



FACULTAD DE TURISMO Y FINANZAS

GRADO EN TURISMO

Certificado COVID Digital de la UE: Un Análisis de Sentimientos y su repercusión en el Turismo.

Trabajo Fin de Grado presentado por Adrián García Pozo, siendo el tutor del mismo el profesor Pedro Ramiro Palos Sánchez.

Vº Bº del Tutor:

Alumno/a:

D. Pedro Ramiro Palos Sánchez

D. Adrián García Pozo

Sevilla. Diciembre de 2021



**GRADO EN TURISMO
FACULTAD DE TURISMO Y FINANZAS**

TRABAJO FIN DE GRADO

CURSO ACADÉMICO [2020-2021]

TÍTULO:

CERTIFICADO COVID DIGITAL DE LA UE: UN ANÁLISIS DE SENTIMIENTOS Y SU REPERCUSIÓN EN EL TURISMO

AUTOR:

ADRIÁN GARCÍA POZO

TUTOR:

D. PEDRO RAMIRO PALOS SÁNCHEZ

DEPARTAMENTO:

ECONOMÍA FINANCIERA Y DIRECCIÓN DE OPERACIONES

ÁREA DE CONOCIMIENTO:

ORGANIZACIÓN DE EMPRESAS

RESUMEN:

El presente trabajo es un estudio de investigación, que tiene el objetivo de conocer cómo valora la población la implantación del Certificado COVID Digital de la UE a través de la técnica de Análisis de sentimientos. Para llevarlo a cabo, se ha explicado en qué consiste tanto el Certificado mencionado como la técnica a utilizar, y luego se ha expuesto la ejecución del procedimiento realizándose una recolección de datos a través de la herramienta informática denominada Socioviz en la plataforma social Twitter, siendo el patrón de búsqueda utilizado #EUCovidCertificate. Con los datos obtenidos en Socioviz se ha llevado a cabo la detección de la polaridad y de la intensidad de sentimiento con la herramienta Mozdeh y, además, se ha elaborado un grafo utilizando las herramientas que ofrece Gephi.

PALABRAS CLAVE:

Certificado Covid; análisis de sentimientos; twitter; turismo.

ÍNDICE

CAPÍTULO 1: PRESENTACIÓN DEL TRABAJO	1
1.1 INTRODUCCIÓN.....	1
1.2 OBJETIVOS.....	1
1.3 JUSTIFICACIÓN DEL TEMA.....	2
1.4 METODOLOGÍA	3
CAPÍTULO 2: CERTIFICADO COVID DIGITAL DE LA UE	5
CAPÍTULO 3: ANÁLISIS DE SENTIMIENTOS	7
3.1 DEFINICIÓN.....	7
3.2 EL ENFOQUE DE INTERNET HACIA EL BIG DATA	7
3.3 PRIMEROS TRABAJOS E INVESTIGADORES SOBRE ANÁLISIS DE SENTIMIENTOS	9
3.4 TÉCNICAS DEL ANÁLISIS DE SENTIMIENTOS	9
3.4.1 Técnica de aprendizaje automático o machine learning approach.	9
3.4.2 Técnica basada en diccionarios o Lexicon-based approach.....	10
3.4.3 Enfoque semántico.....	10
3.5 TAREAS QUE REALIZA EL ANÁLISIS DE SENTIMIENTOS	11
3.6 ETAPAS PARA REALIZAR UN ANÁLISIS DE SENTIMIENTOS	11
3.7 APLICACIONES DEL ANÁLISIS DE SENTIMIENTOS	12
3.8 ANÁLISIS DE REDES	13
3.8.1 Definición.....	13
3.8.2 Estructura red.....	13
3.8.3 Características y propiedades de la red	16
3.8.3.1 Indicadores de estructura	16
3.8.3.2 Indicadores de centralidad.....	17
CAPÍTULO 4: EJECUCIÓN DEL ANÁLISIS DE SENTIMIENTOS.....	19
4.1 RECOLECCIÓN DE DATOS CON SOCIOVIZ	19
4.1.1 Definición de Socioviz	19
4.1.2 ¿Cómo funciona?.....	19
4.1.3 Análisis de resultados de Socioviz.....	21
4.2 ANÁLISIS DE NUESTRA RED CON GEPHI	29
4.2.1 Definición de Gephi.....	29
4.2.2 ¿Cómo funciona? Construcción y análisis de nuestra red paso a paso	
29	
4.2.2.1 Vista general.....	29
4.2.2.2 Laboratorio de datos	39
4.2.2.3 Previsualización	40

4.3 MOZDEH	43
4.3.1 Definición	43
4.3.2 ¿Cómo funciona?	43
4.3.3 Análisis de resultados con Mozdeh	45
CAPÍTULO 5: CONCLUSIONES	49
Limitaciones del estudio	52
Estudios futuros	53
Glosario de términos	54
Bibliografía	57

1

ÍNDICE DE ILUSTRACIONES

Figura 3.1. Grafo de ejemplo	14
Figura 3.2. Multitudes polarizadas	15
Figura 3.3. Multitudes unificadas	15
Figura 3.4. Clústeres fragmentados.....	15
Figura 3.5. Conglomerados múltiples.....	15
Figura 3.6. Red de amplificación.....	16
Figura 3.7. Red de apoyo	16
Figura 3.8. Nodo con centralidad por grado	17
Figura 3.9. Grupos de nodos más roles de intermediación	18
Figura 4.1. Filtros de búsqueda en Socioviz	19
Figura 4.2. Resúmenes de búsqueda en Socioviz	20
Figura 4.3. Resumen de búsqueda de #EUCovidCertificate en Socioviz	21
Figura 4.4. Captura de Gráfica time-line (tuits).....	22
Figura 4.5. Captura de pantalla de gráfica time-line (retuits).....	23
Figura 4.6. Captura de listado de tuits y retuits	24
Figura 4.7. Grafo de usuarios	26
Figura 4.8. Grafo Top 10 palabras más usadas	27
Figura 4.9. Captura I de la herramienta Gephi.	30
Figura 4.10. Captura II de la herramienta Gephi	31
Figura 4.11. Gráfica de grado medio.....	32
Figura 4.12. Gráfica de excentricidades.....	33
Figura 4.13. Gráfica de Closeness Centrality	34
Figura 4.14. Gráfica de valores de intermediación.	35
Figura 4.15. Gráfica de grupos o comunidades.....	36
Figura 4.16. Gráfica de centralidad de vector propio.....	37
Figura 4.17. Captura III de la herramienta Gephi	38
Figura 4.18. Captura IV de la herramienta Gephi.....	38
Figura 4.19. Captura de la tabla de datos de Gephi (nodos).....	39
Figura 4.20. Captura de la tabla de datos (aristas) de Gephi	40
Figura 4.21. Red acabada	41

Figura 4.22. Red de nodos más importantes según valor de grado.	42
Figura 4.23. Valor de intermediación de #europe.....	43
Figura 4.24. Pestaña de búsqueda de Mozdeh.....	44
Figura 4.25. Filtro de búsqueda por frecuencia de sentimiento de Mozdeh.....	44
Figura 4.26. Biblioteca de tuits de Mozdeh	45
Figura 4.27. Captura de gráfica de la puntuación conjunta de los tuits	47

ÍNDICE DE TABLAS

Tabla 4.1. Usuarios más activos	24
Tabla 4.2. Usuarios más influyentes	25
Tabla 4.3. Palabras más usadas.....	26
Tabla 4.4. Hashtags más importantes Gephi	28
Tabla 4.5. % de tuits negativos y positivos en cada intensidad	45

ÍNDICE DE GRÁFICAS

Gráfica 4.1. % de tuits negativos y positivos en cada intensidad	46
--	----

CAPÍTULO 1: PRESENTACIÓN DEL TRABAJO

1.1 INTRODUCCIÓN

Con la problemática pasada y presente que ha generado y sigue generando el Covid-19 se han visto perjudicados prácticamente todos los sectores de la sociedad. En este contexto podemos decir que, el sector turístico ha sido uno de los sectores económicos más castigados por las obligadas y necesarias restricciones ejecutadas por los gobiernos que se han vivido y se siguen viviendo hoy día con el objetivo de conseguir que el impacto negativo sanitario y social sea el mínimo posible.

Una de estas consecuencias, derivada por el principio básico de no transmisión y de distanciamiento social, ha sido la dificultad para realizar viajes turísticos, por una parte, por problemas estructurales y de logística obvios, pero sobre todo por las duras restricciones de movilidad que se fueron imponiendo en casi la totalidad de los países a raíz de la pandemia.

Con el paso de los meses y gracias al avance en cuanto a la detección de positivos por Covid-19 a través de pruebas de antígenos y PCR (Reacción en Cadena de la Polimerasa, o en inglés; Polymerase chain reaction), entre otras, pero sobre todo gracias al desarrollo y la distribución de las efectivas vacunas, los gobiernos de algunos países han podido relajar restricciones de movilidad y han tenido la posibilidad, al fin, de desarrollar medidas que afectan positivamente al sector económico y a la libre circulación de las personas.

Una de estas novedosas medidas es el denominado Pasaporte Covid, o Certificado Covid Digital para la Unión Europea. Se trata de una medida que tiene el objetivo de agilizar los tiempos de espera en los aeropuertos y de unificar requisitos burocráticos entre países a través de un certificado que autoriza a la persona que cumpla los requerimientos de poseerla poder viajar a otros países con un menor tiempo de espera y sin la necesidad de realizarse otro test adicional para entrar a otro país europeo.

Llegados a este punto, hemos tenido en cuenta que, esta implantación del Pasaporte Covid ha podido percibirse como una medida discriminatoria entre un sector de la población, ya que existen países y personas (junio de 2021) que por diferentes motivos tienen una menor capacidad para acceder a las vacunas y a los test de detección del virus Covid-19 o que directamente no pueden actualmente o simplemente no quieren, por lo que este trabajo va a estar enfocado en averiguar qué opina la población en general sobre la implantación de este certificado que, al fin y al cabo, da una serie de privilegios a todo aquel que lo posee.

Para ello, se ha realizado un estudio de Análisis de Sentimientos en la red social Twitter, tomando como referencia los tweets publicados por diferentes usuarios que han usado el hashtag "#EUCovidCertificate".

1.2 OBJETIVOS

- Objetivos principales:

- Definir qué es el Certificado COVID digital para la UE y qué implica su implantación.

- Descubrir, en base a un análisis de los resultados obtenidos, qué piensa la población sobre la implantación de esta medida.

- Objetivos secundarios:

- Explicar la importancia que ha adquirido el análisis de macrodatos para cualquier institución con afán de conocer el mercado y su propio contexto, haciendo un recorrido desde la aparición de la Web 2.0 hasta la técnica de Análisis de Sentimientos que conocemos hoy día.
- Conocer en qué consiste la técnica de Análisis de sentimientos.
- Descubrir qué son las redes o grafos y de qué están compuestas.
- Realizar un estudio en el que se explican las herramientas utilizadas.

1.3 JUSTIFICACIÓN DEL TEMA

El Turismo se ha convertido en los últimos años en uno de los principales sectores económicos de nuestra era, tal es así que, en España, sin ir más lejos, llegó a convertirse en el sector económico que más aportación realizaba al PIB nacional antes de la pandemia, alcanzando un 15% en agosto de 2018 y cerrando el ejercicio de 2019 (último año sin pandemia) con un 12,4% de aportación al PIB español significando un valor de 154.487 millones de euros y concentrando el 12,9% del empleo (2,72 millones de puestos) según datos de la Cuenta Satélite del Turismo, publicados en el Instituto Nacional de Estadística (INE) (Estadística, 2021).

Pero con el inicio de la pandemia, esos datos tan positivos que hacían del sector turístico un sector clave de nuestra economía, además de muy próspero, se desplomaron en 2020 pasando, por ejemplo, de haber un gasto por parte de los turistas que visitaban nuestro país de 91.912 millones de euros en 2019 a 19.740 millones de euros en 2020.

Viendo la importancia vital que tenía este sector económico en la época de prepandemia para nuestro país, se antoja clave intentar recuperarlo en la medida de lo posible. ¿Pero cómo? El sector turístico poco puede hacer por la mejora de la salud pública, pero dadas las circunstancias, debe tratar de adaptarse, y es aquí donde encontramos una de las medidas más esperanzadoras que se han ejecutado para relanzar en la medida de lo posible al sector turístico; Certificado COVID Digital de la UE.

Presuponiendo la importancia que debe tener el lanzamiento de este novedoso certificado para la mejora del sector, hemos decidido estudiar la aceptación que tiene dicha medida entre la población, ya que, al fin y al cabo, de la población depende, a través de sus viajes que el turismo reflote.

Este estudio, como hemos comentado, lo vamos a realizar a través de un Análisis de Sentimientos, adquiriendo los datos de la red social Twitter, ya que existen evidencias que emiten que la opinión de los usuarios en las diferentes redes sociales de Internet tiene un impacto real e importante para el sector turístico en términos de ventas y posicionamiento en el mercado, sobre todo en el subsector hotelero. Además, según apunta una encuesta realizada por el "Observatorio de las Redes Sociales" el 85% de los encuestados confiesa que confía en los comentarios aportados por los usuarios en las diferentes redes sociales y que perciben esas opiniones o comentarios como más

objetivos que los aportados por la propia empresa en cuestión, por lo que, gran parte de la población que contestó que les parecía confiables los comentarios de otros usuarios confiarán más en esos usuarios que en la propia empresa turística que vende.

Por lo tanto, con lo expuesto, debe de haber una interesante cantidad de opiniones e información a analizar sobre el turismo y su relación con el Pasaporte COVID en la red Internet.

1.4 METODOLOGÍA

En los capítulos 2 y 3 se ha realizado una revisión bibliográfica obteniéndose la información de diferentes medios y plataformas. En el capítulo 2: Certificado COVID digital de la UE, se ha realizado una investigación de fuentes externas y secundarias, basándose en un seguimiento de las noticias publicadas en los distintos medios digitales como periódicos o revistas tales como El Diario o La Vanguardia y estando al corriente de las nuevas medidas publicadas por el Parlamento Europeo a través de sus medios oficiales como el Centro Multimedia del Parlamento Europeo sobre el Certificado COVID digital de la UE.

