporción importante de la muestra expresando miedo o tristeza en el caso

español lo cual se puede estudiar en profundidad para entender la razón de estos sentimientos.

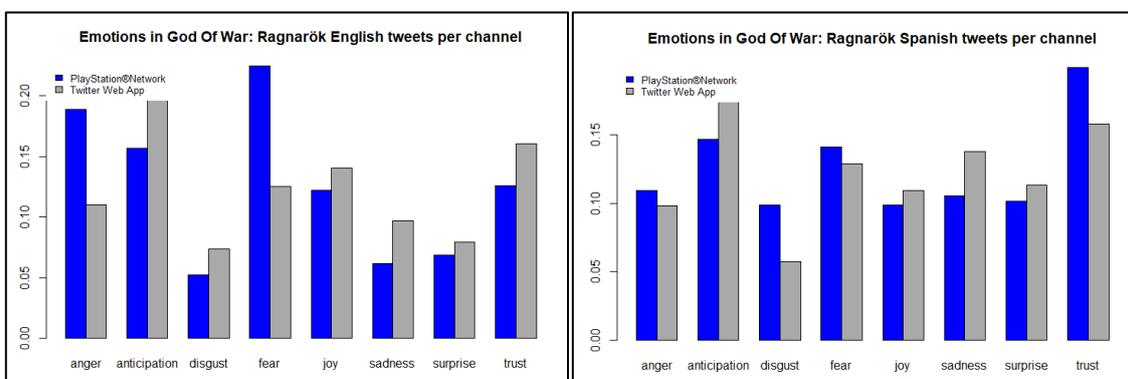
Pasando al análisis por plataformas con las ilustraciones 22 y 23 es extremadamente notable el hecho de que en la muestra en español no haya presencia de tuits provenientes de Xbox, lo cual es una clara prueba del dominio que tiene la consola de Sony, PlayStation, sobre el mercado español. También se puede comprobar como ciertas emociones negativas se disparan al comprobar los tuits emitidos desde consolas contra los tuits de la aplicación web, que seguramente se deba al hecho de que la gente tuiteando desde sus consolas lo haga porque se ha producido un hecho que les haga tuitear, ya sea un error en el juego o algo que no les haya gustado y por ende se sientan inclinados a criticar.

### Para el lanzamiento de God Of War: Ragnarök

*Ilustración 24 y 25: Análisis de sentimiento en tuits sobre God Of War: Ragnarök en inglés y en español*



*Ilustración 26 y 27: Emociones por canal de tuits para God Of War: Ragnarök en inglés y en español*



En las figuras 24 y 25 se vuelven a observar ciertos patrones como es que confianza y anticipación sean en ambos casos las emociones que más peso tienen en el texto, esto no

deja de dar credibilidad al análisis puesto que es un hallazgo lógico a la situación. En cuanto a diferencias entre los tuits ingleses y españoles, se repite el hecho de que la comunidad de habla hispana parezca ser más negativa, repitiéndose miedo y tristeza como algunas de las emociones principales.

También se repiten patrones observados en los lanzamientos anteriores cuando se profundiza en el análisis por plataforma (figuras 26 y 27). En el caso de los tuits en inglés es evidente que emociones como enfado y miedo ven comportamientos mucho más extremos en consola (PlayStation) que en la aplicación web. Sin embargo, esto no se da para los datos en español, que ven un comportamiento similar entre plataformas.

### Para el lanzamiento de Pokémon Escarlata y Púrpura:

Ilustración 28 y 29: Análisis de sentimiento en tuits sobre Pokémon Escarlata y Púrpura en inglés y en español

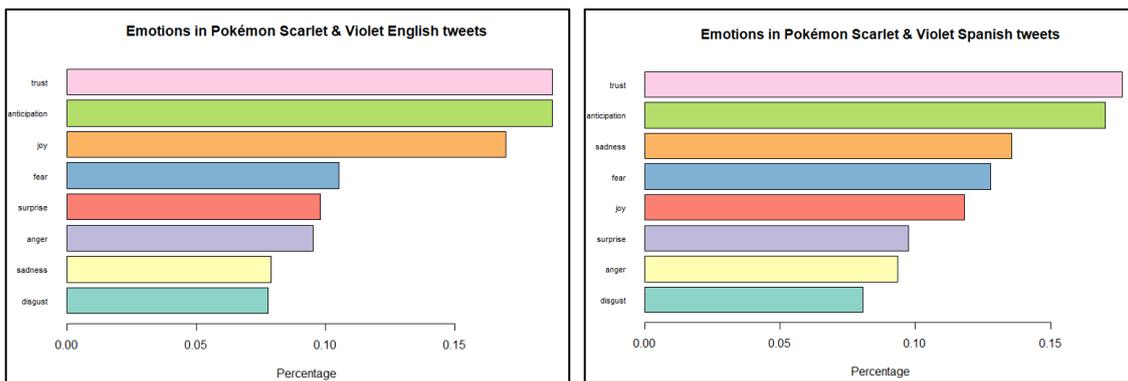
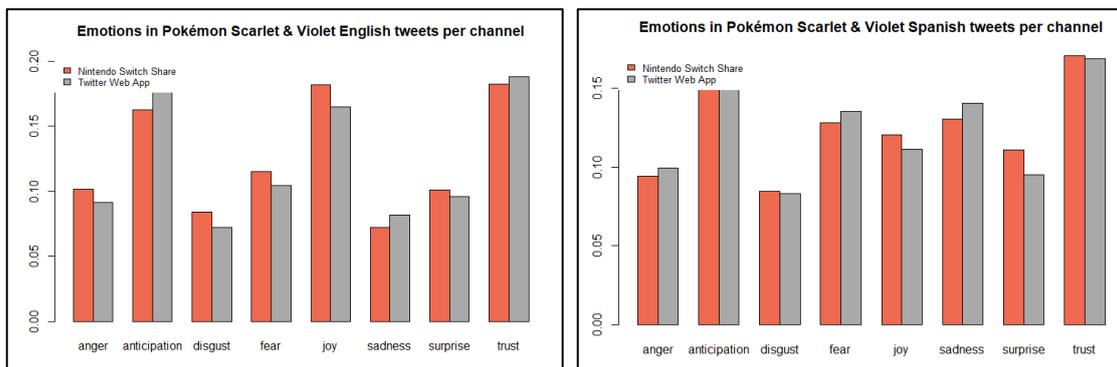


Ilustración 30 y 31: Emociones por canal de tuits para Pokémon Escarlata y Púrpura en inglés y en español



Como se aprecia en las figuras 28 y 29, en este último lanzamiento los resultados obtenidos en ambos casos son muy similares. La excepción de esto se da al encontrar tristeza en el top 3 de emociones de los tuits en habla hispana, frente a alegría en los tuits ingleses, repitiéndose el mismo patrón que con el resto de los lanzamientos.

Por otro lado, si se analizan las distintas plataformas en las ilustraciones 30 y 31 se puede comprobar que en este caso hay muy poca diferencia entre fuentes, significando que la comunidad es extremadamente similar transversalmente en las plataformas. También es destacable que mientras que en los tuits en inglés hay mucha diferencia de peso entre las 3 emociones principales y el resto, en el caso de los tuits en español existe una relación mucho más equilibrada.

En definitiva, a parte de la capacidad que ofrece de estudio sobre emociones particulares, ha quedado demostrado que este algoritmo ayuda a filtrar comentarios de interés para los desarrolladores, tanto buenos como malos. Así se pueden identificar comentario informativamente valiosos para el análisis, y usuarios que pueden ser notablemente beneficiosos o perjudiciales para la conversación en redes. También es de utilidad la diferenciación entre plataformas, que supone que en el caso de unos resultados muy diferentes entre estas se pueda focalizar ciertas acciones, ya sean de marketing como arreglos en el juego, para solucionar los problemas específicos que pueda tener esa parte de la comunidad.

### **3.5 Redes de términos y temas**

Este apartado supone una primera aproximación al concepto de topic modelling discutido con anterioridad al hablar de aplicaciones previas del análisis de redes sociales. La elaboración de estas redes permite encontrar temas y agrupar estos en clústeres.

La primera parte va a consistir en utilizar los datos limpios para producir una TDM (Term Document Matrix), que contendrá todos los términos que están conectados entre sí de forma binaria. Es decir, no se va a observar cuantas veces un término aparece con otro sino cuántas relaciones tiene un término a lo largo del corpus.

Para que los datos sean interpretables a nivel visual, una vez producida la TDM, esta va a ser reducida a un tamaño gestionable según el caso. Se va a utilizar la librería `igraph` para pasar de esta matriz a una red según teoría de grafos, y se utilizará un algoritmo de clustering que utiliza `edge betweenness` como medida para identificar grupos o comunidades. Esto significa que los elementos de la red más importantes de cara a formar clústeres serán aquellos vínculos que más alta tengan esta métrica: que conectan grupos que si no fuese por ese vínculo no tendrían conexión. Estos vínculos “conectores” se utilizarán como señal de que se ha completado un clúster, puesto que la lógica que sigue el algoritmo es la de agrupar vínculos entre nodos cuya `betweenness` es relativamente baja (que están aislados respecto al resto de la red), una vez se han agrupado todos estos nodos y se llega a un vínculo en la que la métrica es alta, significa que se ha encontrado otro clúster y se cierra el primer grupo para comenzar a agrupar el siguiente. Esto implica que normalmente se irán incorporando vínculos o edges (y los nodos que estos conectan) desde el exterior hacia el interior de la red (Luo, Zhong, Ying, & Fu, 2011). Así se han producido las figuras 32 a 39.

En las redes representadas se observan la existencia de ciertos patrones que se dan en la industria, como es la existencia de clústeres de temas relacionados con la retransmisión en directo para casi todos los casos. Se mencionan plataformas como YouTube, Twitch, PSLive, etc. Sin embargo, estas figuras también sirven para apreciar particularidades para cada título. Un ejemplo es el caso de las redes relacionadas con Pokémon (ilustraciones 38 y 39) en las que se ve como la comunidad se divide en conversaciones sobre la serie de animación, el contenido del juego una vez completada la historia (los Pokémonshinies) y la discusión sobre que versión del juego comprar. Los patrones, clústeres y particularidades de cada red pueden encontrarse comentados después de las ilustraciones.