Para la elaboración del capítulo 3 se ha recurrido en gran parte al catálogo Fama de la biblioteca de la Universidad de Sevilla, la cual da acceso a bases de datos que han sido primordiales para la elaboración de este trabajo como son Web of Science o Scopus en mayor medida. Entre las principales obras de referencia en este capítulo, cabe destacar el trabajo de (Cordón García, 2014) para la comprensión de las redes y sus características, especialmente, y el trabajo académico de (Sobrino Sande, 2018) para la comprensión del Análisis de sentimientos.

Para la realización del capítulo 4 (Ejecución del Análisis de Sentimientos) se han utilizado las herramientas Socioviz, Gephi y Mozdeh, cada una cumpliendo una función de acuerdo con el contenido y con los objetivos del trabajo. La información que ha sido analizada ha sido recolectada de Twitter.

CAPÍTULO 2: CERTIFICADO COVID DIGITAL DE LA UE

El Reglamento sobre el certificado COVID digital de la UE ha entrado en vigor el 1 de julio de 2021. En los próximos puntos vamos a desglosar qué es el certificado COVID Digital de la UE y cuáles son sus utilidades e implicaciones.

Se trata de un certificado que puede portarse tanto en formato digital como en papel, tiene un código QR que sirve de autenticación, es gratuito y es expedido en la lengua nacional del poseedor, además de en inglés y será válido en todos los países de la UE.

Un ciudadano podrá portar esta acreditación si cumple alguno de estos tres requisitos:

- Ha sido vacunado contra la COVID-19 con la pauta completa con una de las vacunas autorizadas por la Agencia Europea de Medicamentos (EMA).
- Se ha realizado una prueba cuyo resultado ha sido negativo en las últimas 72 horas.
- Se ha recuperado de la COVID-19 en los últimos 6 meses y muestra anticuerpos.

A diferencia de lo que pueda pensarse, el Certificado COVID digital de la UE no es un requisito indispensable para viajar dentro de los estados miembros de la UE, ya que viajar se trata de un derecho fundamental en la UE. El propósito de este certificado es facilitar la libre circulación dentro de la UE, pero ¿Cómo? Al viajar, el poseedor del Certificado COVID Digital de la UE queda, en principio, exento de las restricciones a la libre circulación. Esto quiere decir que los países receptores no podrán imponer restricciones adicionales a las personas titulares de este certificado, salvo para salvaguardar la salud pública por causas mayores, como, por ejemplo, la aparición de nuevas variantes.

Aunque cabe la posibilidad de que la exigencia de portar este certificado o uno similar por parte de los ciudadanos no sólo quede para el ámbito de viajar a otros países, ya que, cada vez más países están pidiendo una prueba o certificado similar para poder asistir a grandes eventos como conciertos, para la entrada a museos y espectáculos o incluso para comer en interiores de restaurantes y bares. De hecho, a principios de julio de 2021 la Comisión Europea propuso una serie de recomendaciones a los países para realizar una reapertura segura de los sectores cultural y creativo, proponiendo la posibilidad de que los asistentes tuvieran que mostrar una prueba negativa, una prueba de vacunación, o haber pasado la enfermedad, es decir, uno de los tres requisitos necesarios para obtener el Certificado COVID digital de la UE.

La información sobre la normativa vigente respecto al Certificado COVID digital de la UE se encuentra actualizada en la sección Reabrir la UE dentro de la página web oficial de la Unión Europea. (Europa U. , 2021)

CAPÍTULO 3: ANÁLISIS DE SENTIMIENTOS

3.1 DEFINICIÓN

La definición más aceptada por la comunidad de investigadores es la que ofrecieron los investigadores (Pang & Lee, 2008); “Tratamiento computacional de opiniones, sentimientos y subjetividad en textos”. Pero existen otras definiciones que concretan más el significado de lo que es el análisis de sentimientos y las diferentes partes que lo componen, como la definición de (Cambria & Amir, 2012).; “Conjunto de técnicas computacionales para la extracción, clasificación, comprensión y evaluación de opiniones expresadas en fuentes publicadas en Internet, comentarios en portales web y en otros contenidos generados por usuarios”.

El Análisis del Sentimiento es una forma de analizar las redes sociales. Aplica una serie de técnicas del procesamiento del lenguaje natural y de la lingüística computacional, como la minería de datos para extraer grandes cantidades de datos y procesarlos, o el machine learning para evaluarlos, y también la minería de textos, cuyo objetivo es analizar una información subjetiva para extraer un valor tangible, como puede ser “positivo o negativo” (en el contexto turístico suele dividirse entre “satisfecho e insatisfecho”) a partir de los comentarios de usuarios en redes sociales. En el caso específico del turismo, esas reseñas sacadas de las diferentes redes sociales plasman la satisfacción/insatisfacción, felicidad, frustración o decepción de los clientes con los productos y servicios turísticos.

Algunas de las muchas ventajas que tiene la técnica de análisis de sentimientos son:

- Puede realizar una investigación aplicando un análisis mixto sobre datos cualitativos y cuantitativos. Este tipo de análisis mixto es el más empleado en investigaciones realizadas en redes sociales y en investigaciones turísticas en concreto.
- Además, el análisis de sentimientos asistido por ordenador tiene una serie de ventajas únicas si lo comparamos con los métodos tradicionales de recolección y análisis de datos, como encuestas o grupos de debate, ya que gracias a su automatización puede procesar datos que por volumen, variedad y velocidad serían casi imposible o muy poco eficiente de evaluar por medios humanos, por lo que se trata de una técnica que atenúa el costo, el tiempo y la mano de obra de una manera considerable, y tiene una mayor capacidad de superar los análisis de codificación de contenidos en términos de eficiencia y fiabilidad de los resultados. Es por esto, que, los sistemas de análisis de sentimientos se están usando en casi todos los sectores comerciales y sociales actualmente.

En los siguientes párrafos vamos a desglosar los diferentes elementos que componen el análisis de sentimientos, las fases en las que se divide, así como la manera en la que funciona y sus aplicaciones más importantes.

3.2 EL ENFOQUE DE INTERNET HACIA EL BIG DATA

Para conocer la importancia que ha ocupado el análisis de sentimientos y del del Big data en la actualidad, debemos dar primero un repaso a la evolución que se ha dado en internet, teniendo a los usuarios como protagonistas, desde la entrada de este presente siglo.

En la primera década del siglo XXI, con la aparición de plataformas sociales como Facebook en el año 2004 o Youtube en el año 2005, el concepto de Internet tomó una nueva dimensión. Esa nueva dimensión se conceptuó en 2003 con el término Web 2.0 que hacía una distinción entre una anterior época de la web donde el usuario era básicamente un sujeto pasivo, pues sólo recibía información y no podía apenas interactuar con la propia web, de la época actual en la que el usuario puede interactuar con la web, así como con otros usuarios aportando e intercambiando información.

Este contenido aportado o depositado por los usuarios en Internet puede encontrarse en formato de texto, imagen, vídeo o audio (podcasts), y tiene la trascendental peculiaridad de que puede permanecer en las redes durante un largo periodo de tiempo.

Esto comenzó a desencadenar una transformación en el papel que tenían los usuarios dentro de la red, ya que adquirieron la capacidad de influenciar sobre otros usuarios con la exposición de sus opiniones en páginas web, por ello se empezó a dejar atrás el concepto clásico de boca-oreja donde el posible poder de influencia de los sujetos se reducía a su círculo social, a dar paso al denominado boca-oreja-electrónico, donde el poder de difusión y por lo tanto de influencia era y es mucho mayor, ya que la opinión aportada por el usuario puede ser vista por prácticamente cualquier persona del mundo, y además, esa información, como se ha comentado anteriormente, permanece durante un gran periodo de tiempo en las redes.

El libre acceso a la red para cualquier persona que poseyera un dispositivo capaz de conectarse a Internet y la libertad de poder publicar su opinión en las diferentes plataformas sociales hizo que poco a poco se fuera acumulando en las redes una ingente cantidad de información que empezó a denominarse “Big Data” o macrodatos por su enorme tamaño y porque era prácticamente imposible de analizar por un humano.

Esta desmedida cantidad de datos que se encuentra en las redes hizo que las empresas llevaran a cabo una transformación en cuanto al valor que le daban a los diferentes elementos que componen su negocio, puesto que a finales del siglo XX, el sector turístico, como era lógico, sólo centraba sus esfuerzos, o sus inversiones en las oportunidades para las ventas, los productos y los servicios turísticos, elementos que, hoy en día, siguen siendo de vital importancia, pero a estos elementos se les unió uno más con el comienzo de la Web 2.0 del que dependen todos los mencionados anteriormente; el Análisis del Big Data.

Teniendo claro en este punto la gran influencia que pueden tener los usuarios con sus aportaciones en las diferentes plataformas sociales que conforman la denominada Web 2.0, además de, haber una mayor facilidad para acceder a grandes cantidades de datos de libre acceso o de propiedad, es lógico pensar que una empresa que no quiera estancarse y quiera mejorar y sobre todo saber qué es lo que demandan sus clientes y potenciales clientes deba utilizar esa información generada por el usuario denominada en inglés como “User Generated Content” (UGC) en su propio beneficio. Aún más si hablamos de la industria del turismo y los viajes, ya que está caracterizada por tener una naturaleza dinámica y volátil, y por vivir en un entorno muy competitivo. Además, el turismo ha sido reconocido como el sector número uno en términos de participación en línea. Este entorno y estas características exigen de una mayor competencia y preparación en análisis de marketing, es por todo esto que, en la última

década, el análisis de datos a través del UGC se ha convertido en un asunto vital entre los estudiosos del turismo y la hostelería.

Para darle uso a esa UGC es imprescindible recabar toda la información posible que nos pueda ser útil para la tarea, estudio o investigación que queramos llevar a cabo como empresa, bien para saber con una perspectiva más amplia si nuestra empresa funciona correctamente, además de saber la imagen que proyectamos como marca de cara al exterior, pero sobre todo para poder identificar problemas, posibles anomalías o tendencias futuras con el objetivo de poder adelantarnos a esos supuestos contextos o situaciones y estar adecuadamente preparados para lo que va a ocurrir.

3.3 PRIMEROS TRABAJOS E INVESTIGADORES SOBRE ANÁLISIS DE SENTIMIENTOS

Primero cabe decir que existe tal cantidad de trabajos relacionados con el tema que existen estudios cuyo principal objetivo es recopilar, analizar y mostrar las publicaciones más importantes. Dicho esto, los primeros trabajos relacionados con el análisis de sentimientos de los que se tiene constancia y que dieron el primer paso para configurar las diferentes partes y variantes que lo conforman hoy se muestran en los siguientes puntos:

- Pan et al., 2002: Se trata del primer trabajo conocido que usó algoritmos de aprendizaje automático con el fin de clasificar textos, en su caso los usó con críticas de películas extraídas de un sitio Web.
- P.D. Turney, 2002: Mostró un sistema capaz de clasificar opiniones de usuarios sobre diversos productos a partir de un análisis gramatical de las oraciones y una serie de consultas en el motor de búsqueda AltaVista.
- Alexander Pak y Patrick Paroubek, 2010: Presentó un sistema que crea automáticamente un corpus para entrenar los algoritmos de aprendizaje automático con el objetivo de clasificar mensajes de Twitter.
- Maite Taboada, Julian Brooke, Milán Tofiloski, Kimberly D. Voll y Manfred Stede 2011: Hizo uso de diccionarios con palabras etiquetadas en diferentes polaridades para la clasificación de textos

3.4 TÉCNICAS DEL ANÁLISIS DE SENTIMIENTOS

Encontramos que existen dos enfoques principalmente al aplicar el Análisis de Sentimientos, las técnicas que se basan en aprendizaje automático (machine learning approach) y las que se basan en diccionarios (lexicon-based approach).

3.4.1 Técnica de aprendizaje automático o machine learning approach.

El Machine Learning es una rama de la inteligencia artificial que empezó a adquirir importancia a partir de la década de 1980. Consta de un conjunto de algoritmos caracterizados por tener un alto componente estadístico y algebraico, por lo que cuentan con una gran capacidad de cálculo. Estos algoritmos están programados para estar constantemente aprendiendo patrones y aplicando ese aprendizaje. Este tipo de técnica funciona de la siguiente manera:

1. Primero analiza la información de las entradas o inputs que recibe y extrae de esos inputs unos patrones de comportamientos sintácticos.
2. En base a esos patrones aprende y asimila una información generando un output.

3. Finalmente, con esa información aprendida y asimilada o con esos outputs generados realiza la evaluación de las nuevas entradas.

Esta técnica está dividida en dos variantes o en dos procedimientos con pequeñas discrepancias:

- Técnica de aprendizaje automático supervisado: los algoritmos de aprendizaje automático supervisado parten con un conocimiento o “corpus” manualmente clasificado con el cual el algoritmo lleva a cabo el proceso de encontrar los mejores parámetros para sí mismo y evaluar el nivel de fiabilidad con esos parámetros. Es curioso, porque la técnica de aprendizaje automático supervisado no ha sido muy utilizada en estudios relacionados con el turismo, sin embargo, suele mostrar un resultado mayor a diferencia de otras técnicas. Además, cabe decir que la técnica de aprendizaje automático supervisado es la más usada en la red social Twitter.
- Técnica de aprendizaje automático no supervisado: Los algoritmos de aprendizaje automático no supervisado realizan su función o procesamiento en base únicamente a las entradas que les llegan, y se van programando a medida que analiza las observaciones que va recibiendo.

3.4.2 Técnica basada en diccionarios o Lexicon-based approach

Esta técnica está compuesta de algoritmos más sencillos que los empleados en la técnica basada en el aprendizaje automático. Realiza su trabajo apoyándose en una serie de diccionarios confeccionados, estos diccionarios están compuestos por un listado de palabras y multipalabras, teniendo cada una de ellas una orientación de sentimiento y una intensidad asignada. (Una multipalabra es un conjunto de palabras que emiten un significado unitario).

Para llevar a cabo su trabajo, el diccionario evalúa las palabras o multipalabras que se pretenden estudiar tomando como referencia las palabras y multipalabras guardadas en el propio diccionario. Es decir, si por ejemplo en el diccionario está incluida la palabra “decepción”, la cual obviamente estaría asociada a un sentimiento negativo, en el momento en el que se quiera evaluar un texto que contenga esa palabra, como por ejemplo un tweet, esta herramienta nos la valorará como un sentimiento negativo, asignándole la puntuación de intensidad que tenga establecida.

3.4.3 Enfoque semántico

Esta técnica basada en diccionarios se mejoró con la incorporación de métodos de análisis basados en la semántica, los cuales tienen la capacidad de resolver las ambigüedades que pueda presentar el lenguaje, como, por ejemplo, una palabra o expresión puede tener diferentes significados dependiendo de las circunstancias en las que se usen, este hito consiguió que se añadiera la facultad de detectar de manera mucho más precisa las diferentes polaridades en segmentos de texto.

Se pueden mencionar dos tipos de análisis de sentimientos semánticos; el análisis de sentimientos semántico contextual y el conceptual.

Así que, continuando con el ejemplo descrito en el apartado anterior, ahora no sólo tomaría como referencia la palabra “decepción”, sino que también podrá desempeñar una evaluación global del texto (siempre que las palabras y multipalabras estén guardadas en el diccionario) para determinar la polaridad e intensidad de ese segmento de texto.

Estos diccionarios de palabras y multipalabras asociadas a diferentes sentimientos podemos encontrarlos en aplicaciones o páginas web tanto para búsquedas de palabras en inglés como SentiWordNet, Wordnet, Framenet o Roget's Thesaurus o para búsquedas de palabras en español como el diccionario Elhuyar o el diccionario CRiSOL.

Muchos investigadores realizan sus trabajos a partir de técnicas basadas en diccionarios construyendo un nuevo diccionario a partir de las combinaciones de algunos de los mencionados anteriormente para conseguir unos resultados más amplios y completos

También existe el enfoque híbrido, el cual consiste en realizar la técnica basada en diccionarios y la basada en aprendizaje automático de forma paralela y combinar los resultados de polaridades de ambos.

3.5 TAREAS QUE REALIZA EL ANÁLISIS DE SENTIMIENTOS

- Detección de la polaridad: básicamente lograr averiguar si una opinión es negativa o positiva, o en algunos casos, también se puede utilizar la variante neutra.
- Detección de la intensidad: Se trata de aplicar un valor numérico según la intensidad del sentimiento, dividiéndose en intensidad más o menos negativa o intensidad más o menos positiva.
- Clasificar el texto según el tipo de emoción que exprese, como, por ejemplo; ira, felicidad, tristeza, etc.

3.6 ETAPAS PARA REALIZAR UN ANÁLISIS DE SENTIMIENTOS

1. Recopilación de datos: El primer paso para realizar un correcto Análisis de sentimientos es reunir un conjunto de datos. Esto se lleva a cabo a través de herramientas informáticas aplicadas en alguna de las redes sociales que existen. En nuestro caso la herramienta utilizada ha sido SocioViz y la red social ha sido Twitter. En esta fase de la operación es importante seleccionar bien qué palabra o multipalabra vamos a usar como patrón de búsqueda para que los resultados obtenidos sean amplios y estén estrechamente ligados al tema que queremos investigar
2. Identificación de opiniones: Esta identificación se divide en tres fases:
 - a. Los comentarios deben ser separados de los textos presentados.
 - b. Los comentarios deben ser procesados con el objetivo de desechar los comentarios inapropiados y falsos.
 - c. Detectar las opiniones, es decir, las emociones individuales sobre el tema que queremos investigar.
3. Extracción de aspectos: En esta tercera fase se extraen todos los aspectos existentes según los procedimientos que se hayan llevado a cabo. Hasta aquí podríamos hablar de preprocesamiento.
4. Clasificación de la opinión según la técnica que se desee emplear. Estos tres últimos puntos los llevará a cabo la herramienta Mozdeh.

5. Resumen de producción: Como su propio nombre indica, se trata de hacer un resumen de los resultados de las opiniones. El resumen del presente trabajo va a ser presentado en los apartados de cada herramienta, en forma de gráficas, imágenes, texto, tablas y grafos.
6. Evaluación de los resultados, el cual está presentado al final del Capítulo 4 y en las conclusiones.

3.7 APLICACIONES DEL ANÁLISIS DE SENTIMIENTOS

Hoy en día el análisis del Big Data se usa para innumerables campos y motivos, desde directivos de equipos de fútbol que quieren saber a qué futbolistas deben contratar para sus clubs, hasta para en ámbitos de la agricultura realizar, por ejemplo, unas adecuadas medidas de prevención de incendios, o para detectar cual es la calidad del agua.

Podemos mencionar las siguientes aplicaciones como algunas de las más importantes y utilizadas hoy día dividiéndolas en tres grandes sectores:

- a) Empresas-ventas-marketing
 - Valoración de la opinión sobre productos y servicios: es posible que sea la aplicación más práctica y directa del análisis de sentimientos. Gracias a esta aplicación, las empresas pueden saber en todo momento cuál es la imagen que transmiten sus productos o servicios a los usuarios, dando la posibilidad de replantear su estrategia de marketing sin necesidad de estudios lentos y tradicionales como encuestas, adquiriendo así grandes ventajas competitivas.
 - Corrección de opinión errónea: ocurre en muchas ocasiones que los usuarios expresan su opinión en sitios de compra online escribiendo una reseña y aplicando una puntuación, pero puede ocurrir que esa reseña no se corresponda con la puntuación indicada porque el usuario no haya indicado correctamente dicha puntuación, por lo que la aplicación del análisis de sentimientos sobre la reseña escrita podría corregir automáticamente dicha puntuación.
 - Recomendaciones mejoradas: una empresa online recomendará ciertos productos teniendo en cuenta la opinión positiva de los usuarios, y así mismo, dejará de recomendar otros tantos productos si la opinión es negativa.
- b) Aplicación con fines políticos: además de estas aplicaciones que suelen hacer las empresas para incrementar su número de ventas y mejorar su imagen, hay que resaltar la utilización política que se hace hoy en día del análisis de sentimientos para conseguir objetivos tales como; saber qué opinión tiene la población sobre un determinado partido político o candidato político, averiguar en la medida de lo posible la intención de voto de la población en fechas próximas a las elecciones, o conocer cómo es el recibimiento por parte de la población de una nueva medida adoptada por el gobierno y ver la progresión de la misma a lo largo del tiempo.
- c) Análisis de situación financiera: a través del análisis de un conjunto de datos extraídos de diferentes páginas web y redes sociales sobre una empresa en concreto, se puede prever en cierta medida cuál será su futuro financiero.

Dentro de este apartado, debemos hacer una mención especial a la manera en la que se aplica el análisis de sentimientos en el sector turístico. En los siguientes puntos se van a mostrar algunas de las aplicaciones más habituales con algunos trabajos de ejemplo:

- Valoración de la opinión sobre productos y servicios.
 - o Estudio para conocer la polaridad de las reseñas en línea de los turistas en TripAdvisor. (Mirzaalian & Halpenny, 2021).
 - o Estudio en el que se trata de averiguar cuáles son los elementos que hacen que el turista perciba la experiencia negativamente. (Kim, Park, Yun, & Yun, 2017).
 - o Estudio que trata de descubrir qué factores influyen más en la satisfacción de los clientes que usan Airbnb (Chen & Jin, 2019).

- Corrección de opinión errónea.
 - o En este trabajo se ha llevado a cabo un estudio para detectar reseñas falsas en la industria del turismo. (Reyes-Menéndez, Ramón Saura, & Filipe, 2019)

- También se puede usar para la aplicación con fines políticos, ya que ayuda a comprender la situación turística de ciertos territorios a través del estudio de fenómenos que preocupan al sector turístico actualmente.
 - o Tratar de averiguar si el big data existente en las redes sociales es un causante del sobre turismo existente en ciudades como Barcelona (Alonso Almeida, Borrajo Millán, & Yi, 2019).
 - o Conocer cuál es la apreciación que tiene la población sobre el turismo sostenible y cómo puede ayudar la aplicación de este tipo de turismo a las empresas turísticas. (Brzustewicz & Singh, 2021).

3.8 ANÁLISIS DE REDES

Para realizar un análisis de sentimientos más completo, vamos a aprender cómo se construye y se analiza una red o grafo, para así tener un resumen de producción más completo con una mayor cantidad de datos para analizar.

3.8.1 Definición

El análisis de las redes es un campo de investigación dentro del PLN (Procesamiento del lenguaje natural) que se preocupa por estudiar los vínculos que surgen entre una serie de elementos o nodos que constituyen una red, extrayendo de ellos una serie de datos relacionales, es decir, datos relacionados entre sí a partir de una información que puede ser objetiva o subjetiva.

3.8.2 Estructura red

Existen diferentes estructuras de red. En este trabajo vamos a tratar con una estructura de tipo red social o social media denominada Twitter. Estas redes están constituidas en forma de grafos, los cuales están compuestos por una serie de elementos denominados nodos y por las relaciones que se dan entre ellos. Estos nodos normalmente aparecen representados en forma de círculo y su función es almacenar diferentes tipos de datos y de información que el investigador habrá tenido que recolectar con anterioridad.

Entre cada pareja de nodos puede haber uno, varios o ningún vínculo representado por aristas o flechas. La diferencia entre que haya aristas o flechas reside en el tipo de red con que estemos tratando, si se trata de una red dirigida, los nodos que la componen estarán conectados por flechas, lo que significa que el desplazamiento de un nodo a otro sólo es posible en un sentido, mientras que, en las redes no dirigidas, los nodos están conectados por aristas, lo que significa que el desplazamiento entre dos nodos es posible en ambos sentidos. Como ejemplo de red no dirigida de la red social Twitter podría ser una red en la que los nodos serían hashtags, y como ejemplo de red dirigida podría ser una red conformada por los usuarios más activos.

Estas flechas y aristas pueden valorarse de alguna forma, siendo el investigador el responsable de averiguar cuáles son de mayor interés, teniendo en cuenta la dirección que tienen y su fuerza o grosor.

Para comprender mejor qué es un grafo, vamos a ver un ejemplo simple de uno conformado por 5 nodos:

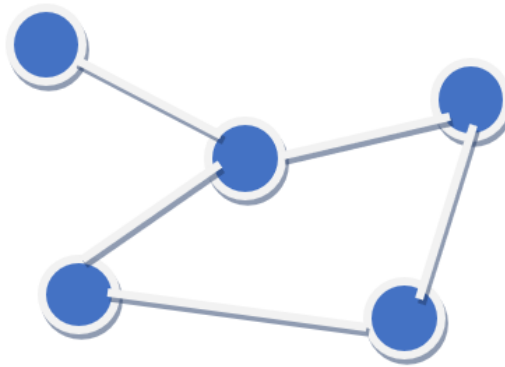


Figura 3.1. Grafo de ejemplo

Fuente: Elaboración propia a partir de (Stuart, y otros, 2019)

La estructura de la red es la que determina los fenómenos que ocurren en ella, por lo que, lo más importante de este punto es tener claro con qué red queremos trabajar (Twitter) y saber cómo analizarla y delimitarla para que nos proporcione la máxima información posible con la menor cantidad de información residual.

Wasserman y Faust (1994) sugirieron dos enfoques para la definición de los límites de una red:

- La primera sería un caso en el que una definición externa determine los elementos que pertenecen a la red.
- En el segundo enfoque sería el propio investigador el que tuviera que determinar ese límite. Esta limitación podría tener un criterio realista, es decir, si la pertenencia a esa red es identificada y reconocida por los mismos actores (nodos), o nominalista, si se basa en una elección que impone el criterio del investigador en la selección de los miembros de la red, ya sea por supuestos teóricos o por conveniencia en la investigación.

Según Smith (2014), existe una serie de estructuras, que, aunque pueden presentar variedades y no cumplir con un patrón exacto, aparecen con una mayor asiduidad en Twitter. Estas estructuras vienen definidas y representadas por las figuras de los siguientes puntos:

- Multitudes polarizadas, donde aparecen grandes grupos de nodos divididos, es decir, sin tener relación.

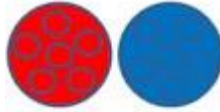


Figura 3.2. Multitudes polarizadas

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

- Multitudes unificadas; se trata de una gran cantidad de nodos divididos en grupos, pero que tienen una relación intermedia.



Figura 3.3. Multitudes unificadas

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

- Clústeres fragmentados: muchos conglomerados de nodos entre los que se da poca interacción.



Figura 3.4. Clústeres fragmentados

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

- Conglomerados múltiples: nodos agrupados por intereses comunes entre los que se da poca interacción.



Figura 3.5. Conglomerados múltiples

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

- Red de amplificación: un nodo central que es poco activo pero que recibe muchas interacciones de otros muchos nodos.



Figura 3.6. Red de amplificación

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

- Red de apoyo: al contrario que la red de amplificación se trata de un nodo central con mucha actividad pero que genera poca respuesta.



Figura 3.7. Red de apoyo

Fuente: (Smith, Rainie, Himelboin, & Shneiderman, 2014)

3.8.3 Características y propiedades de la red

Al realizar un análisis de redes en Twitter podemos encontrarnos con una serie de propiedades que la componen y la caracterizan, pudiendo diferenciar entre las que actúan como indicadores de estructura, las cuales analizan la red en su totalidad, y las que actúan como indicadores de centralidad, las cuales examinan la posición y la importancia de los nodos de manera individual:

3.8.3.1 Indicadores de estructura

- I. Diámetro: se trata de la longitud que se da entre los nodos más separados de la red, es decir, entre los dos nodos que están separados por un mayor número de conexiones.
- II. Densidad de la red: se trata del coeficiente entre el número de vínculos existentes y el número máximo de vínculos que podría existir, es decir, si todos los nodos estuvieran conectados directamente.
- III. Distancia media: se trata de la distancia media que hay entre los nodos de la red, es decir, la media del número mínimo de conexiones o saltos que hay que dar para conectar un nodo con otro.
- IV. Grupos: en las redes existen los denominados grupos, se trata de una serie de nodos que están relacionados entre sí de una manera directa o indirecta. Dentro de estos grupos encontramos dos tipos; los grupos donde todos los nodos están relacionados entre sí de forma directa, es decir, donde cada nodo tiene una vinculación directa con todos y cada uno de los demás nodos que integran el grupo; a este fenómeno se le denomina “camarilla”. Por otra parte, encontramos los grupos donde los nodos pueden estar vinculados entre sí de forma directa o indirecta, a este fenómeno se le denomina “componente”.
- V. Grado medio: se trata del número medio de enlaces que tienen los nodos.