## Call Of Duty: Modern Warfare 2

Ilustración 34: Redes de términos y comunidades para Call Of Duty: Modern Warfare 2 en inglés

Term communities Network for CoD: MW2 English tweets

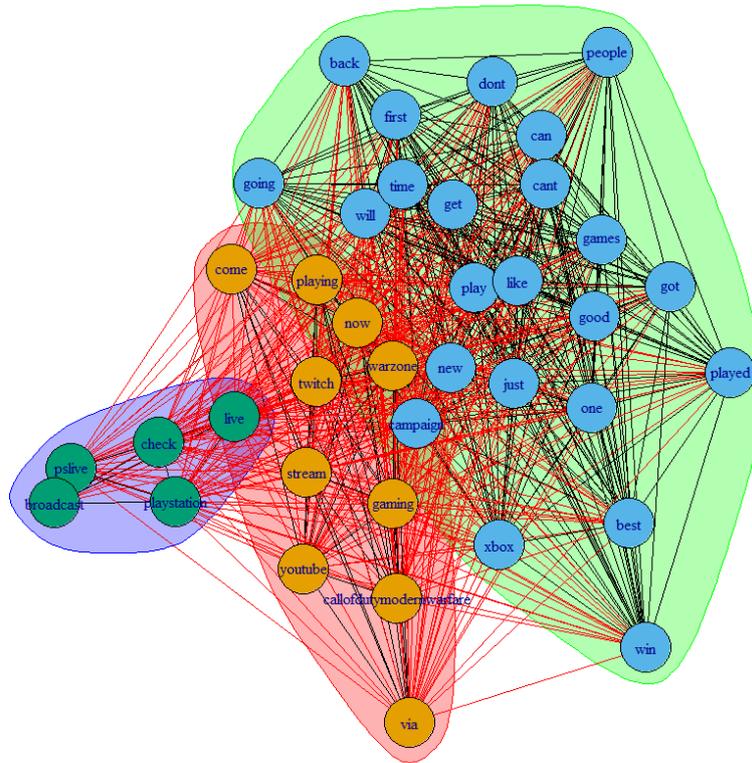
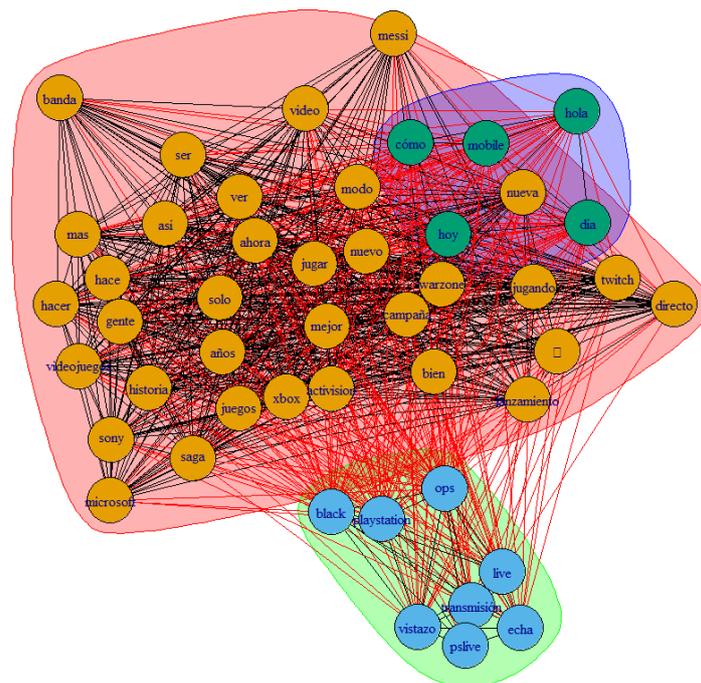


Ilustración 35: Redes de términos y comunidades para Call Of Duty: Modern Warfare 2 en español

Term communities Network for CoD: MW2 Spanish tweets







## **Información obtenida de las redes:**

Para todos los casos se han podido identificar 3 o más clústeres o comunidades, diferenciados por colores, que se van a explicar por videojuego a continuación. Las explicaciones han sido desarrolladas gracias a un conocimiento general de todos los títulos, así como al análisis previo del contenido que se ha realizado en este proyecto.

### **Caso 1: lanzamiento de FIFA 23**

En inglés: ilustración 32.

- Rojo: modo de juego de FUT, o 'FIFA Ultimate Team', algunas de las palabras clave que delatan esta información son: "coins", "packs", "totw" o "fut".
- Morado: emisión en directo de contenido, con las palabras "live", "today" o "now", en esta comunidad seguramente se encontrarían otras palabras que hiciesen alusión a plataformas para compartir contenido como YouTube o Twitch, si no se hubiese reducido el número de términos.
- Verde: jugabilidad normal o problemas de ejecución del juego, destacan los términos "cant", "dont", "game", "please", "play".

En español: ilustración 33.

- Rojo: las novedades del juego en cuanto a jugadores y modos y la jugabilidad normal de este, con palabras como "nuevo", "jugadores", "modo" o "año".
- Morado: modo de juego de FUT, con "fut", "futchampions" o "totw".
- Verde: retransmisión del juego o contenido en directo, con "twitch" o "directo".

### **Caso 2: lanzamiento de Call Of Duty: Modern Warfare 2**

En inglés: ilustración 34.

- Rojo: retransmisión de contenido en directo y modo de juego de Call Of Duty Warzone con "warzone", "youtube", "twitch" y "live".
- Morado: entorno alrededor de PlayStation y sus noticias, algunos tokens destacados son "pslive", "playstation" o "broadcast".
- Verde: comentarios sobre los modos del juego y la jugabilidad normal, destacan "campaign", "new", "play", "games" o "good".

En español: ilustración 35.

- Rojo: Verde: comentarios sobre los modos del juego y el juego en general, destacan “campaña”, “lanzamiento”, “historia” o “jugar”
- Morado: parece tener que ver con el juego en móvil y la retransmisión con “mobile”, “hoy” y estando muy cercano a “twitch” y “directo” (aunque estos dos últimos han sido clasificados en el clúster rojo).
- Verde: entorno alrededor de la plataforma PlayStation y sus noticias, algunos tokens destacados son “pslive”, “playstation”, “transmisión” y “live”.

Cabe destacar que en ambos casos para este juego la plataforma Xbox no logra formar un clúster propio y forma parte del clúster más general, mientras que PlayStation (en ambos) y móvil (en español) sí. Esto puede ser de utilidad para segmentación de usuarios.

### **Caso 3: lanzamiento de God Of War Ragnarök**

En este juego las comunidades están peor separadas entre sí, pero aun así se pueden identificar ciertas narrativas entre grupos.

En inglés: ilustración 36, en esta instancia se observan 4 clústeres.

- Rojo: parece centrarse alrededor de los adelantos previos al juego que se emitieron en YouTube sobre la historia y los personajes. Esto queda evidenciado por alguno de sus tokens más relevantes como “youtube”, “thor”, “kratos”, “atreus” o “video”.
- Morado: retransmisión en directo del juego en Twitch, con términos como “twitch”, “come”, “stream”, “now” o “live”.
- Verde: anticipación al lanzamiento, gente que lleva esperando mucho tiempo y no puede esperar a jugar: “cant”, “wait”, “will”, “play” o “day”.
- Azul: usuarios que han terminado o han jugado al juego y opinan sobre él con palabras como “finished”, “best”, “story”, “amazing”, “love” o “hours”.

En español: ilustración 37.

- Rojo: comunidad alrededor PlayStation, el lanzamiento y la retransmisión en directo del creador de contenido TheGrefg. Esto se ve en tokens como “playstation”, “thegrefgsgowragnarok”, “lanzamiento” o “twitch”.

- Morado: usuarios en general, hablando sobre el querer jugar. Destacan: “jugar”, “quiero”, “ahora” y “nuevo”.
- Verde: conversación sobre cuál va a ser el mejor juego del año (Game Of The Year o GOTY), enfrentando a este lanzamiento con otro popular: Elden Ring. Algunas de las palabras clave son “goty” “elden”, “ring”, “historia” y “mejor”.

#### **Caso 4: lanzamiento de Pokémon Escarlata y Púrpura**

En inglés: ilustración 38

- Rojo: los temas generales de esta comunidad se refieren al juego, y al contenido de este una vez completada la historia principal, siendo importantes palabras como “shiny” o “shinypokemon” que se refieren a capturar Pokémon de color diferente al normal, y “charizard” que tiene que ver con un evento que emitió el desarrollador en el lanzamiento para poder capturar a un Pokémon especialmente fuerte.
- Morado: tiene que ver con la serie de animación de Pokémon, y en especial con los eventos que se desarrollaron en esta durante el lanzamiento de los juegos, donde el personaje principal consiguió la meta que llevaba persiguiendo desde que empezó la serie hace más de 20 años. Esto se ve evidenciado por palabras clave como “ash”, “finally” o “best”.
- Verde: discusión sobre que juego escoger, si la edición escarlata o la edición púrpura, los tokens que señalan esto son “scarlet” y “violet”.

En español: ilustración 39

- Rojo: Tiene que ver con la serie de animación de Pokémon: “anime”, “ash” o “campeón”.
- Morado: Discusión sobre las versiones del juego y cual escoger: “pokemonviolet”, “escarlata” o “púrpura”.
- Verde: Se mencionan temas sobre el juego en general, pero destaca la presencia de tokens como “primer” o “shiny” evidenciando que hay cierto revuelo en la comunidad del juego alrededor de este evento (capturar Pokémon de color distinto al original y en concreto que sea el primero que obtienes).

Es especialmente significativo que se siga la misma estructura en ambos casos: tiene lugar una división en dos grandes comunidades interconectadas y la existencia de una comunidad pequeña dentro de otra. De los 4 casos observados, este es el único en el que no se puede observar alguna diferencia fundamental entre los datos en inglés y los datos en español.

### **3.6 Identificación de influencers**

Pese a que en una primera instancia del proyecto se planteaba realizar un proceso parecido al del apartado anterior para los usuarios de los datos obtenidos, esto finalmente no ha sido posible por la naturaleza de los datos del estudio. Esto se debe a que, debido a que es más conveniente para los análisis de contenido realizados en el estudio, como son el sentiment analysis o la elaboración de redes de términos, no se incluyeron retuits (que otro usuario publique nuevamente tuits de otro usuario, destacándolo) en la query o llamada de extracción de datos. Se ha incluido un apartado en **anexos** explicando el cambio que se tiene que hacer a la extracción, así como el proceso para elaborar redes de usuarios según retuits.

En vez de esto, para este caso de estudio se han valorado otras metodologías de detección de influencers que no estén relacionadas con teoría de grafos. Así, se ha desarrollado un protocolo simple en base al número de retuits, de favoritos (que un usuario marque el tuit de otro usuario como favorito) y el número de seguidores de cada cuenta que forma parte de los datos de cada lanzamiento, y se han señalado usuarios, separando los verificados de los no verificados, según la variable “verified”.

La estructura de usuarios de Twitter podría llevar a entender que únicamente aquellas personas o cuentas verificadas son relevantes para entender cómo se desarrollan ciertos eventos. Esto se debe a que, por definición, en el momento en el que se obtuvieron los datos únicamente estaban verificadas las personas que eran consideradas “celebridades” y por lo tanto necesitaban que Twitter acreditase la veracidad de su cuenta. Esta funcionalidad ha sido vital tanto para instituciones como para individuos, como puede ser el caso de creadores de contenido, como los que seguramente se vean señalados en el análisis. Sin embargo, para publishers de videojuegos puede ser quizás de igual o mayor importancia conocer a los líderes “informales” de sus comunidades, ya que mueven el



Ilustración 43: Influencers no verificados para tuits de FIFA 23 en inglés

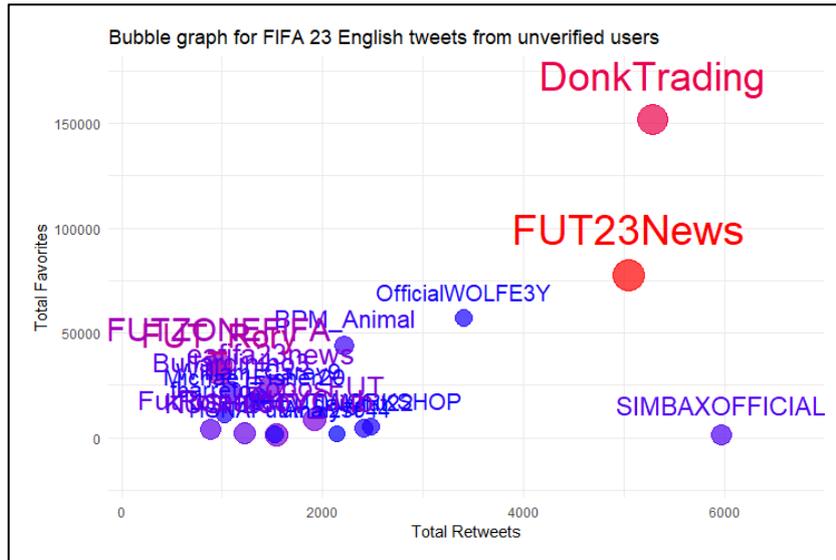


Ilustración 44 y 45: Top 20 usuarios verificados y no verificados con más retuits para FIFA 23 en español

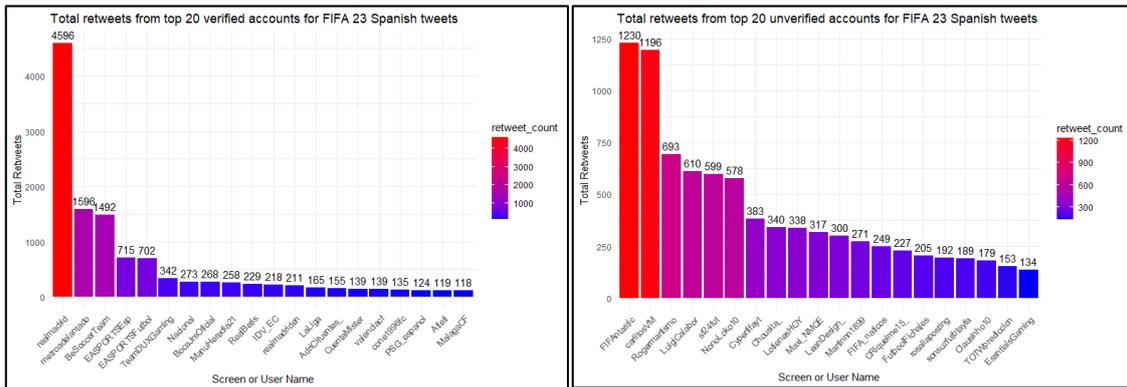


Ilustración 46: Influencers verificados para tuits de FIFA 23 en español

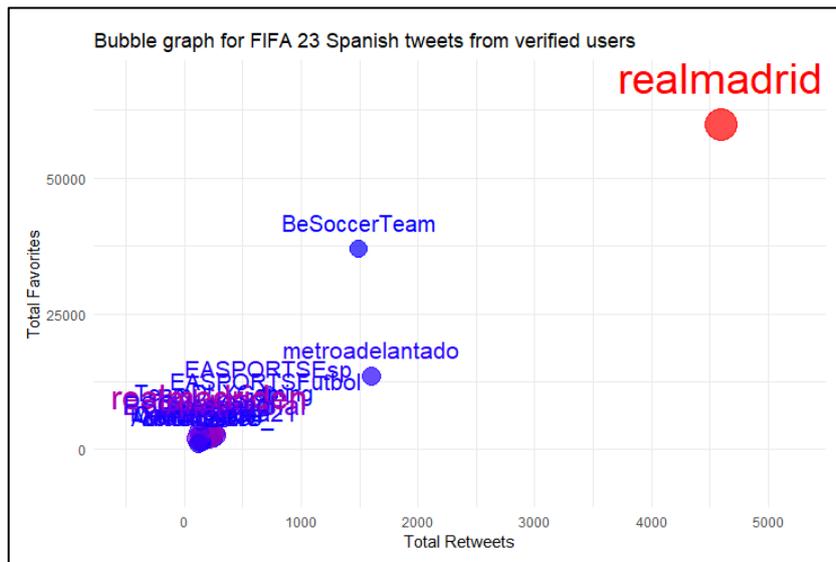
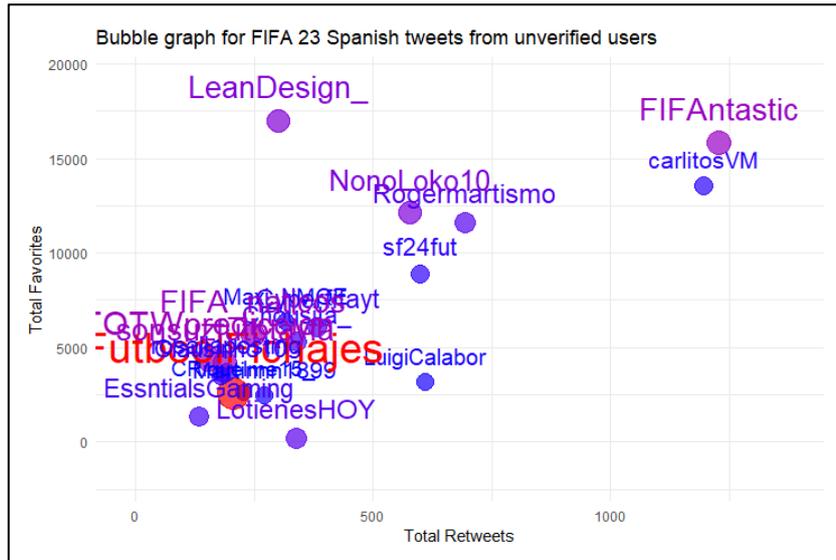


Ilustración 47: Influencers no verificados para tuits de FIFA 23 en español



## Call Of Duty Modern Warfare 2

Ilustración 48 y 49: Top 20 usuarios verificados y no verificados con más retuits para Call OfDuty: MW2 en inglés

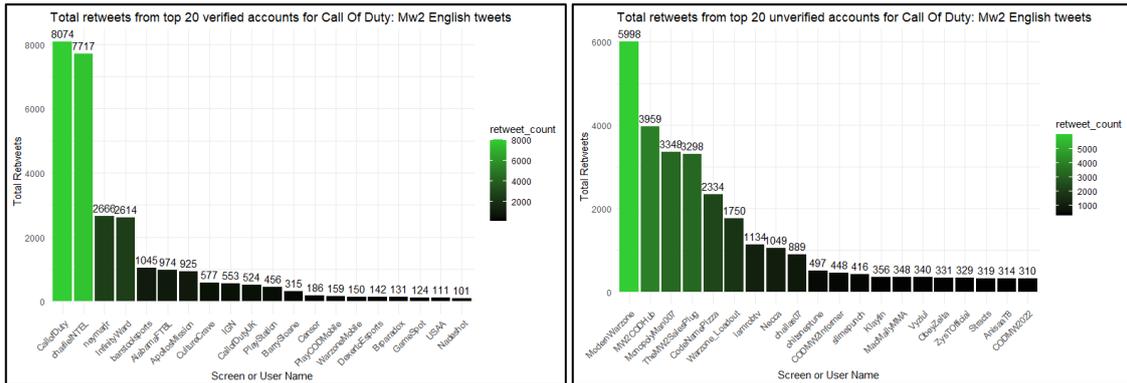


Ilustración 50: Influencers verificados para tuits de Call OfDuty: MW2 en inglés

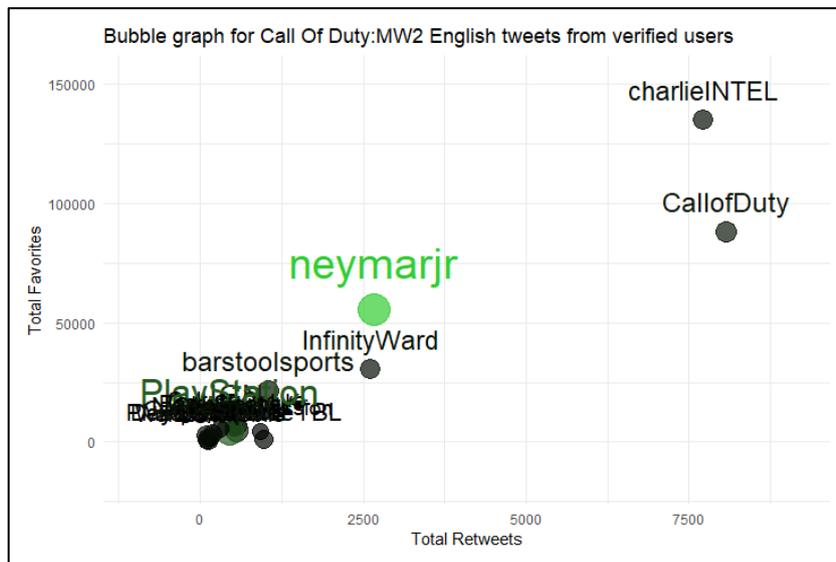


Ilustración 51: Influencers no verificados para tuits de Call Of Duty: MW2 en inglés

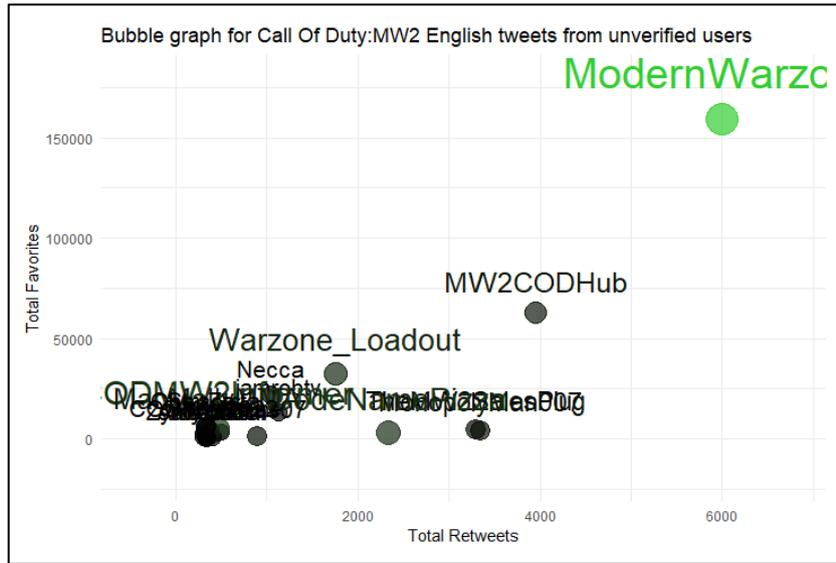


Ilustración 52 y 53: Top 20 usuarios verificados y no verificados con más retuits para Call Of Duty: MW2 en español