3.8.3.2 Indicadores de centralidad.

Existen varias formas de medir esta centralidad:

- I. Grado: se trata del protagonismo que tienen algunos nodos dentro de una red, este protagonismo se traduce en el número de conexiones que tienen con los demás nodos. Cuanto mayor sea el número de conexiones, mayor centralidad tendrá, lo que se traducirá en un poder mayor, obteniendo así una situación ventajosa.

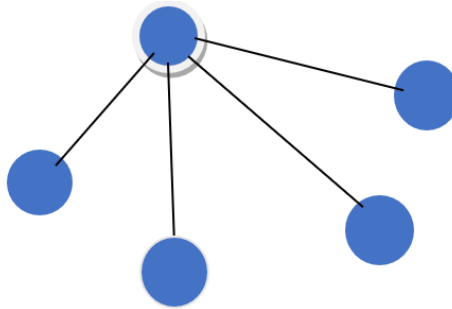


Figura 3.8. Nodo con centralidad por grado

Fuente: Elaboración propia a partir de (Stuart, y otros, 2019)

- II. Cercanía o closeness centrality: también es importante analizar las conexiones indirectas que tienen los nodos con los demás nodos, es decir, la cercanía que tiene con otros nodos con los que, aunque no está directamente conectado, pero de los que puede obtener información por su cercanía. Aquí encontramos relacionada la centralidad de vector propio: la centralidad de un nodo depende de cómo sean de centrales sus nodos vecinos, es decir, de la importancia de sus conexiones, por lo que, en este caso, a diferencia de la centralidad de grado, no se tiene en cuenta la cantidad sino la calidad de las conexiones.
- III. Rol de intermediación: desarrollada por Burt (1992) a través del concepto de agujeros estructurales. Consiste en que hay nodos dentro de la red que conectan diferentes puntos de esta, lo que quiere decir que también existen nodos que necesitan de estos denominados nodos focales para conectar con otros nodos. En la siguiente imagen podemos observar a la izquierda un ejemplo de grupo de nodos denominados camarilla y en la derecha un grupo de nodos denominados componente, además de los nodos (grises) con rol de intermediación que los unen:

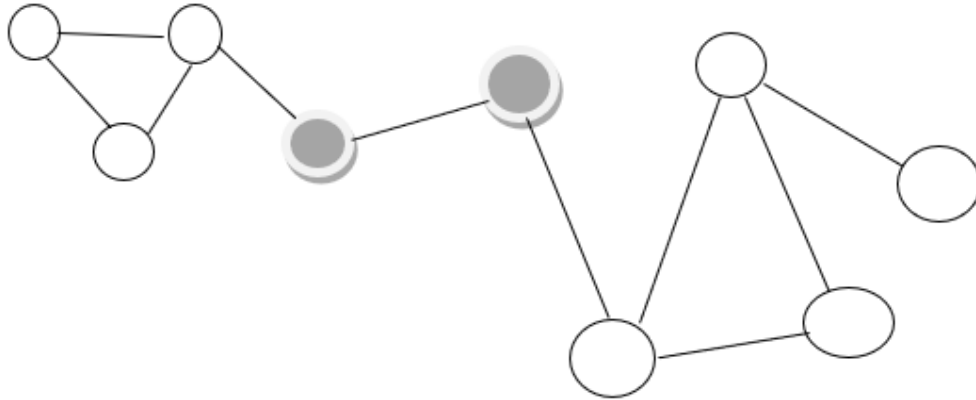


Figura 3.9. Grupos de nodos más roles de intermediación

Fuente: Elaboración propia a partir de (Stuart, y otros, 2019)

- IV. **Excentricidad:** se trata de la distancia desde un nodo a el nodo más alejado de él en la red. Es decir, calcula la distancia más larga para cada nodo. Los nodos periféricos de la red suelen tener un mayor valor de excentricidad, mientras que los nodos más cercanos al centro de la red suelen tener un valor menor de excentricidad. El valor mínimo de excentricidad lo da el radio de la red y el valor máximo lo da el diámetro de la red.

Una vez hemos conocido qué es la red o grafo, qué elementos lo componen y cómo funcionan, tenemos que ver cómo se recopilan los datos que van a generar nuestro grafo, ver cómo se construye un grafo que nos sea útil y realizar las tareas de detección de polaridad y de intensidad para alcanzar el objetivo principal de este trabajo.

Existen muchos métodos y herramientas que realizan esa función de recolección de datos y de información, de hecho, algunos de los más importantes intermediarios turísticos que existen en la actualidad, como Tripadvisor, realizan ese trabajo ofreciendo la posibilidad a las diferentes empresas turísticas de conocer qué opinión tienen los consumidores sobre sus productos o servicios mediante la acción de rastreo, conteo y división en subgrupos, a través de los comentarios que dejan los clientes en sus foros, ya que esos comentarios poseen un título, el cual tiene la capacidad de sintetizar los sentimientos vividos por el cliente durante su estancia en el hotel, los cuales pueden ser analizados para extraer sus patrones semánticos.

CAPÍTULO 4: EJECUCIÓN DEL ANÁLISIS DE SENTIMIENTOS

4.1 RECOLECCIÓN DE DATOS CON SOCIOVIZ

4.1.1 Definición de Socioviz

Para el rastreo y recolección de datos de la red social Twitter hemos utilizado la herramienta informática Socioviz. Socioviz es una herramienta de análisis de medios sociales que está impulsado por métricas de análisis de redes sociales (Centralidad, centralidad de intermediación, densidad...) y que es capaz de generar archivos gexf.

4.1.2 ¿Cómo funciona?

Vamos a realizar nuestra recolección de datos a la vez que explicamos cómo funciona la herramienta Socioviz paso a paso.

1. Para acceder a ella primero hay que crear una cuenta de usuario y elegir un plan de búsqueda entre los varios que ofrece la herramienta que se ajuste mejor a la investigación que se quiera realizar.
2. Una vez elegido el plan hay que vincular Socioviz con una cuenta de Twitter y/o de Facebook. En nuestro caso la hemos vinculado con una cuenta de Twitter.
3. Para realizar las búsquedas en cualquiera de estas dos redes sociales es necesario activar la opción “Gorjeo” y/o “Facebook” y rellenar unos parámetros de búsqueda de una manera precisa para que se aproxime lo máximo posible a lo que deseamos encontrar:

Figura 0.1. Filtros de búsqueda en Socioviz

Fuente: Captura de pantalla de la herramienta Socioviz.

- a) Consulta: Esta pestaña es la más importante de todas, pues en ella hay que escribir patrón de búsqueda, este puede ser un Hashtag, nombre de usuario o una palabra o multipalabra. En nuestro caso, hemos utilizado un hashtag. Este debe estar estrechamente relacionado con nuestro tema de investigación, debe ser el que mejor represente la información que queremos recolectar y debe estar presente en una gran cantidad de Tweets, para que el posterior análisis que se realice sea lo más representativo y completo posible. Para encontrar el hashtag más adecuado podemos realizar varias consultas con diferentes hashtags hasta encontrar una que nos aporte

la cantidad de datos que necesitamos. Nosotros hemos utilizado el hashtag #EUCovidCertificate.

- b) En las pestañas de fecha, se puede ordenar que realice la búsqueda sólo con los tweets que se publicaron en un periodo de tiempo concreto o bien, si esto no resulta importante, se puede dejar en blanco. Después de esto sólo quedaría elegir el idioma deseado y elegir la preferencia de búsqueda en la pestaña de tipo de resultado.
- c) Por último, tenemos la posibilidad de aplicar la opción “Buscar”, la cual recolecta tuits que se publicaron con anterioridad al momento de la búsqueda, o bien, ejecutar la posibilidad de realizar una búsqueda que recolecte los tuits que sean publicados a partir de ese momento, es decir, una búsqueda en tiempo real con la opción “Calendario”.

Nuestra búsqueda fue entre las fechas del 2 de Julio de 2021 y del 10 de Julio de 2021, el idioma elegido fue inglés y el tipo de resultado fue “resultados más recientes”, puesto que nos aportaba un mayor número de resultados.

4. En el momento que iniciamos la búsqueda, la herramienta comienza a trabajar recolectando los datos. Cuando la búsqueda y recolección finaliza, obtenemos lo siguiente:



Figura 4.2. Resúmenes de búsqueda en Socioviz

Fuente: Captura de pantalla de la herramienta Socioviz.

En la imagen, podemos apreciar varios cuadrados, dos de ellos contienen los resúmenes con los resultados de búsqueda; uno denominada “pasaporte covid”, el cual ejecutamos cuando estábamos delimitando qué consulta era más adecuada para nuestra investigación, pero que ya no nos sirve puesto que la cantidad de datos que arrojó era insuficiente. Por otra parte, aparece el resumen de la consulta #EUCovidCertificate que sí es la que hemos utilizado para realizar nuestro análisis de sentimientos.

Estos resúmenes de búsqueda reflejan el número de tweets, retweets y usuarios que ha podido recolectar usando nuestro patrón de búsqueda en las fechas señaladas. En nuestro caso, como podemos ver en la siguiente imagen, ha conseguido agrupar 2.056 tweets, 1.791 retweets y 1.521 usuarios.



Figura 4.3. Resumen de búsqueda de #EUCovidCertificate en Socioviz

Fuente: Captura de pantalla de la herramienta Socioviz.

Además de esto, Socioviz ofrece un práctico informe listo para descargar en formato Excel y en formato PNG entre otros, en el que se incluye una pequeña, pero práctica variedad de datos, gráficas y grafos con información sobre los usuarios más activos, las fechas y horarios claves, los hashtags más frecuentes, los tweets redactados, etc.

4.1.3 Análisis de resultados de Socioviz

Vamos a analizar y comentar los aspectos más importantes que nos aportan las tablas, las gráficas y los grafos que nos ha proporcionado Socioviz, relacionándolos con acontecimientos destacados que se han dado en las últimas fechas en relación al Certificado COVID digital de la UE.

- i. Gráfica time-line: la gráfica representada en la figura 4.3. nos muestra el día y la hora que tuvo una mayor actividad nuestro hashtag.
Por una parte, en esta primera imagen podemos observar que, dentro de los parámetros de fecha de nuestra búsqueda, el día 9 de Julio de 2021 a las 07:00 A.M. fue el momento en que más tweets fueron publicados (13).

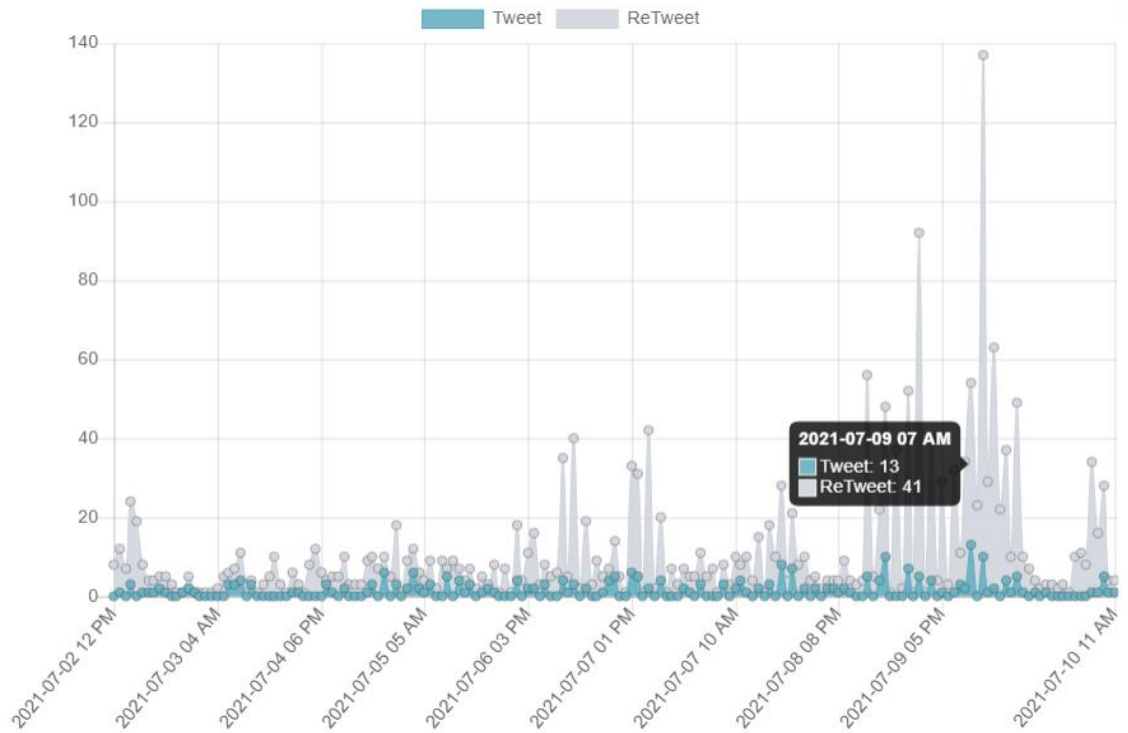


Figura 4.4. Captura de Gráfica time-line (tuits)

Fuente: Captura de pantalla de la herramienta Socioviz.

Por otra parte, la gráfica de la siguiente imagen nos muestra que el momento con mayor número de retweets de nuestro hashtag fue el 9 de Julio de 2021 a las 08:00 A.M. con 127 retweets.