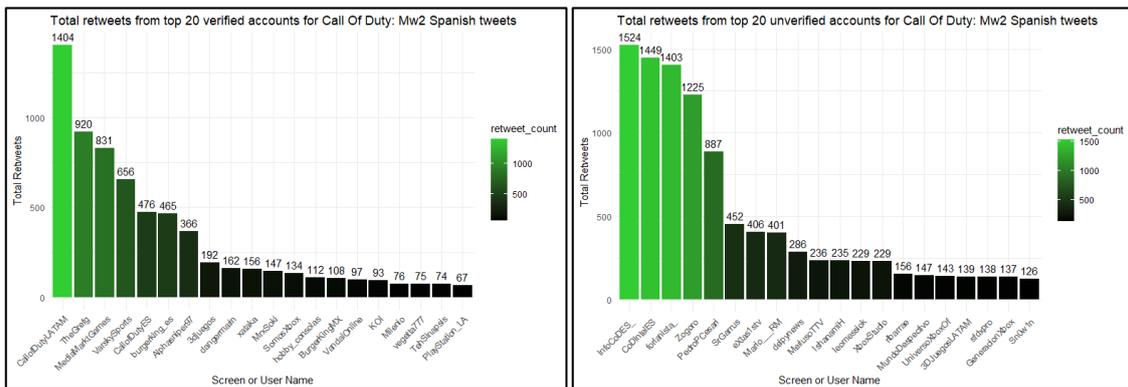


Ilustración 54: Influencers verificados para tuits de Call Of Duty: MW2 en español

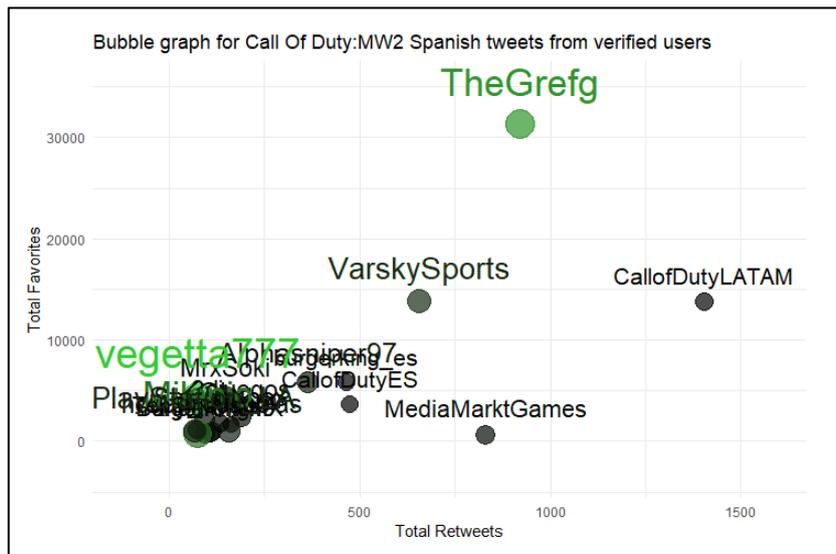
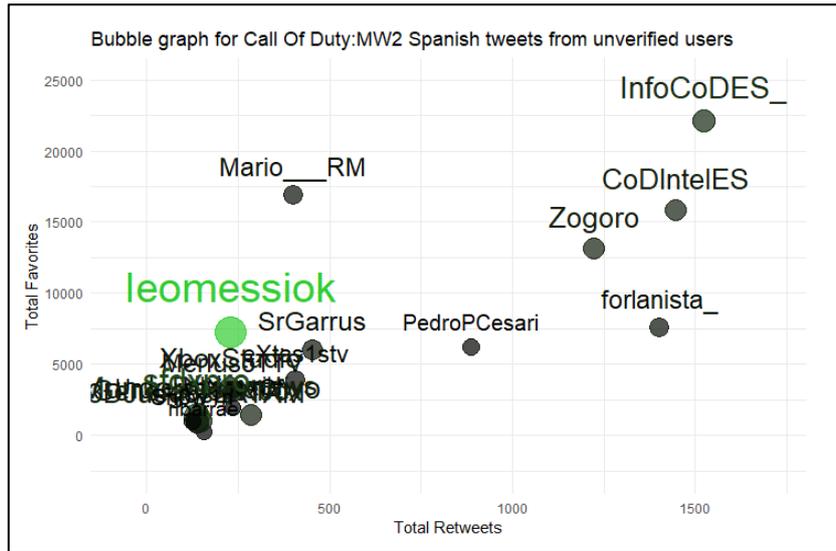


Ilustración 55: Influencers no verificados para tuits de Call Of Duty: MW2 en español



### God Of War: Ragnarök

Ilustración 56 y 57: Top 20 usuarios verificados y no verificados con más retuits para God Of War: Ragnarök en inglés

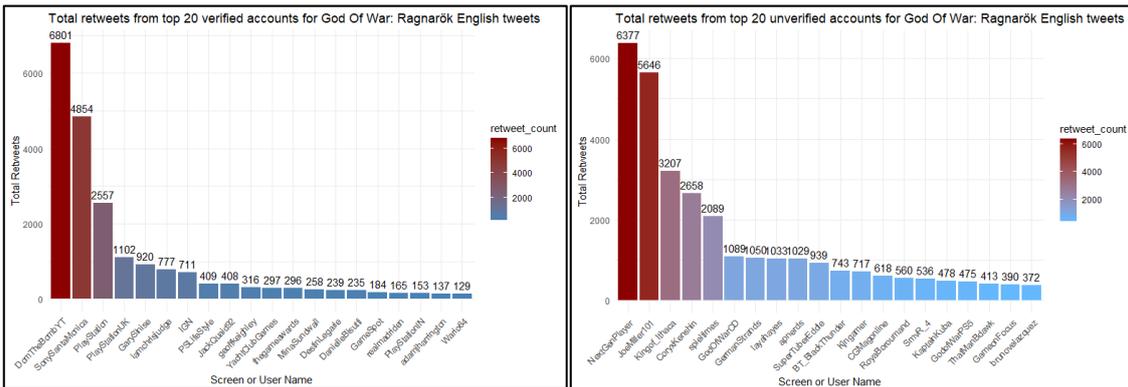


Ilustración 58: Influencers verificados para tuits de God Of War: Ragnarök en inglés

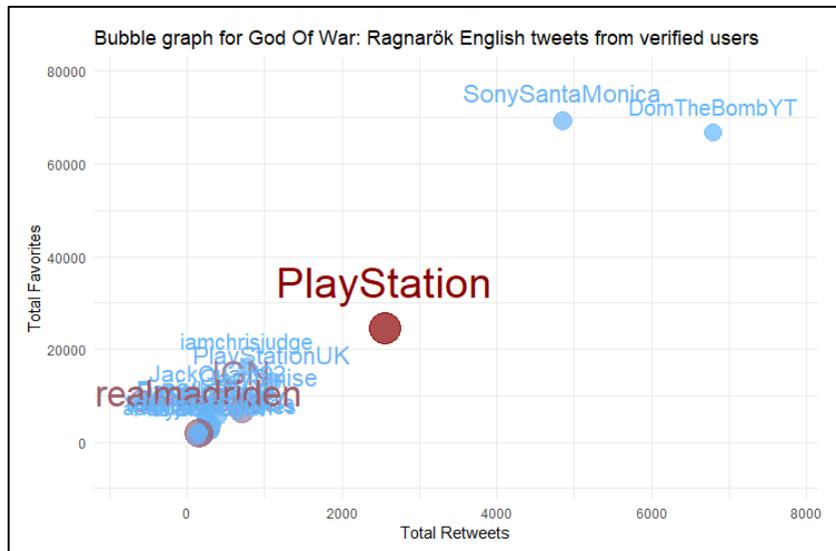






Ilustración 67: Influencers no verificados para tuits de Pokémon Escarlata & Púrpura en inglés

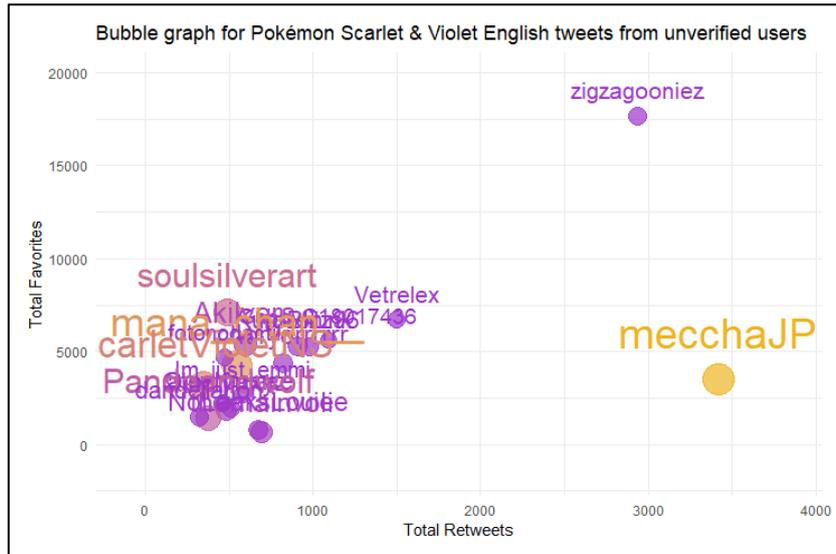


Ilustración 68 y 69: Top 20 usuarios verificados y no verificados con más retuits para Pokémon Escarlata & Púrpura en español

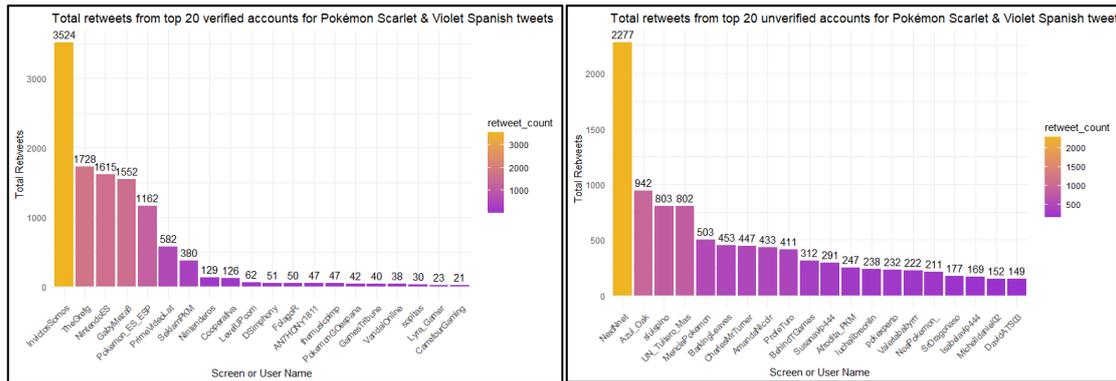


Ilustración 70: Influencers verificados para tuits de Pokémon Escarlata & Púrpura en español