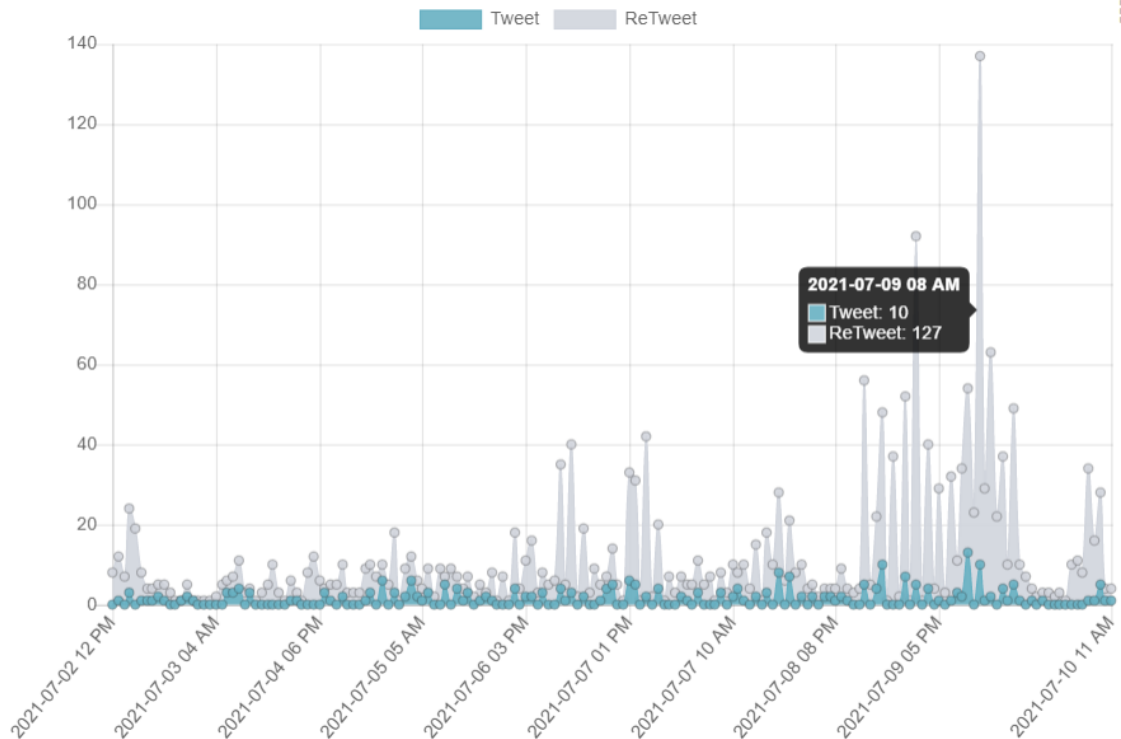


Figura 4.5. Captura de pantalla de gráfica time-line (retuits).

Fuente: Captura de pantalla de la herramienta Socioviz.

Esta concentración de interacciones, en forma tanto de tweets como de retweets, que se dan entre los días 8 de Julio de 2021 y 9 de Julio de 2021 coinciden justo con los días posteriores a la celebración de la sesión plenaria que empezó el 5 de Julio de 2021 y acabó el 8 del mismo mes, en el Parlamento Europeo en Estrasburgo, donde los eurodiputados debatieron la implementación de las regulaciones del Certificado Covid Digital de la UE.

Esto nos enseña que es muy importante saber elegir las fechas para realizar nuestra investigación o, al menos, en este caso, saber elegir cuando realizar la tarea de recolección de la información, puesto que, como hemos visto, hay momentos de gran agitación interactiva en las redes como consecuencia de acontecimientos puntuales, lo que desemboca en un gran vertido de información a las redes.

- ii. Lista de tuits y retuits: El análisis antes descrito no hubiera sido posible sin la lista de tweets y retuits completos y ordenados por fecha y hora que Socioviz nos ofrece.

Hemos podido comprobar, a través de esta lista, que a partir del día 8 de Julio de 2021 se da una gran cantidad de tweets y retuits por parte de los usuarios con preguntas y comentarios sobre las medidas adoptadas en la sesión plenaria. Además, la cuenta oficial de Twitter de la Comisión Europea (@EU_Comission) estuvo los días 7,8 y 9 de Julio de 2021 dando respuesta (como podemos comprobar con el siguiente enlace de Twitter <https://t.co/42b8L5hH0S>) a las dudas y preguntas más frecuentes que podían surgir con relación a las medidas adoptadas sobre el Certificado COVID Digital de la UE, generando así, una gran cantidad de interacciones.



Figura 4.6. Captura de listado de tuits y retuits

Fuente: Captura de pantalla de la herramienta Socioviz.

iii. Tablas y grafo de los usuarios más activos y de los usuarios más influyentes:

- Usuarios más activos: se trata de aquellos usuarios que han interactuado más en la red social, creando una mayor cantidad de contenido a través de la publicación de tweets, y contestando y citando a otros usuarios en relación con el tema que se está tratando, sin embargo, ese contenido creado no tiene por qué haber generado una reacción similar por parte de los otros usuarios.

Usuarios más activos	N.º de tuits enviados
@louise94055458	46
@philipp82621183	32
@robertpernod	26
@joyofhazel	11
@carold82472627	11
@ccbagle	11
@tnteachertim	11
@biddypat	11
@sinead_1981	10
@michaelbann94	10

Tabla 4.1. Usuarios más activos

Fuente: Elaboración propia a partir de los datos aportados por Socioviz.

- Usuarios más influyentes: Se trata de aquellos usuarios que, a través del contenido que han creado, han generado una reacción mayor por parte de los otros usuarios que han participado en la red. Esta reacción se ve reflejada a través del número de retuits o RT recibidos y por las menciones en tuits de otros usuarios.

Usuarios más influyentes	N.º de RT/Menciones recibidas
@eu_commission	754
@grahambsi	223
@jb_bax	156
@eu_consumer	114
@euinmalaysia	90
@michalisrokaseu	89
@vonderleyen	63
@ottocrat	36
@campbellclaret	26
@riffmandale	25

Tabla 4.2. Usuarios más influyentes

Fuente: Elaboración propia a partir de los datos aportados por Socioviz.

Como podemos observar en la segunda tabla, el usuario que más retuits ha generado con sus tuits y más menciones ha recibido ha sido @eu_commission, esto vuelve a tener una relación estrecha con el acontecimiento que tuvo lugar entre los días 7 y 9 de julio de 2021, puesto que fue el usuario que se encargó de informar y dar respuesta a las dudas en lo relacionado a la sesión plenaria que hemos comentado antes.

El grafo siguiente representa las relaciones o conexiones/interacciones que se dan entre los usuarios; como podemos observar en el grafo, los nodos que representan a los tres mayores influenciadores de nuestra red (@eu_commission, @grahambsi y @jb_bax) se encuentran muy cercanos dentro de la red pudiendo afirmar que pertenecen al mismo grupo.

En cuanto al análisis de las aristas o flechas del grafo podemos ver que la mayoría de las flechas que tocan al nodo que representa al usuario más activo (@louise94055458) parten del mismo, lo que nos confirma que es un usuario que ha contactado con muchos usuarios, pero de los que apenas ha recibido o generado respuesta, representando un tipo de estructura de red de apoyo. Por otra parte, encontramos aristas o flechas muy gruesas entre algunos de los usuarios más influyentes como @eu_commission y @jb_bax, además de @eu_health lo que significa que se ha producido una gran interacción entre esos nodos.

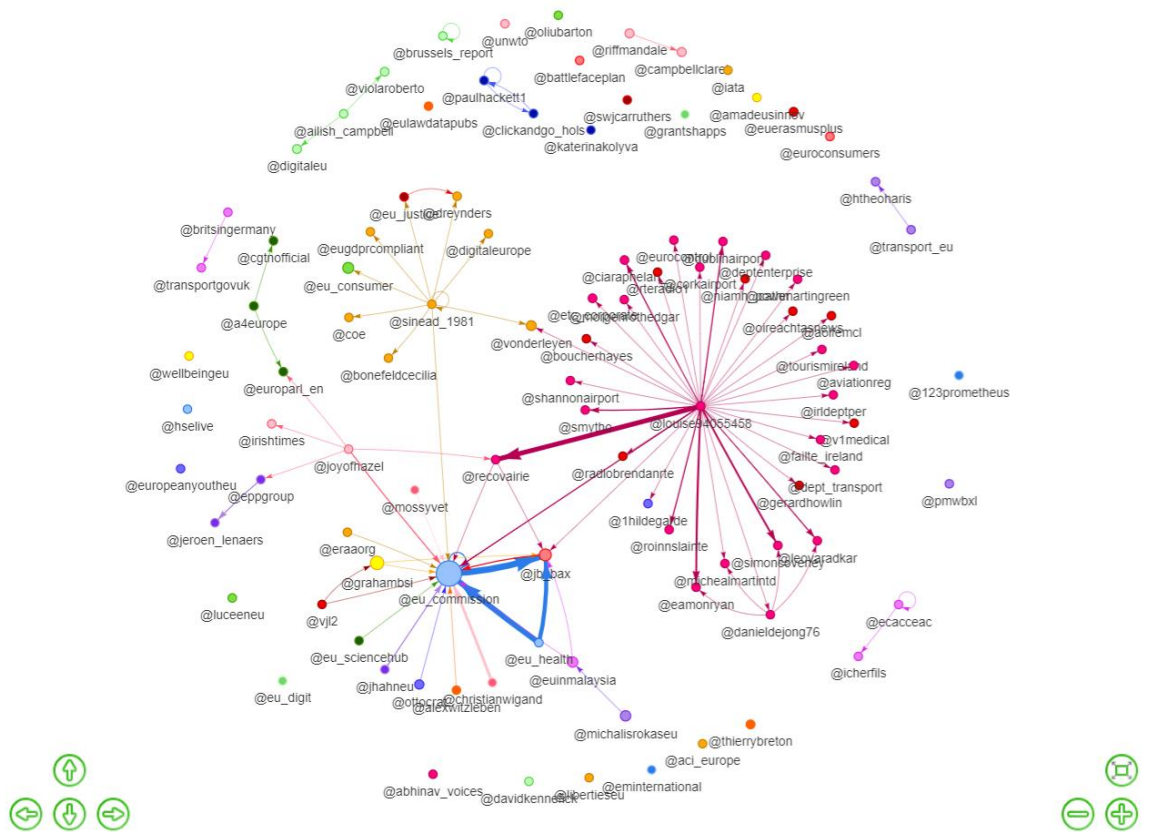


Figura 4.7. Grafo de usuarios

Fuente: Captura de pantalla de la herramienta Socioviz.

- iv. Tabla y grafo de palabras más usadas: Vamos a ver y comentar una tabla y una red o grafo con las diez palabras que más se repiten en los tweets que hemos recolectado.

Palabra	Frecuencia
summer	1141
travel	1107
swiss	639
today	620
countries	561
gateway	440
switzerland	430
joined	415
travellers	415
great	407

Tabla 4.3. Palabras más usadas

Fuente: Elaboración propia a partir de los datos aportados por Socioviz.

Esta tabla ya nos va dando una pista de la opinión que tienen los usuarios sobre el Certificado COVID digital de la UE, ya que encontramos la palabra “great” entre una de las diez más repetidas, lo que previsiblemente hace indicar que la tendencia del sentimiento será positiva en su mayoría.

Por otra parte, en el grafo podemos observar que tres de los nodos más importantes como son “summer” (verano), “travel” (viajar) y “countries” (países) tienen una fuerte conexión ya que las aristas que los unen son bastante gruesas, lo que nos indica que una de las cosas que más les preocupa a los usuarios está relacionado con la posibilidad de poder viajar a otros países este verano.

Además, cabe destacar la presencia de las palabras Swiss (suizo/a) y Switzerland (Suiza), esto se debe a la aprobación que se hizo el 8 de julio para que los ciudadanos suizos puedan viajar dentro de la UE y viceversa con sus respectivos certificados, siendo una de las medidas aprobadas más importantes.

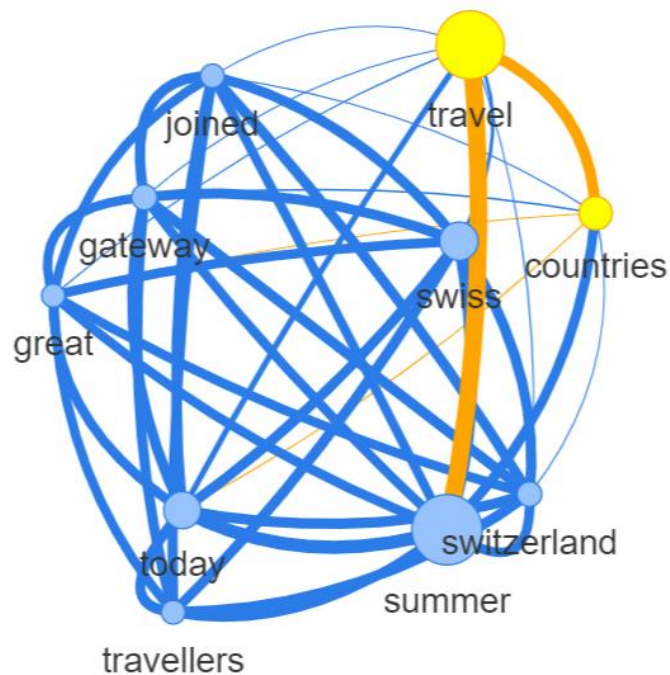


Figura 4.8. Grafo Top 10 palabras más usadas

Fuente: Captura de pantalla de la herramienta Socioviz.

- v. Tabla y grafo de hashtags más importantes: Vamos a ver y comentar una tabla y una red o grafo con los diez hashtags más usados por los usuarios en los tweets que hemos recolectado.

Como es lógico, el hashtag más repetido es #EUCovidCertificate, ya que es el patrón que escogimos para que se repitiera en todos los tweets que queríamos recolectar. Cabe decir que, Twitter no hace distinción entre minúsculas y

mayúsculas en la búsqueda de hashtags, por lo que en la práctica los hashtags #EUCovidCertificate y #eucovidcertificate darán los mismos resultados de búsqueda.

El segundo hashtag más repetido es #kitajagakita, está escrito en el idioma malayo y significa “nos cuidamos”. Esta fuerte presencia como uno de los hashtags con mayor número de interacciones se debe a que el 7 de julio de 2021, el embajador de la Unión Europea en Malasia, Michalis Rokas, con nombre de usuario en Twitter de @MichalisRokasEU, compartió un comunicado en esta red social elaborado por la Delegación de la Unión Europea en Malasia, en relación a los viajes hacia la UE por parte de los habitantes malayos, y para dar respuesta a las cuestiones en relación con las vacunas producidas por fabricantes que aún no han solicitado, o no tienen la intención de solicitar, una autorización de la Agencia Europea de Medicamentos (EMA).

*Frecuencia simboliza el número de veces que se usó cada hashtag.

Hashtag	Frecuencia
#eucovidcertificate	2056
#kitajagakita	90
#strongtogether	90
#strongertogether	47
#reopeneu	46
#covid19	45
#travel	37
#eudigitalcovidcertificate	26
#brexitreality	26
#etoncurtain	26

Tabla 4.4. Hashtags más importantes Gephi

Fuente: Elaboración propia a partir de los datos aportados por Socioviz.

Por otra parte, hay que mencionar la presencia de los hashtags #strongtogether #strongertogether en el tercer y cuarto puesto del top diez de hashtags con más interacciones, que significan en español “juntos somos más fuertes”, lo que hace presagiar de nuevo que existe una tendencia al alza de pensamiento positivo con el Certificado COVID digital de la UE por parte de los usuarios.

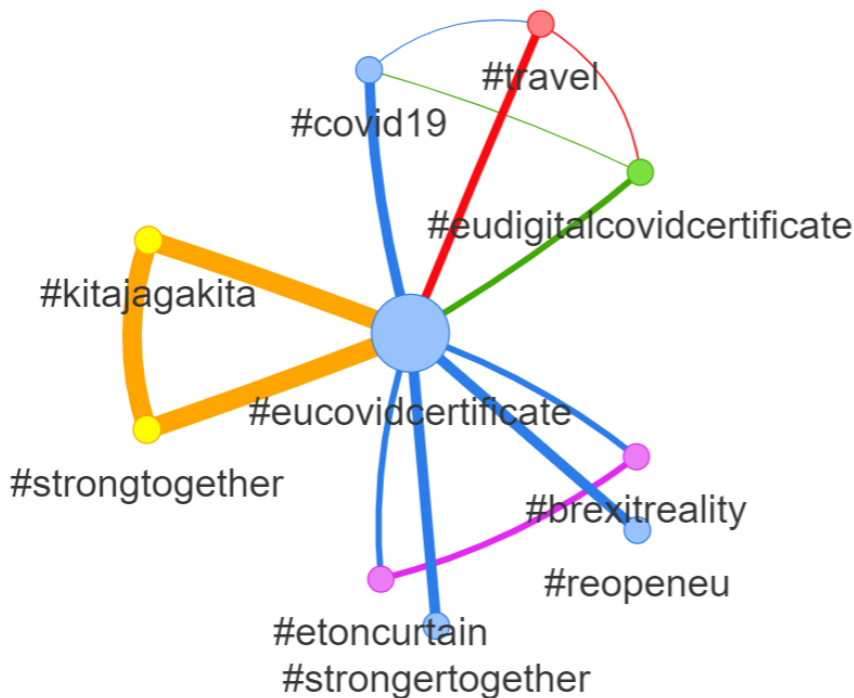


Figura 4.8. Grafo Top 10 hashtags más importantes

Fuente: Captura de pantalla de la herramienta Socioviz.

Por último, para acabar con Socioviz y dar paso a un análisis más exhaustivo de nuestras redes, hay que descargar el informe que nos ha dado Socioviz para exportarlo a la herramienta Gephi. Lo más probable es que al abrir los archivos descargados se nos abren directamente en Gephi siempre que lo tengamos instalado.

4.2 ANÁLISIS DE NUESTRA RED CON GEPHI

4.2.1 Definición de Gephi

Gephi es un software de código abierto, cuya finalidad de uso es la configuración, visualización y análisis de redes. Se caracteriza por su utilidad para buscar patrones y tendencias en una base de datos. (Armetrics, 2021)

4.2.2 ¿Cómo funciona? Construcción y análisis de nuestra red paso a paso

El primer paso para poder empezar a usar Gephi es tener una matriz de datos, la cual puede ser creada manualmente o puede ser importada desde otro programa. En nuestro caso se han importado los datos desde Socioviz, concretamente la matriz de datos de hashtags. A partir de este momento comienza la configuración y análisis de nuestra red de hashtags a través de tres ventanas en las que está dividida Gephi:

4.2.2.1 Vista general

En esta pestaña es donde aparece representado nuestro grafo y donde es posible darle forma y calcular sus métricas. Nuestro grafo está compuesto de 141 nodos y 560 aristas, como muestra la ventana denominada "Contexto" que se encuentra arriba a la derecha de la pestaña "Vista General".

Al abrir la matriz de datos en Gephi, es muy probable que aparezca un grafo con nodos muy próximos entre sí y del que resulta difícil sacar alguna conclusión, al menos visualmente. En esta pestaña se va a tratar de convertirlo en un grafo mucho más visible y revelador aplicándole una serie de herramientas que funcionan con algoritmos:

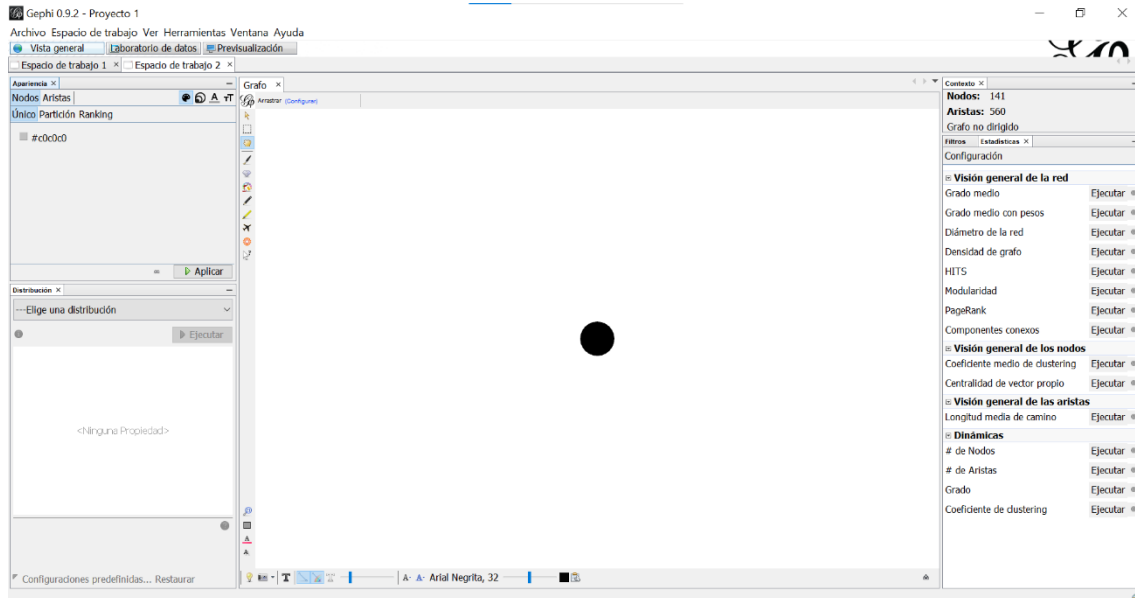


Figura 4.9. Captura I de la herramienta Gephi.

Fuente: Captura de pantalla de la herramienta Gephi.

- i. Modificar la distribución: en la parte izquierda de la imagen encontramos una opción denominada “Elige una distribución”, la cual contiene una serie de algoritmos de distribución. Nosotros escogemos “Force Atlas” y seguidamente aumentamos la “Fuerza de repulsión” de los nodos a 5.000. Para entender porque debemos hacer esto, hay que explicar que la mayoría de los algoritmos que encontramos aquí funcionan como algoritmos de mínima energía, esto se entiende mejor a través de la hipótesis en la cual los nodos son cargas que se repelen entre sí y los enlaces o las aristas son muelles que tratan de atraerlos y mantenerlos juntos. Es por esto, que, al aumentar esa fuerza de repulsión los nodos se distancian unos de otros, pero no se llegan a separar. Volviendo a nuestro objetivo, para asegurarnos que ningún nodo tape a otro en la visualización de la red, vamos a marcar la opción “Ajustar por tamaño”. El resultado de este primer paso lo vemos en la siguiente imagen.

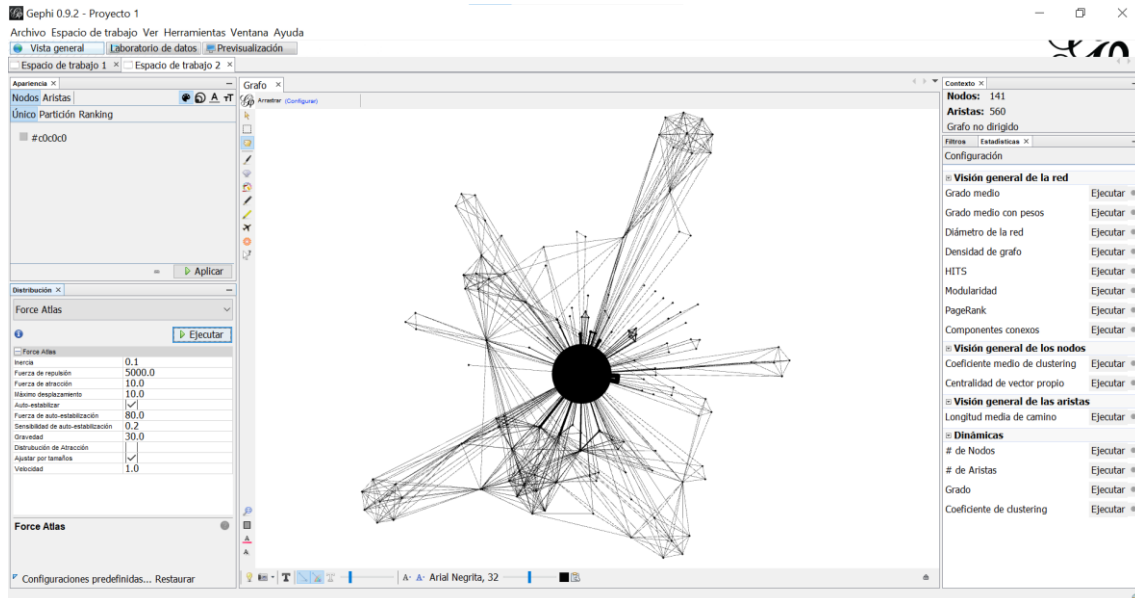


Figura 4.10. Captura II de la herramienta Gephi

Fuente: Captura de pantalla de la herramienta Gephi.

- ii. Estadísticas: se encuentra en la parte central derecha. Está compuesto por una serie de algoritmos que se pueden ejecutar para calcular características de la red, que hemos visto en capítulos anteriores, y expresadas en gráficas, tanto de manera global como de los nodos en particular:
 - Grado medio: Al seleccionar este algoritmo, nos da el número medio de enlaces que tienen los nodos, siendo de 7'943 enlaces de media por nodo. Además, nos ofrece la siguiente gráfica con la distribución de los nodos según su grado. Podemos observar que hay un nodo que tiene 140 enlaces, siendo el que tiene el mayor grado, esto nos dice que, al haber 141 nodos, este nodo tiene conexión con todos los demás de la red. Y por otro lado encontramos que hay 17 nodos con un solo enlace, siendo estos los que tienen el menor grado de toda la red.

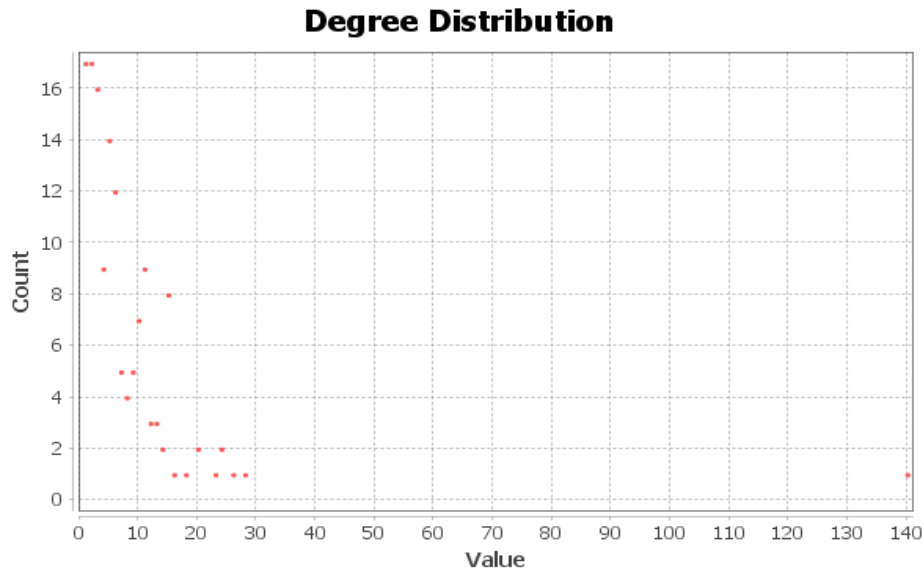


Figura 4.11. Gráfica de grado medio.

Fuente: Captura de pantalla de la herramienta Gephi.

- Diámetro de la red: aunque esta opción se llame “Diámetro de la red”, aplica algoritmos que calculan y analizan varias características además de la mencionada:
 - Por una parte, nos proporciona efectivamente el diámetro; 2, esto significa que la distancia entre los dos nodos más alejados de la red es de dos aristas, siendo la distancia más larga de la red.
 - Excentricidad: la excentricidad media es de 1,9434650455927052. La siguiente gráfica de excentricidades nos muestra que hay un nodo cuya excentricidad es 1, la más baja posible, ya que coincide con el radio, y la del resto de nodos es 2, la más alta posible, ya que coincide con el diámetro de la red. Esto nos vuelve a indicar que, el nodo cuyo valor de excentricidad es uno está conectado con todos los demás nodos que componen la red, ya que su mayor distancia en la red es 1.