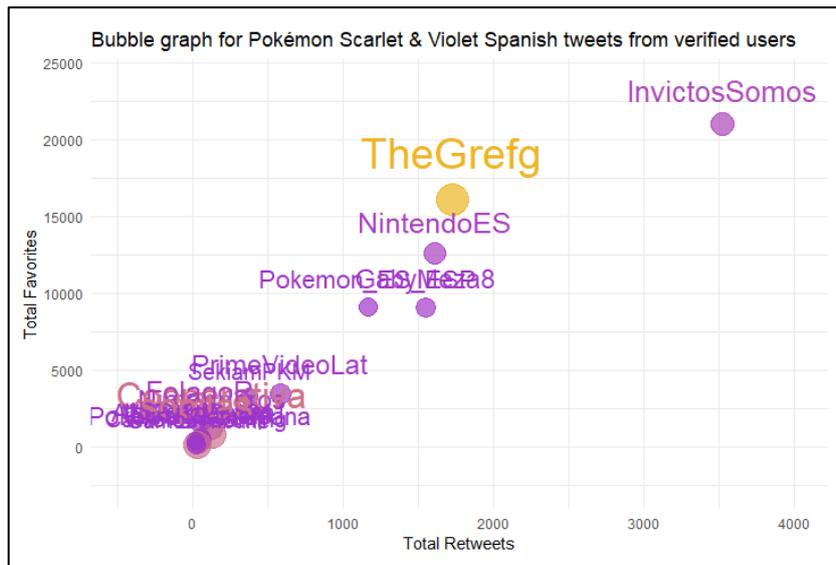
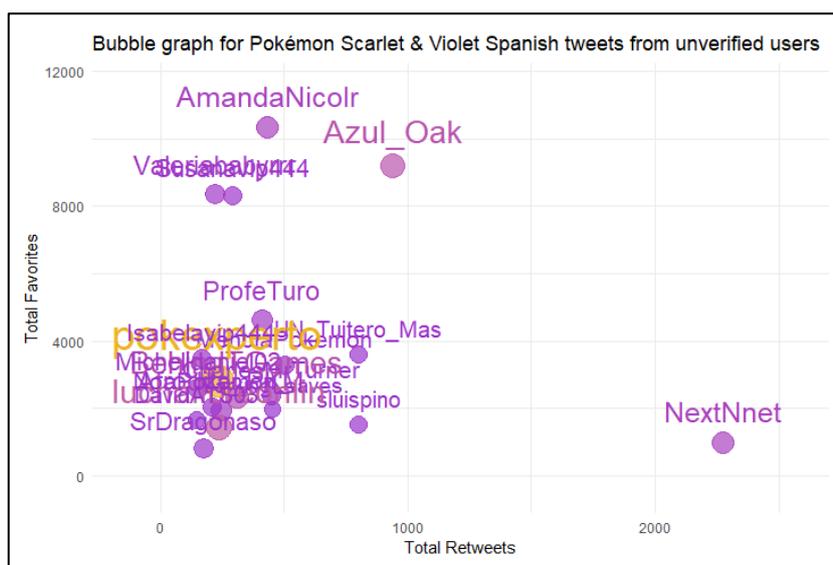


Ilustración 71: Influencers no verificados para tuits de Pokémon Escarlata & Púrpura en español



Todos los usuarios que han sido señalados en estos gráficos son de relevancia para el lanzamiento del videojuego, ya que son los 20 que más interacciones han recibido por medio de retuits. Sin embargo, observando los gráficos, fácilmente se puede singularizar a unos pocos. Las cuentas oficiales de desarrolladores o publishers como Santa Monica Studio y PlayStation en el caso de God Of War o Nintendo en el caso de Pokémon han sido ignoradas para estos resultados, puesto que su importancia en la red es evidente. En el caso de los tuits en inglés de Pokémon Escarlata & Púrpura se han tenido que eliminar esas cuentas de los datos porque eclipsaban al resto de usuarios. Los resultados obtenidos se han señalado en las siguientes tablas:

Tabla 7: Resumen de influencers más relevantes para tuits en inglés

TÍTULO	VERIFICADOS	NO VERIFICADOS
<b>FIFA 23</b>	realmadrid, AdamEShelton	DonkTrading, FUT23News, SIMBAXOFFICIAL
<b>Call Of Duty: MW2</b>	charlieINTEL, neymarjr	ModernWarzone, MW2CODHub
<b>God Of War: Ragnarök</b>	DomTheBomBYT	JoeMiller101, NextGenPlayer, CoryxKenshin
<b>Pokémon Escarlata &amp; Púrpura</b>	pokejungle, PokemonNewsUK	zigzagooniez, mecchaJP, Vetrelex

Tabla 8: Resumen de influencers más relevantes para tuits en español

TÍTULO	VERIFICADOS	NO VERIFICADOS
<b>FIFA 23</b>	realmadrid, metroadelantado, BeSoccerTeam	FIFAntastic, carlitosVM, Rogermartismo
<b>Call Of Duty: MW2</b>	TheGrefg, VarskySports, MediaMarktGames	InfoCoDES_, CoDIntelES, Zogoro, forlanista_
<b>God Of War: Ragnarök</b>	realmadrid, vegetta777, TheGrefg	GodofWar_ES, Dios_Manco, JWulen
<b>Pokémon Escarlata &amp; Púrpura</b>	InvictosSomos, TheGrefg, GabyMeza8	NextNnet, Azul_Oak, AmandaNicolR

En este apartado llama especialmente la atención la presencia constante de usuarios que actúan como medios de noticias, o de información sobre el juego estudiado. Estos suponen algunos de los usuarios no verificados más importantes de los lanzamientos observados. Algunos ejemplos son, en español “InfoCoDES\_” o “FIFAntastic” para el Call of Duty y el FIFA respectivamente. Otro hallazgo relevante es la existencia de influencers, que, pese a que se trate juegos muy diferentes, son relevantes para casi todos los juegos: esto es el caso de TheGrefg, que se encuentra entre los perfiles verificados más notables en 3 de los 4 títulos analizados en español. Las implicaciones comerciales que puede tener algunos de estos hallazgos y la identificación de estas personas de gran interés en la comunidad se desarrollarán en el siguiente apartado.

## **4. DISCUSIÓN DE LOS RESULTADOS Y APLICACIÓN PRÁCTICA**

### **4.1 Resultados del caso**

Una vez finalizada la fase de análisis del proyecto se han obtenido una serie de resultados que pueden proporcionar conocimientos clave de lanzamientos de algunos de los juegos más populares en la actualidad, y de la industria de los videojuegos en general. Algunos de los hallazgos más importantes son:

1. Existen evidencias para afirmar que en la mayoría de los casos observados la comunidad de Twitter de habla inglesa tiene comportamientos diferentes a la de habla hispana. Se han visto ejemplos concretos a lo largo de todo el análisis, algunos de estos son:
  - a. Análisis de frecuencias de términos: en los juegos más de masas se identifican intereses principales diferentes: por ejemplo, en cuanto al FIFA 23 en los tuits en español la retransmisión de contenido parece tener más relevancia relativa que en los datos en inglés que parecen más centrados en la jugabilidad del título.
  - b. Análisis de sentimiento: en los datos en español se perciben emociones más negativas que en los datos en inglés. Esto se confirma en el análisis de polaridades, en la que para todos los casos la media española es más negativa que la inglesa a lo largo del mes observado.
  - c. Análisis de redes de términos y clústeres: se puede observar cómo en 3 casos de los 4 estudiados, cuando se comparan los datos en español con los escritos en inglés, las redes tienen diferencias relevantes entre sí. Un ejemplo concreto de esto se da con el título God Of War: Ragnarök, donde en ambos idiomas se pueden encontrar clústeres exclusivos para estos como es el caso del clúster sobre la discusión de los videos en YouTube previos al lanzamiento para los datos en inglés, o el clúster sobre la discusión del premio al mejor juego del año (GOTY) en los datos en español. Otro caso concreto de esto es la presencia de un clúster de la versión móvil del juego para Call Of Duty:MW2 en español. La excepción de esto es Pokémon, donde ambas redes siguen la misma estructura,

señalando una gran similitud entre los jugadores de este juego, independientemente de si son de habla hispana o inglesa.

2. Para todos los lanzamientos se ha obtenido información relevante que podría ayudar a los desarrolladores y publishers de estos juegos a mejorar su contenido o lograr mejores relaciones con su comunidad, tanto a nivel español como inglés.

Algunos ejemplos de esto para cada juego son:

- a. FIFA 23: se ha podido comprobar la tremenda relevancia de ciertos modos de juego como es el caso de FUT (FIFA Ultimate Team).
- b. Call Of Duty: MW2: ha quedado demostrada la importancia que tuvo la campaña de este juego a la hora de lograr captar atención para el lanzamiento. Además de esto, ciertas interacciones con otros juegos o modos de juego como es Call Of Duty Warzone han supuesto nexos clave en la comunidad de esta franquicia durante el mes observado.
- c. God Of War: Ragnarök: se ha podido observar la reacción de la comunidad ante actividades promocionales de la comunidad como el contenido en YouTube previo al lanzamiento y con emisiones en directo de usuarios muy influyentes en la red, con el caso de TheGrefg. Otro caso del impacto de estos eventos es la conversación que genera el que la comunidad pueda considerar que un juego aspire a un premio como el GOTY.
- d. Pokémon Escarlata & Púrpura: se han dado sinergias entre el anime o la serie de Pokémon, con el lanzamiento de esta nueva generación de la saga. A esto se suma que se ha podido identificar la importancia que tiene la comunidad dedicada al contenido que tiene el juego después de la historia, como es capturar Pokémon “shinies”. Se ha podido observar que tienen una comunidad muy homogénea tanto cuando se comparaban datos en inglés versus datos en español, como si se profundizaba en análisis más técnicos como analizando el sentimiento por plataforma (los datos de Nintendo Switch y de aplicación web eran prácticamente idénticos).
- e. Se han observado diferencias claras entre el comportamiento de las comunidades los distintitos juegos. Por ejemplo, se ha contrastado que las comunidades de juegos como FIFA 23 y Pokémon tienen particularidades diferentes.

Habiendo visto resultados concretos para cada título cabe destacar la presencia de ciertos patrones que se han comprobado a lo largo de todo el análisis y para todos los juegos:

1. Importancia de la retransmisión en directo y la anticipación que eso genera para la comunidad: en casi todos los juegos se han visto clústeres dedicados al seguimiento de estos eventos en vivo en plataformas como Twitch, YouTube o PSLive.
2. Presencia clara y objetiva de influencers o usuarios influyentes tanto en la categoría de verificados como en la de no verificados: en todos los casos se ha podido identificar a uno o más usuarios clave para la red.
3. Énfasis en el juego en sí: pese a que en todos los casos de este estudio siempre se han encontrado ciertas particularidades propias de cada lanzamiento, en todos los textos analizados hay una importancia fundamental en el diseño de cada juego, ya sea por la historia, por la competitividad o por la propiedad intelectual, entre otros.
4. Relevancia general de medios de noticias dedicados exclusivamente a cada juego observado para todas las comunidades, que actúan como influencers no verificados en todas las comunidades. Como se mencionaba en el apartado **3.6**, se pueden observar para todos los casos entre las cuentas más relevantes a usuarios que actúan como medios informativos sobre las novedades de cada juego.
5. Existencia de conversación por parte de la comunidad que va más allá del juego y trata temas como las acciones estratégicas de desarrolladores y comparaciones entre estos.

#### **4.2 Aplicación práctica a la industria: uso empresarial**

Los procedimientos desarrollados y la información obtenida a lo largo de este proyecto son útiles a nivel empresarial de la siguiente forma:

En primer lugar, el identificar de forma precisa los temas más relevantes alrededor del lanzamiento de un videojuego genera información clave sobre los consumidores de dicho título y sus intereses. Esta información es fundamental para desarrollar la estrategia general y las campañas de marketing, mediante la elaboración de acciones que aumenten el grado de compromiso de los usuarios. Además, el conocer estos intereses y las personas concretas que están interesadas en ellos puede ayudar a mejorar el diseño del propio juego. Un ejemplo de esto es en el caso estudiado de Pokémon Escarlata & Púrpura con

los Pokémon “shinies”. Vista la relevancia que tienen en la comunidad, podrían incluirse en el diseño del juego ciertos eventos relacionados con estos, asegurando así un aumento de la actividad de los jugadores.