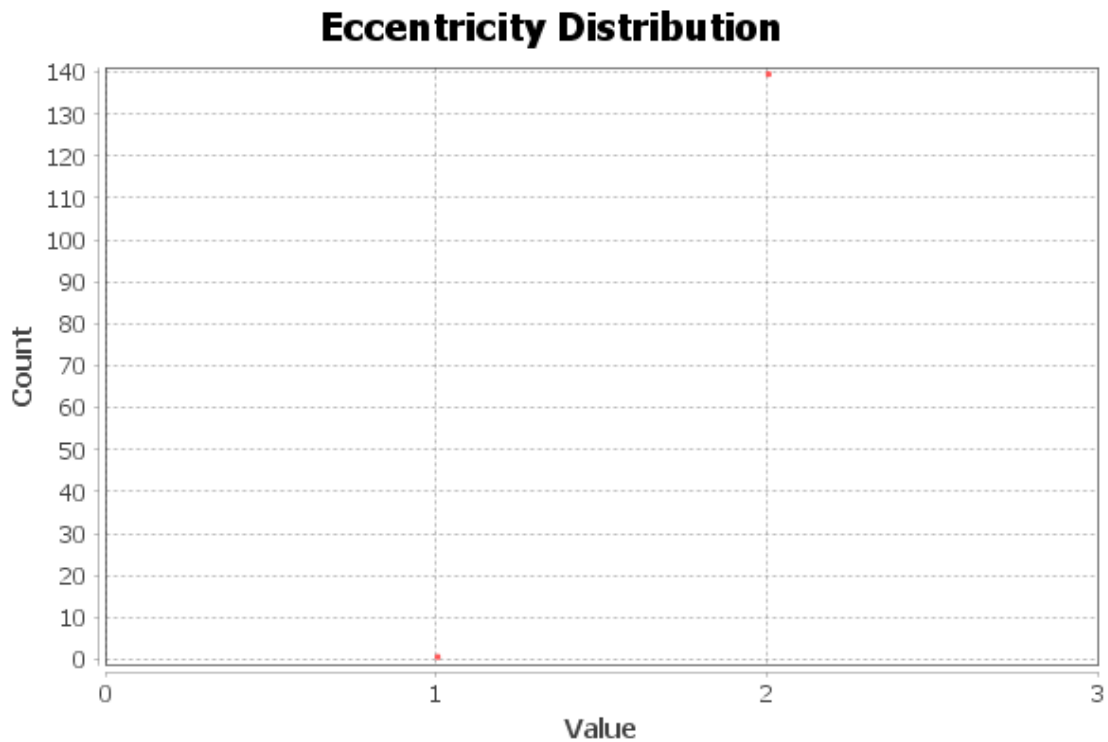


Figura 4.12. Gráfica de excentricidades.

Fuente: Captura de pantalla de la herramienta Gephi.

- **Closeness Centrality:** se trata de la distancia media que tiene un nodo con todos los demás nodos de una red calculada a la inversa, para que los nodos que estén más centrados en la red tengan un mayor valor. En la siguiente gráfica vemos que hay un nodo que tiene un valor de closeness centrality de 1 y los demás están en una horquilla entre 0,5 y 0,6, teniendo en cuenta que el valor mínimo es 0 y el valor máximo es 1. Ese nodo con valor uno es el que tiene una distancia media menor con los demás nodos de la red.

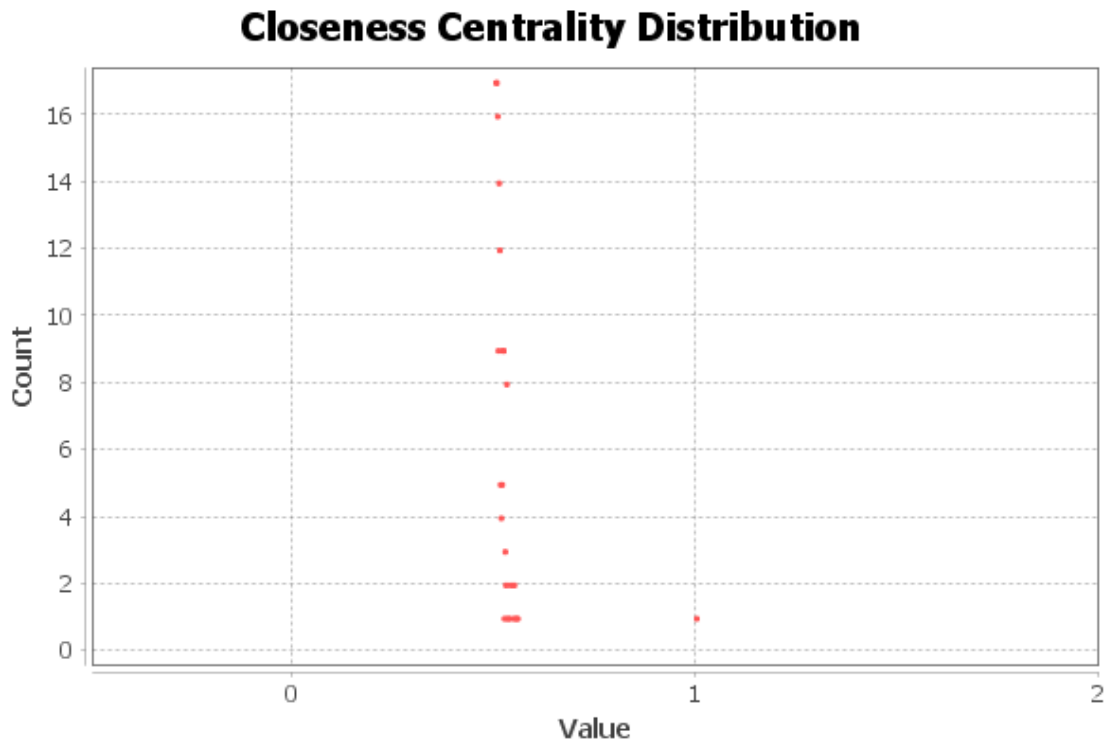


Figura 4.13. Gráfica de Closeness Centrality

Fuente: Captura de pantalla de la herramienta Gephi.

- Betweenness o poder de intermediación: trata de calcular qué nodo es más importante en base al número de caminos más cortos que pasan por él y que conectan diferentes partes del grafo. La siguiente gráfica nos muestra el valor de intermediación que tienen los nodos de nuestro grafo, siendo el valor mínimo 0 y el valor máximo 1. Vemos que la mayoría están agrupados entre 0 y 0,1 y hay un nodo que está cerca del valor máximo, por lo que conecta todas las partes de la red, siendo el camino más rápido para muchos nodos.

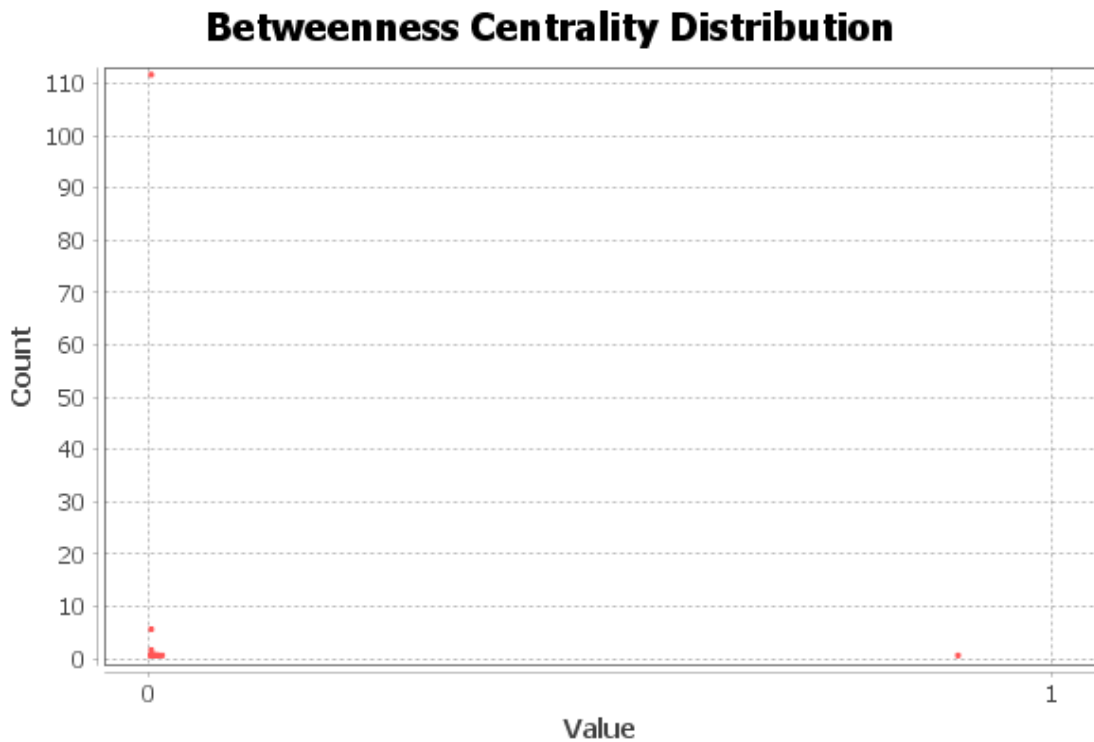


Figura 4.14. Gráfica de valores de intermediación.

Fuente: Captura de pantalla de la herramienta Gephi.

- Densidad: nuestra red tiene una densidad de 0,057, por lo que está lejos de estar completo, es decir, de que hubiera al menos una arista entre cada uno de los nodos.
- Modularidad: los algoritmos de esta opción dividen el grafo en diferentes comunidades o grupos cuyos nodos tienen una mayor conexión entre sí. Aporta, además, el dato de que en nuestro grafo hay 15 comunidades distribuidas de la siguiente manera; como se puede ver en la gráfica, la abscisa representa las comunidades o diferentes grupos que hay y la ordenada representa el número de nodos, por lo que, por ejemplo, en la comunidad denominada "0" hay 58 nodos, siendo la mayor, y en las comunidades denominadas "1", "2", "7", "10" y "12" hay 2 nodos, siendo la menores en número de nodos.

Size Distribution

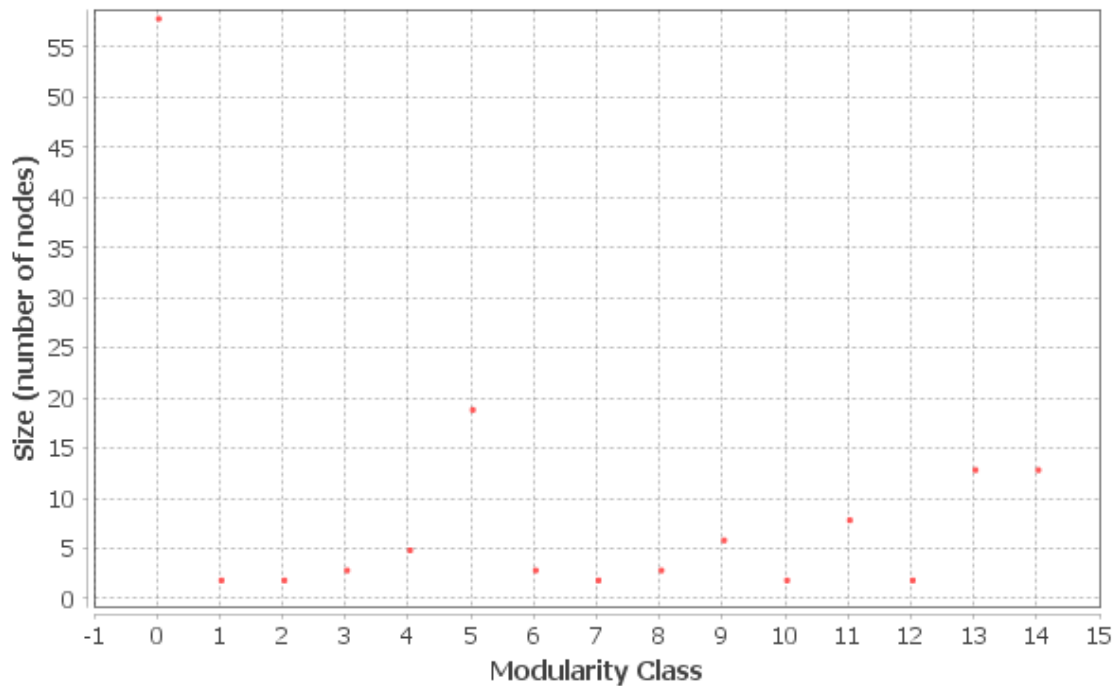


Figura 4.15. Gráfica de grupos o comunidades

Fuente: Captura de pantalla de la herramienta Gephi.

- Centralidad de vector propio: Esta centralidad sirve para medir la calidad de las conexiones del nodo y no tanto la cantidad, como hace el grado. En la siguiente gráfica vemos como están repartidos los nodos según el valor de sus conexiones, siendo el valor mínimo 0 y el valor máximo 1. Vemos que de nuevo hay un nodo que sobresale de los demás, pero en esta gráfica, el valor de los nodos está un poco más repartido que en la gráfica de intermediación.

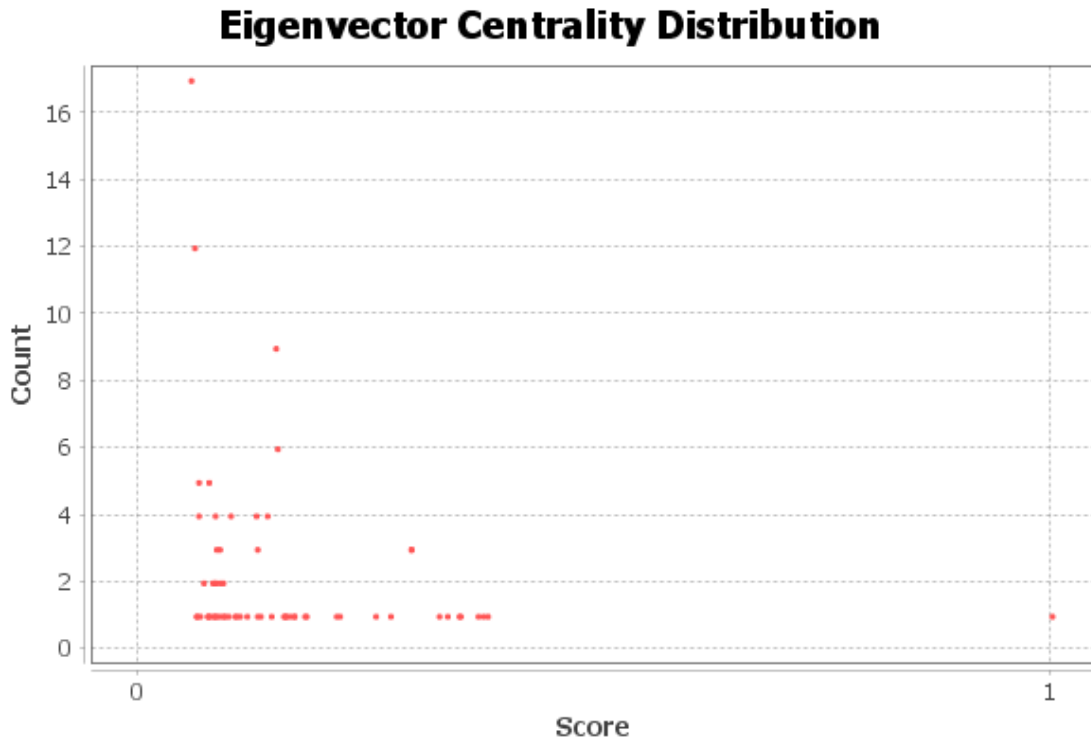


Figura 4.16. Gráfica de centralidad de vector propio

Fuente: Captura de pantalla de la herramienta Gephi.