En segundo lugar, tener la capacidad de señalar personas de interés en la red, ya sean influencers famosos o usuarios normales que forman una parte esencial de la comunidad puede ser muy beneficioso a la hora de gestionar, tanto la publicidad del juego como el propio estado de la comunidad. Mediante la colaboración con estos, ya sea creando contenido en otras plataformas o directamente en Twitter, se puede llegar de forma directa y efectiva a la comunidad alrededor de un juego. También es relevante que estos usuarios seguramente no participen únicamente en la conversación del juego analizado, sino que formarán parte de muchas otras comunidades con temas y usuarios similares. Esto permitiría al publisher del juego analizado alcanzar a otros usuarios de las redes sociales con atributos muy similares a los de su propia audiencia.

En tercer lugar, aplicar un análisis de sentimiento profundo que permita diferenciar entre emociones diversas y relevantes para una ocasión como puede ser el lanzamiento del juego ayudaría a la empresa que publica el juego a entender realmente cómo este es recibido por los jugadores. De esta forma puede detectar posibles comentarios tanto positivos, como seguramente más importante, negativos, que ayuden al desarrollador a encontrar problemas con el juego como los que se han visto en ciertos lanzamientos de este estudio. Teniendo esto en cuenta se podrá mejorar la satisfacción general de sus consumidores. No solo esto, sino que, tener información sobre el sentimiento promedio de la comunidad durante un periodo de tiempo, como se ha visto con el análisis de polaridad, permitiría a las empresas entender en detalle la evolución del estado de sus consumidores, y que eventos pueden afectarles.

Finalmente, el poder señalar los temas de mayor interés permitiría a los desarrolladores centrar su atención en dichos problemas u oportunidades. Para juegos tan grandes como los que se han observado, con comunidades de tamaños muy difíciles de manejar, el tener datos reales sobre qué ajustes en el juego pueden ser prioritarios es de gran interés para asegurar mayor retención de jugadores.

En resumen, con la aplicación de estas técnicas de análisis de datos las compañías de videojuegos pueden tomar decisiones informadas y adaptar sus juegos y acciones de marketing a las necesidades y preferencias de la comunidad de jugadores.

## **5. CONCLUSIÓN**

Volviendo a los objetivos que se sentaron en la propuesta se puede afirmar que estos han sido cumplidos con éxito. El objetivo principal era demostrar el valor que pueden tener las técnicas desarrolladas para realizar recomendaciones en el caso del lanzamiento de videojuegos. Su cumplimiento se puede comprobar revisando los subobjetivos:

Para el primer subobjetivo, se ha desarrollado un marco conceptual extenso, revisando conceptos teóricos necesarios para entender el funcionamiento de las técnicas planteadas y utilizando casos concretos para ejemplificar su utilización. El segundo subobjetivo se ha cumplido aplicando con éxito el análisis de sentimiento y el análisis de redes (representando las redes de términos y sus clústeres). Otros métodos desarrollados han sido: métodos de detección de influencers y formas de seguir el lanzamiento mediante series temporales (el estudio de polaridad). Además de esto se ha incluido cómo elaborar un análisis de redes de usuarios con unos datos que sí contengan retuits, en **anexos**. Finalmente, cumpliendo el tercer subobjetivo, se han realizado recomendaciones útiles a nivel de marketing, estrategia general, desarrollo del producto, etc. No solo esto, sino que se han podido encontrar diferencias clave entre cada juego y entre las comunidades de habla hispana y de habla inglesa, como es el hecho de que, en cuanto a polaridad y emociones, los datos en español siempre resultaban más negativos.

No obstante, se deben mencionar una serie de limitaciones y restricciones que han afectado al curso del trabajo. Estas son la de la extensión máxima del TFG y el uso de una cuenta de desarrollador de Twitter básica (en el momento de la extracción de datos, gratuita: noviembre 2022), que limitaba las capacidades del estudio.

En cuanto a recomendaciones futuras o áreas de mejora, con la información obtenida tras el desarrollo del caso de estudio se pueden llevar a cabo muchos otros análisis y para muchos otros títulos, productos y horizontes. Entre estas destacan:

1. Desarrollar el análisis, no para el conjunto de los datos agregados, sino por semana, para estudiar cómo cambian los comportamientos de los usuarios (sentiment analysis), las comunidades, las personas influyentes o los temas de interés.
2. Llevar a cabo otras técnicas, como la ya mencionada elaboración de la red de usuarios, estudiar la ubicación de los usuarios y realizar topic modelling extenso. Estas técnicas se han visto utilizadas en otros proyectos de investigación similares y supondrían una continuación de la lógica seguida y de los resultados obtenidos en este trabajo.

## 6. BIBLIOGRAFÍA

- Adamic, L. (2014). *SNA 3A centrality*. Obtenido de GitHub:  
<https://github.com/ladamalina/coursera-sna/tree/master/Week%203.%20Centrality>
- Belli, S., & Raventós, C. (2008). A brief history of videogame. *Athenea Digital: Revista de Pensamiento e Investigacion Social*(14), 159-179. Obtenido de <https://ddd.uab.cat/record/29923>
- Bloehdorn, S., & Hotho, A. (2004). Text classification by boosting weak learners based on terms and concepts. *In Proc. IEEE Int. Conf. on Data Mining (ICDM 04)* (págs. 331–334). IEEE Computer Society Press.
- Buijsman, M. (12 de Mayo de 2022). *The Top 10 Public Game Companies Generated \$126 Billion in 2021 as Subscriptions and M&A Shake up the Market*. Obtenido de newzoo: <https://newzoo.com/insights/articles/the-top-10-public-game-companies-generated-126-billion-in-2021-as-subscriptions-and-ma-shake-up-the-market>
- Castillo Fadić, M. N. (2020). Corpus Básico del Español de Chile: metodología de procesamiento y análisis. *Lexis*, 44(2), 483-523. Obtenido de <https://doi.org/10.18800/lexis.202002.004>
- Chafale, D., & Pimpalkar, A. (2014). Review on developing corpora for sentiment analysis using plutchik's wheel of emotions with fuzzy logic. *International Journal of Computer Sciences and Engineering (IJCSE)*, 2(10), 14-18.
- Cioffi-Revilla, C. (2014). *Introduction to computational social science*. London y Heidelberg: Springer.
- Clements, M. T., & Ohashi, H. (2005). Indirect network effects and the product cycle: video games in the US, 1994–2002. *The Journal of Industrial Economics*, 53(4), 515-542.
- Feldman, R., & Dagan, I. (1995). Knowledge Discovery in Textual Databases (KDT). *KDD*, 95, págs. 112-117.
- Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1, 215–239.

- Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. *Journal for Language Technology and Computational Linguistics*, 20(1), 19-62.
- Jivani, A. G. (2011). A comparative study of stemming algorithms. *Int. J. Comp. Tech. Appl*, 2(6), 1930-1938.
- Jovanovic, B. (7 de Abril de 2023). *Gamer Demographics: Facts and Stats About the Most Popular Hobby in the World*. Obtenido de Dataprot: <https://dataprot.net/statistics/gamer-demographics/#:~:text=Among%20the%20adult%20crowd%2C%20people,gamers%20are%20older%20than%2055>.
- Leonhardt, M., & Overå, S. (2021). Are There Differences in Video Gaming and Use of Social Media among Boys and Girls?-A Mixed Methods Approach. *International journal of environmental research and public health*, 18(11), 6085. Obtenido de <https://doi.org/10.3390/ijerph18116085>
- Lochbaum, K. E., & Streeter, L. A. (1989). Comparing and combining the effectiveness of latent semantic indexing and the ordinary vector space model for information retrieval. *Information Processing & Management*, 25(6), 665-676.
- Lu, C., Li, X., Nummenmaa, T., Zhang, Z., & Peltonen, J. (2020). Patches and Player Community Perceptions: Analysis of No Man's Sky Steam Reviews. *DiGRA '20- Proceedings of the 2020 DiGRA International Conference*. DiGRA.
- Luo, T., Zhong, C., Ying, X., & Fu, J. (2011). Detecting community structure based on edge betweenness. *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2, págs. 1133-1136. Shanghai.
- McDonald, E. (7 de Diciembre de 2017). *Newzoo and Google Play White Paper: Gender and Inclusivity for Women in Mobile Gaming*. Obtenido de Newzoo: <https://newzoo.com/resources/blog/newzoo-and-google-play-white-paper-women-and-mobile-gaming>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council*, 2, pág. 234. Canada.

- Needham, M., & Hodler, A. E. (2018). *A comprehensive guide to graph algorithms in neo4j*. Neo4j.com.
- Ortiz, L., Tillerias, H., C., C., & Toaza, V. (2020). Impact on the video game industry during the COVID-19 pandemic. *Athenea Engineering sciences journal*, 1(1), 5-13.
- Petrova, E., & Gross, N. (Junio de 2017). *4 reasons people watch gaming content on YouTube*. Obtenido de Think With Google:  
<https://www.thinkwithgoogle.com/marketing-strategies/video/statistics-youtube-gaming-content/>
- Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55, 16-24.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc."
- Putera, T. (20 de Enero de 2022). *LBB: Social Network Analysis*. Obtenido de RPubS:  
[https://rpubs.com/TeraPutera/social\\_network\\_analysis](https://rpubs.com/TeraPutera/social_network_analysis)
- Rai, B. (13 de Octubre de 2021). *Social Network Analysis of Tweets Using R*. Obtenido de GitHub: <https://github.com/bkrai/R-files-from-YouTube/blob/main/Social%20Network%20Analysis%20of%20Tweets%20Using%20R>
- Scott, J. (2011). Social network analysis: developments, advances, and prospects. *Social network analysis and mining*, 1, 21-26.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*.
- Zhao, L., & Min, C. (2019). The Rise of Fashion Informatics: A Case of Data-Mining-Based Social Network Analytics in Fashion. *Clothing and Textiles Research Journal*, 37(2), 87-102.
- Zhao, Y. (2013). Analysing twitter data with text mining and social network analysis. *In 11th Australasian Data Mining Conference (AusDM 2013)*, (págs. 41-47).

## 7. ANEXOS

### 7.1 Código utilizado

Se ha dejado el código ajustado para un solo juego (FIFA 23 en inglés), pero este proceso se ha aplicado para todos los casos observados

#### **7.1.1 Extracción de datos de Twitter**

```
library(rtweet)
```

```
H1<-search_tweets2(c("#fifa23","fifa23"), n=5000, include_rts=FALSE,  
lang="en")
```

Si se alcanza el máximo volumen permitido, se puede incorporar el siguiente parámetro para automatizar que vuelva a extraer datos cuando se pueda:

```
#, retryonratelimit = TRUE
```

Guardar los datos, en este caso para la semana 4:

```
library(writexl)
```

```
write_xlsx(H1, "FIFA23_en_w4.xlsx")
```

Otra función útil, para obtener los datos de un único usuario es:

```
tmls <- get_timeline(c("EASports"), n = 3200)
```

#### **7.1.2 Unión de estructuras de datos**

```
library(writexl)
```

Hay que situarse en una carpeta en la que se encuentren todos los datos que se deseen juntar (los datos de las 4 semanas)

```
rm(list = ls())
```

```
lista<-list.files()
```

```
Total=data.frame()
```

```
for (i in seq_along(lista)) {
```

```
  a=read_excel(lista[i])
```

```
  Total=rbind(Total, a)
```

```
}
```

```
data <- Total[!duplicated(Total$status_id), ]
```

```
write_xlsx(data, "FIFA_ENG.xlsx")
```

### 7.1.3 Limpieza de datos

Para echar un vistazo a los datos:

```
library(readxl)
Tweets <- read_excel("FIFA_ENG.xlsx")
head(Tweets$text, 10) # Vemos qué pinta tienen
View(Tweets)
```

A continuación, se realizan los procesos de limpieza, según lo que se necesite:

```
# Quitamos menciones @
Tweets$text<-gsub("@\\w+", "", Tweets$text)

# Pasamos a minúsculas
Tweets$text<-tolower(Tweets$text)

# Eliminación de URLs del tipo https://t.co/
Tweets$text<- gsub("https://t.co/[A-Za-z\\d]+", "", Tweets$text)

# En Twitter es muy común encontrar la expresión &
Tweets$text<- gsub("&", "", Tweets$text) # eliminamos esos &

# Es la forma que tiene de expresar el & de un texto.

# Ejemplo
# grep("&", Tweets$text)

# Quitamos puntuación, exclamaciones, etc. (!"#%&'()*+,-
./:;<=>?@[\\^_`{|}~)
Tweets$text<-gsub("[[:punct:]]", "", Tweets$text)

# quitar digitos
Tweets$text<-gsub("[[:digit:]]", "", Tweets$text)

#Eliminamos algunos tweets cuyo mensaje está duplicado
Tweets<-Tweets[!duplicated(Tweets$text),]

#Quitar emojis
Tweets$text<-gsub("[^\\x01-\\x7F]", "", Tweets$text) #Para datos en
inglés

Tweets$text<-gsub('\\p{So}|\\p{Cn}', '', Tweets$text, perl = TRUE)
#Para datos en español
```

Para eliminar palabras en concreto se utiliza una función propia:

```
library(stringr)

word_match_list = function(...) {
  words = c(...)

  word_options = paste(words, collapse = "|") # combine the words w/ |
  between them

  paste0("\\b(?:", word_options, ")\\b")
}

word_list_regex = word_match_list("fifa", "fifa23")

Tweets$text<-str_remove_all(Tweets$text, word_list_regex)
```

### 7.1.4 Análisis de frecuencias

```
library(quanteda)
library(quanteda.textstats)
library(quanteda.textplots)
```

Formación del documento:

```
# Formamos un vector con los documentos
docus<-c(Tweets$text)

View(as.data.frame(docus))

# Ahora construimos el corpus
micorpus<-corpus(docus)

summary(micorpus)
```

Antes de la tokenización se pueden repetir procesos de limpiado para asegurar la validez de los datos:

```
mistoks<-tokens(micorpus,remove_numbers = TRUE, remove_punct = TRUE )
View(mistoks)
```

Construcción de la DFM:

```
midfm<-dfm(mistoks) # Construcción de la matriz document-term
dim(midfm)
```

```
featnames(midfm)[1:500] # vemos cuáles son los primeros términos
```

Si hay stopwords:

```
midfm<-dfm_remove(midfm, pattern = stopwords ("english"))  
#midfm<-dfm_remove(midfm, pattern = stopwords ("spanish")) para  
español  
dim(midfm)  
midfm<-dfm_remove(midfm, "fifa", "fifa23")
```

Si se desea hacer stemming:

```
midfm <- dfm_wordstem(midfm, language = "english")  
#midfm2 <- dfm_wordstem(midfm, language = "spanish") para español  
dim(midfm)  
featnames (midfm)
```

Para ver términos más populares y nubes de palabras:

```
topfeatures (midfm,20) # los 20 términos más frecuentes  
topfeatures (midfm[9,],20) # los términos más frecuentes del documento  
9  
set.seed(132)  
textplot_wordcloud(midfm, max_words=150, color = c('orange', 'blue',  
'red'))
```

Obtención de pesos tf-idf:

```
mitfidf<-dfm_tfidf (midfm, scheme_tf="prop", base=2)
```

### 7.1.5 Análisis de sentimiento

```
library (syuzhet)
```

Para ver cómo funcionan los diccionarios:

```
get_sentiment_dictionary('afinn')  
get_sentiment_dictionary('nrc', language = "spanish")  
get_sentiment_dictionary('nrc', language = "english") %>%  
filter (sentiment=="trust")
```

```
get_sentiment_dictionary('bing')
get_sentiment(Tweets$text, method="afinn") #aplicado a los datos
```

#### Calcular e incorporación de polaridades en los datos:

```
Tweets<-Tweets %>% mutate(nrc_polarity= get_sentiment(text,
method="nrc", language="spanish")
average<-mean(Tweets$nrc_polarity)
```

#### Polaridades por Fuente:

```
library(dplyr)
library(stringr)
Tweets_pol_an<- subset(Tweets, source == 'PlayStation@Network' | source
== 'Xbox One Social'|source == "Nintendo Switch Share" | source
=="Twitter Web App")
Tweets_pol_an %>% group_by(source) %>%
  summarise(positivos = 100 * sum(nrc_polarity > 0) / n(),
            neutros = 100 * sum(nrc_polarity == 0) / n(),
            negativos = 100 * sum(nrc_polarity < 0) / n())
# box plot para mostrar la distribución de polaridad según la fuente
ggplot(Tweets_pol_an, aes(x=source, y=nrc_polarity, fill=source)) +
  geom_boxplot()
```

#### Cálculo e incorporación de emociones en los datos

```
emotions<-get_nrc_sentiment(Tweets$text, language="english")
#emotions<-get_nrc_sentiment(Tweets$text, language="spanish") para
español

# Pero vamos a incorporar esas emociones al data frame Tweets
Tweets<-Tweets %>% mutate (anger=emotions$anger,
                           anticipation=emotions$anticipation,
                           disgust=emotions$disgust,
                           fear=emotions$fear,
                           joy=emotions$joy,
                           sadness=emotions$sadnes,
                           surprise=emotions$surprise,
```

```

        trust=emotions$trust,
        negative=emotions$negative,
        positive=emotions$positive
    )

library(writexl)
write_xlsx(Tweets, "FIFA23engCSA.xlsx")

```

#### Identificar tuits con un sentimiento específico:

```

summary(Tweets$fear)
which(Tweets$fear==3)
which.max(Tweets$sadness)
summary(Tweets_pol_an$anger)
which(Tweets_pol_an$anger==3)
Tweets_pol_an$text[33]
Tweets_pol_an$source[2576]

```

#### Graficar emociones:

```

library(RColorBrewer)
barplot(
  sort(colSums(prop.table(Tweets[, 94:101]))),
  horiz = TRUE,
  cex.names = 0.7,
  col = brewer.pal(n = 8, name = "Set3"),
  las = 1,
  main = "Emotions in FIFA 23 English tweets", xlab="Percentage")
Pesos<-
rbind(colSums(prop.table(Tweets[Tweets$source=='PlayStation@Network',
94:101])),
      colSums(prop.table(Tweets[Tweets$source=='Xbox One
Social', 94:101])),
      colSums(prop.table(Tweets[Tweets$source=="Nintendo Switch
Share", 94:101])),
      colSums(prop.table(Tweets[Tweets$source=="Twitter Web
App", 94:101])))

```

```

rownames(Pesos)<-c('PlayStation@Network','Xbox One Social', 'Nintendo
Switch Share',"Twitter Web App")

# "Nintendo Switch Share", "coral2", "green",'Xbox One Social',
"blue", PlayStation@Network

barplot(Pesos, col = c("blue", "green","coral2","darkgrey"),
beside=TRUE, main = "Emotions in FIFA 23 English tweets per channel")

legend("topleft", legend = rownames(Pesos),

      fill = c("blue", "green","coral2","darkgrey"), box.lty = 0, cex
= 0.8)

```

### 7.1.6 Elaboración de redes de términos y clustering

Este código ha sido obtenido y adaptado de (Rai, 2021)

```

library(readxl)
Tweets<- read_excel("FIFA23engCSA.xlsx")

```

Ampliar el uso de memoria de R:

```

memory.limit()
memory.limit(size = 50000)

```

Construir el corpus:

```

library(tm)
corpus <- iconv(Tweets$text, to = "utf-8") #para inglés
corpus<-Tweets$text #para español
corpus[100]
corpus <- Corpus(VectorSource(corpus))

```

Limpieza de texto (ya está limpio, pero así se puede asegurar):

```

# Clean text
corpus <- tm_map(corpus, tolower)
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, removeNumbers)
cleanset <- tm_map(corpus, removeWords, stopwords('english'))
cleanset <- tm_map(corpus, removeWords, stopwords('spanish'))
removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)

```

```
cleanset <- tm_map(cleanset, content_transformer(removeURL))
cleanset <- tm_map(cleanset, removeWords, c("fifa","fifa23"))
cleanset <- tm_map(cleanset, stripWhitespace)
```

#### Construcción de la Term Document Matrix:

```
tdm <- TermDocumentMatrix(cleanset)
tdm <- as.matrix(tdm)
tdmcomp<-tdm
tdm300<-tdm
```

#### Ajustar la tdm a un tamaño manejable e interpretable:

```
tdm <- tdm[rowSums(tdm)>300,]
tdm <- tdm300[rowSums(tdm300)>550,]
tdm[1:10,1:10] #echar un vistazo
```

#### Crear la red de términos:

```
library(igraph)
tdm[tdm>1] <- 1
termM <- tdm %*% t(tdm)
termM[1:10,1:10]
g <- graph.adjacency(termM, weighted = T, mode = 'undirected')
g <- simplify(g)
V(g)$label <- V(g)$name
V(g)$degree <- degree(g)
```

#### Graficar un histograma del grado de los nodos:

```
hist(V(g)$degree,
     breaks = 100,
     col = 'green',
     main = 'Histogram of Node Degree',
     ylab = 'Frequency',
     xlab = 'Degree of Vertices')
```

Primer gráfico de la red:

```
set.seed(222)
plot(g)
plot(g, layout = layout_with_graphopt,
      vertex.color='green',
      vertex.size = 4,
      vertex.label.dist = 1.5
      )
```

Elaboración de clústeres por betweenness :

```
set.seed(222)
comm <- cluster_edge_betweenness(g)
plot(comm, g)
prop <- cluster_label_prop(g)
plot(prop, g)
greed <- cluster_fast_greedy(as.undirected(g))
plot(greed, as.undirected(g), main="Term communities Network for FIFA
23 English tweets")
```

Graficar ‘hubs’ y ‘authorities’: términos que reciben mucha atención por parte de otros (tienen muchas relaciones desde otros) y términos que afectan mucho al resto (tienen muchas relaciones hacia otras):

```
hs <- hub_score(g, weights = NA)$vector
as <- authority_score(g, weights=NA)$vector
par(mfrow=c(1,2))
plot(g, vertex.size=hs*50, main='Hubs',
      vertex.color=rainbow(50))
plot(g, vertex.size=as*30, main='Authorities',
      vertex.color=rainbow(50))
par(mfrow=c(1,1))
```

### 7.1.7 Detección de influencers

```
library(tidyverse)
library(dplyr)
library(ggplot2)
```

```
library(readxl)
```

Cargar los datos y dividirlos en dos subconjuntos, según si se trata de usuarios verificados o no:

```
Tweets<- read_excel("FIFA23engCSA.xlsx")
tweetsver<-subset(Tweets, verified==TRUE)
tweetsunver<-subset(Tweets, verified==FALSE)
```

Para agregar los datos por nombre de usuario, seleccionando variables concretas:

```
influencersdataver<- aggregate(cbind(retweet_count, favorite_count) ~
screen_name, tweetsver, sum)
```

Obtener los seguidores de cada usuario y unirlos a los datos agregados:

```
datos_seguidores_maxver <- tweetsver %>%
  group_by(screen_name) %>%
  summarise(followers_count = max(followers_count)) %>%
  ungroup()
influencersdataver <- influencersdataver %>%
  left_join(datos_seguidores_maxver, by = "screen_name")
```

Ordenar en orden descendiente de retweet\_count y seleccionar las top 20 cuentas:

```
influencersdataver20<-head(influencersdataver,20)
influencersdataver20 <- influencersdataver[order(-
influencersdataver20$retweet_count),]
```

Gráfico de barras de retuits totales:

```
ggplot(influencersdataver20, aes(x = reorder(screen_name, -
retweet_count), y = retweet_count, fill = retweet_count)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = retweet_count), vjust = -0.5) +
  scale_fill_gradient(low = "blue", high = "red") + #blue, red,
goldenrod, steelblue, coral2
  labs(title="Total retweets from top 20 verified accounts for FIFA 23
English tweets", x = "Screen or User Name", y = "Total Retweets") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Gráfico de burbujas:

```
ggplot(influencersdataver20, aes(x = retweet_count, y =
favorite_count, size = followers_count, color = followers_count)) +
  geom_point(alpha = 0.7) +
  scale_size_continuous(range = c(5, 10)) +
  scale_color_gradient(low = "blue", high = "red") +
  labs(title = "Bubble graph for FIFA 23 English tweets from verified
users", x = "Total Retweets", y = "Total Favorites") +
  theme_minimal() +
  geom_text(aes(label = screen_name), vjust = -1) +
  guides(color = "none", size = guide_legend(title = "Seguidores",
title.position = "top"), alpha = FALSE) +
  theme(legend.position = "none",
        plot.margin = unit(c(1, 1, 1, 1), "cm"), # ajustar el margen
general del gráfico
        panel.margin = unit(c(2, 2, 2, 2), "cm"))+# ajustar el margen
de la zona de trazado del gráfico
  scale_x_continuous(expand = c(0.2, 0)) +
  scale_y_continuous(expand = c(0.2, 0))
```

## **7.2 Procesos de interés para trabajos futuros con código**

### **7.2.1 Elaboración de redes de usuarios**

El método más frecuente para realizar redes de usuarios en redes sociales es el de utilizar interacciones que conectan usuarios entre sí. En Twitter esto es particularmente conveniente puesto que existe la funcionalidad de retuit, que consiste en repostear el tuit de una persona, referenciándolo, y normalmente incluyendo algún comentario propio, de forma que la gente que te siga lo pueda ver. El código y el proceso descrito a continuación ha sido extraído de la página de RPub de Tera Putera, la fuente ha sido incluida en la bibliografía como (Putera, 2022).

Para especificar en la query de extracción de datos el incluir retuits:

```
library(tidyverse)
library(tidygraph)
library(ggraph)
```

```

library(igraph)

library(rtweet)

H1<-search_tweets2(c("#fifa23","fifa23"), n=5000, include_rts=TRUE,
lang="en")

```

Echar un vistazo a los datos:

```

tweets <- read.csv("FIFA_ENG.csv")

head(tweets)

table(tweets$is_retweet)

tweets %>%

  group_by(screen_name) %>%

  summarise(total_retweet = sum(retweet_count)) %>%

  arrange(desc(total_retweet)) %>%

  head()

```

Limpieza de datos:

```

tweets %>%

  select(screen_name, mentions_screen_name)

```

Estructurar los datos de forma que haya dos variables; de y desde (“to” y “from”); la persona que es retuiteada y la persona que retuitea:

```

edge_df <-

tweets %>%

  select(screen_name, mentions_screen_name) %>% #Step 1

  separate_rows(mentions_screen_name, sep = " ") %>% #Step 2

  filter(mentions_screen_name != "") %>% #step 3

  rename(from = screen_name,

         to = mentions_screen_name) #Step 4

edge_df

```

Se crea otro objeto “nodes” con información única (se eliminan duplicados):

```

nodes_df <- data.frame(name = unique(c(edge_df$from, edge_df$to)),

                      stringsAsFactors = F)

```

Preparar los datos que se van a graficar:

```
graph_tweets <- tbl_graph(nodes = nodes_df,
                          edges = edge_df,
                          directed = F)
```

Incorporar medidas de centralidad en los datos:

```
graph_tweets <- graph_tweets %>%
  activate(nodes) %>%
  mutate(degree = centrality_degree(), # Degree centrality
         between = centrality_betweenness(normalized = T), #Betweenness
         closeness = centrality_closeness(), # Closeness
         eigen = centrality_eigen() # Eigen centrality
  )
```

Convertir datos del gráfico a data frame:

```
network_act_df <- graph_tweets %>%
  activate(nodes) %>%
  as.data.frame()
```

Identificar influencers por cada medida de centralidad:

```
pop_username <- data.frame(
  network_act_df %>% arrange(-degree) %>% select(name) %>% head(),
  network_act_df %>% arrange(-between) %>% select(name) %>% head(),
  network_act_df %>% arrange(-closeness) %>% select(name) %>% head(),
  network_act_df %>% arrange(-eigen) %>% select(name) %>% head()
) %>% setNames(c("Degree", "Betweenness", "Closeness", "Eigen"))
pop_username
```

Para ver contenido de los tuits de autores de interés:

```
tweets %>%
  filter(mentions_screen_name == "DonkTrading") %>%
  group_by(mentions_screen_name, text) %>%
```

```

tally() %>%
  arrange(-n) %>%
  pull(text) %>%
  head(4)

```

Realizar el gráfico, identificando clústeres con el método Louvain:

```

set.seed(123)
graph_tweets <- graph_tweets %>%
  activate(nodes) %>%
  mutate(community = group_louvain()) %>% # clustering
  activate(edges) %>%
  filter(!edge_is_loop())
graph_tweets %>%
  activate(nodes) %>%
  as.data.frame() %>%
  count(community)

```

Para asignar usuarios relevantes a cada clúster se desarrollará una función:

```

important_user <- function(data) {
  name_person <- data %>%
  as.data.frame() %>%
  filter(community %in% 1:5) %>%
  select(-community) %>%
  pivot_longer(-name, names_to = "measures", values_to = "values") %>%
  group_by(measures) %>%
  arrange(desc(values)) %>%
  slice(1:6) %>%
  ungroup() %>%
  distinct(name) %>%
  pull(name)
  return(name_person)
}

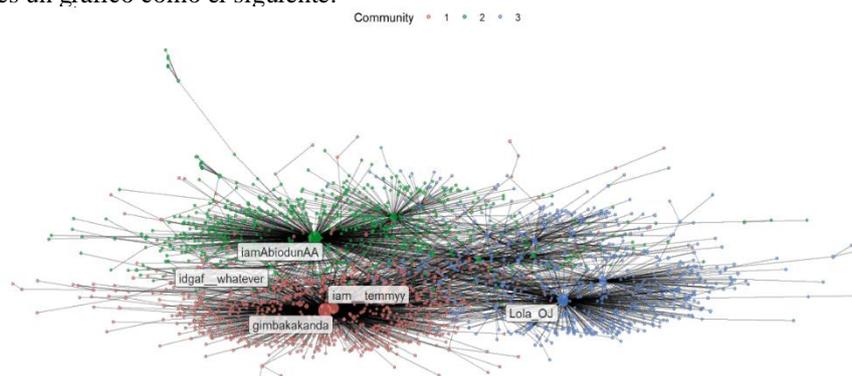
```

```
important_person <-
graph_tweets %>%
  activate(nodes) %>%
  important_user()
```

Realizar el gráfico de redes:

```
set.seed(123)
graph_tweets %>%
  activate(nodes) %>%
  mutate(ids = row_number(),
         community = as.character(community)) %>%
  filter(community %in% 1:3) %>% # number of community.
  arrange(community,ids) %>%
  mutate(node_label = ifelse(name %in% important_person, name,NA)) %>%
  gggraph(layout = "fr") +
  geom_edge_link(alpha = 0.3 ) +
  geom_node_point(aes(size = degree, fill = community), shape = 21,
alpha = 0.7, color = "grey30") +
  geom_node_label(aes(label = node_label), repel = T, alpha = 0.8 ) +
  guides(size = "none") +
  labs(title = "Top 3 Community of #SecureTheTribe",
       color = "Interaction",
       fill = "Community") +
  theme_void() +
  theme(legend.position = "top")
```

El resultado es un gráfico como el siguiente:



Fuente: extraído de (Putera, 2022)

### **7.3 Polaridades medias por plataforma durante el mes de lanzamiento**

<b>PLATAFORMA</b>	<b>POSITIVOS (%)</b>	<b>NEUTROS (%)</b>	<b>NEGATIVOS (%)</b>
<b>FIFA 23 en inglés</b>			
Web App	38,7%	43,5%	17,9%
PlayStation	20,3%	64,7%	15%
Xbox	31,5%	59,3%	9,26%
Nintendo Switch	60%	33,3%	6,67%
<b>FIFA 23 en español</b>			
Web App	31,2%	51,4%	17,4%
PlayStation	12%	73,6%	14,4%
Xbox	31,2%	56,2%	12,5%
Nintendo Switch	0%	100%	0%
<b>Call of Duty: Modern Warfare 2 en inglés</b>			
Web App	28,3%	46,5%	25,3%
PlayStation	11%	76,2%	12,8%
Xbox	14,6%	65,9%	19,5%
<b>Call of Duty: Modern Warfare 2 en español</b>			
Web App	26,8%	49,8%	23,5%
PlayStation	4,35%	92,7%	2,95%
Xbox	0%	100%	0%
<b>God of War: Ragnarök en inglés</b>			
Web App	34,2%	43,3%	22,5%
PlayStation	17,6%	68,7%	13,7%
<b>God of War: Ragnarök en español</b>			
Web App	29,8%	43,3%	26,9%
PlayStation	16,7% %	67,2%	16,1%
<b>Pokémon Escarlata y Púrpura en inglés</b>			
Web App	18,5%	71,7%	9,8%
Nintendo Switch	25,6%	59,2%	15,2%
<b>Pokémon Escarlata y Púrpura en español</b>			
Web App	30,1%	45,9%	24%
Nintendo Switch	16,8%	67,2%	16%

Esta tabla se menciona en el **apartado 3.4.2**