Una vez que tenemos los algoritmos que nos interesan ejecutados, vamos a dar paso a terminar de manipular el aspecto de nuestra red para hacerla lo más representativa y visual posible en base a algunas de las características que hemos obtenido. Esto lo vamos a realizar en la pestaña de “Apariencia”:

- iii. Apariencia: dentro de esta pestaña vamos a modificar dos aspectos visuales de los nodos:
 - Tamaño: el tamaño de los nodos se va a representar en función al grado. El objetivo en esta pestaña es acercar el tamaño de los nodos más pequeños (grado menor) a el tamaño de los nodos más grandes (grado mayor) para conseguir un grafo donde se aprecien bien todos los nodos y sea más fácil de analizar. Para ello tenemos que clicar en nodos y en la opción que hemos redondeado en la siguiente imagen, seguidamente clicamos en ranking de grado, configurando un tamaño mínimo de 10 y un tamaño máximo de nodo de 40, obteniendo lo siguiente.

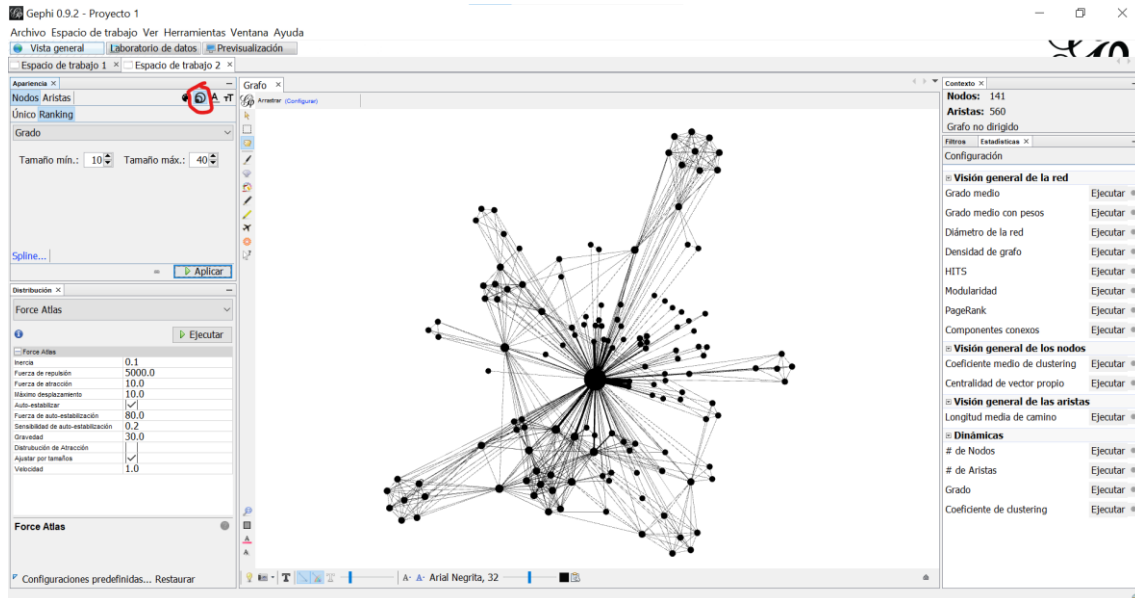


Figura 4.17. Captura III de la herramienta Gephi

Fuente: Fuente: Captura de pantalla de la herramienta Gephi.

- **Color:** en este paso lo que vamos a hacer es partir el grafo en colores para poder diferenciar los diferentes grupos que existen, por lo que cada color representará un grupo distinto. Para ello, tenemos que clicar en nodos y en la opción que hemos redondeado en la imagen, seguidamente clicamos en “partición” y elegimos la opción “Modularity Class”, por último, ejecutamos obteniendo lo siguiente.

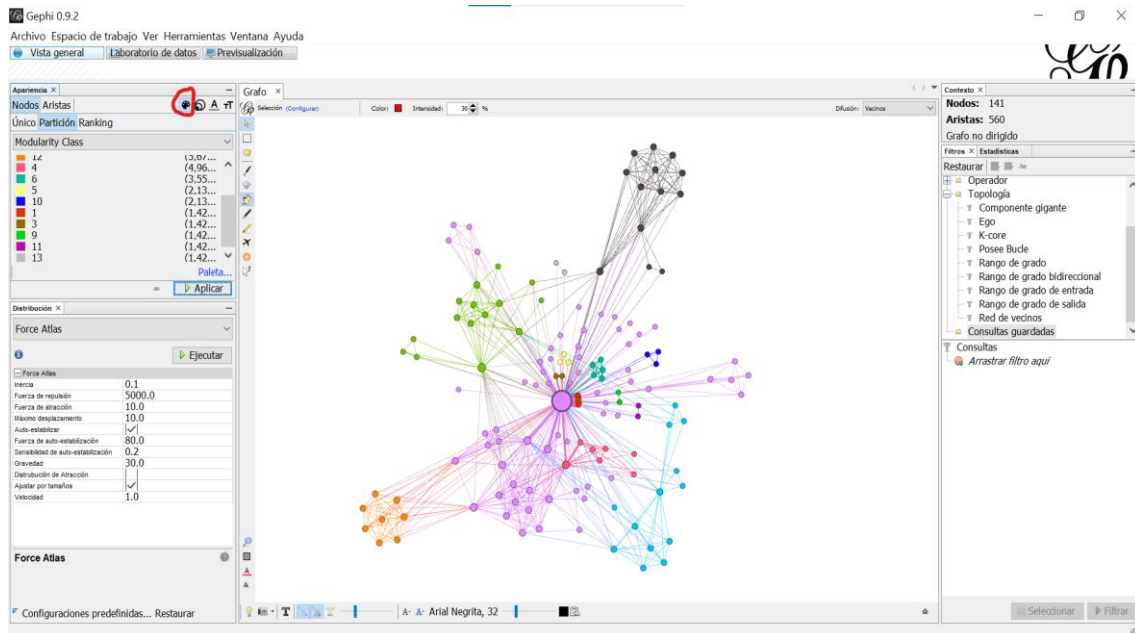


Figura 4.18. Captura IV de la herramienta Gephi.

Fuente: Captura de pantalla de la herramienta Gephi.

4.2.2.2 Laboratorio de datos

En esta sección encontramos dos tablas, una donde se incluyen todos los nodos y las métricas que los caracterizan divididas en columnas, y otra donde aparecen todas las aristas y una serie de características divididas también en columnas. Esta sección por lo tanto sirve para ver las características que se han aplicado antes, pero de manera individual de cada nodo y de cada arista, también sirve para comparar sus valores y su cometido dentro de la red.

- Tabla de nodos:

Id	Label	Timestamp	wordType	Grado	Eccentricity	Closeness Centrality	Betweenness Centrality	Modularity Class	Eigenvector Centrality
#eucovidcertificate	#eucovidcertificate		HASH	140	1.0	1.0	0.894545	0	1.0
#kitajagakita	#kitajagakita		HASH	2	2.0	0.503597	0.0	1	0.061765
#strongtogether	#strongtogether		HASH	2	2.0	0.503597	0.0	1	0.061765
#strongertogether	#strongertogether		HASH	1	2.0	0.501792	0.0	0	0.057873
#reopeneu	#reopeneu		HASH	1	2.0	0.501792	0.0	0	0.057873
#covid19	#covid19		HASH	26	2.0	0.551181	0.007992	0	0.371987
#travel	#travel		HASH	23	2.0	0.544747	0.005808	0	0.351241
#eudigitalcovidcert...	#eudigitalcovidcert...		HASH	24	2.0	0.546875	0.007254	4	0.352599
#brexitreality	#brexitreality		HASH	2	2.0	0.503597	0.0	3	0.061765
#etoncurtain	#etoncurtain		HASH	2	2.0	0.503597	0.0	3	0.061765
#nextgenerationeu	#nextgenerationeu		HASH	7	2.0	0.512821	0.000137	4	0.105954
#vaccination	#vaccination		HASH	6	2.0	0.510949	0.000206	0	0.111443
#19july	#19july		HASH	3	2.0	0.505415	0.0	0	0.081496
#switzerland	#switzerland		HASH	3	2.0	0.505415	0.0	0	0.066232
#europeantourism	#europeantourism		HASH	2	2.0	0.503597	0.0	0	0.06442
#schengen	#schengen		HASH	1	2.0	0.501792	0.0	0	0.057873
#eu	#eu		HASH	28	2.0	0.555556	0.012059	8	0.382476
#antigen	#antigen		HASH	6	2.0	0.510949	0.0	0	0.084359
#interoperableeu	#interoperableeu		HASH	3	2.0	0.505415	0.0	5	0.066232
#interoperability	#interoperability		HASH	3	2.0	0.505415	0.0	5	0.066232
#digitaldigest	#digitaldigest		HASH	5	2.0	0.509091	0.0	6	0.077377
#roaming	#roaming		HASH	5	2.0	0.509091	0.0	6	0.077377
#meluxina	#meluxina		HASH	5	2.0	0.509091	0.0	6	0.077377
#westernbalkans	#westernbalkans		HASH	5	2.0	0.509091	0.0	6	0.077377
#coronavirus	#coronavirus		HASH	20	2.0	0.538462	0.002243	0	0.338538
#covidpassport	#covidpassport		HASH	14	2.0	0.526316	0.002235	8	0.170629
#wellbeing4all	#wellbeing4all		HASH	2	2.0	0.503597	0.0	0	0.07857
#safetravel	#safetravel		HASH	20	2.0	0.538462	0.00305	0	0.329467
#digital	#digital		HASH	5	2.0	0.509091	0.000154	0	0.09886
#opensource	#opensource		HASH	3	2.0	0.505415	0.0	5	0.066232
#digital...	#digital...		HASH	5	2.0	0.509091	0.0	6	0.077377

Figura 4.19. Captura de la tabla de datos de Gephi (nodos)

Fuente: Captura de pantalla de la herramienta Gephi.

En la primera y segunda columna aparecen el ID (identificador) y Label (la etiqueta de los nodos). A partir de la quinta columna es cuando empiezan a aparecer las características de los nodos que nos interesan.

- 5ª columna: aparece el grado de cada uno de los nodos.
- 6ª columna: valor de excentricidad de cada nodo.
- 7ª columna: valor de closeness centrality de cada nodo.
- 8ª columna: valor de betweenness centrality o de intermediación de cada nodo.
- 9ª columna: identifica el grupo o comunidad a la que pertenece cada nodo.
- 10ª columna: muestra el valor de Eigenvector Centrality o centralidad de vector propio de cada nodo.

Podemos ordenar estas columnas de mayor a menor valor dependiendo de la característica que queramos analizar de cada nodo.

- Tabla de aristas: lo más destacable que muestra esta tabla es el peso (weight) que hay entre las diferentes aristas que conforman nuestro grafo,

esto quiere decir, el número de interacciones que hay entre cada uno de los pares de nodos.

Las columnas número 1 y número 2 muestran los nodos que conectan la arista, y la columna número 7 muestra el peso o número de interacciones que hay en esa arista que conecta ese par de nodos. Como ejemplo vamos a observar la arista con mayor peso de la tabla:

Partiendo de la primera fila de la tabla que se ha adjuntado más abajo, vemos que en la columna número 1 denominada “Origen” aparece el hashtag #strongtogether; este es el primer nodo de los dos que conecta la arista que representa esta fila de la columna, el otro nodo se encuentra en la columna número 2 denominada “Destino” donde aparece el hashtag #euocovidcertificate, por último en la columna número 7 denominada “Weight” encontramos el peso de la arista que une los dos nodos antes mencionados (90), lo que significa que hay 90 interacciones entre estos dos nodos señalados, siendo la relación de nodos o hashtags que más se da.

Origen	Destino	Tipo	Id	Label	Timestamp	Weight
#strongtogether	#euocovidcertificate	No dirigida	32			90,0
#strongtogether	#euocovidcertificate	No dirigida	1			47,0
#euocovidcertificate	#reopeneu	No dirigida	0			46,0
#euocovidcertificate	#covid19	No dirigida	6			45,0
#travel	#euocovidcertificate	No dirigida	23			37,0
#brexitreality	#etoncurtain	No dirigida	111			26,0
#euocovidcertificate	#brexitreality	No dirigida	109			26,0
#euocovidcertificate	#etoncurtain	No dirigida	110			26,0
#digitalcovidcertificate	#euocovidcertificate	No dirigida	28			26,0
#nextgenerationeu	#euocovidcertificate	No dirigida	29			21,0
#digitalcovidcertificate	#nextgenerationeu	No dirigida	27			20,0
#euocovidcertificate	#vaccination	No dirigida	53			16,0
#euocovidcertificate	#19july	No dirigida	25			13,0
#travel	#19july	No dirigida	24			13,0
#euocovidcertificate	#switzerland	No dirigida	33			12,0
#euocovidcertificate	#europeantourism	No dirigida	52			12,0
#europeantourism	#vaccination	No dirigida	54			12,0
#digitaldigest	#roaming	No dirigida	208			9,0
#digitaldigest	#meluxina	No dirigida	209			9,0
#digitaldigest	#westernbalkans	No dirigida	210			9,0
#digitaldigest	#euocovidcertificate	No dirigida	211			9,0
#eu	#euocovidcertificate	No dirigida	83			9,0
#euocovidcertificate	#antigen	No dirigida	133			9,0
#euocovidcertificate	#interoperableeu	No dirigida	160			9,0
#euocovidcertificate	#interoperability	No dirigida	161			9,0
#interoperableeu	#interoperability	No dirigida	162			9,0
#meluxina	#westernbalkans	No dirigida	217			9,0
#meluxina	#euocovidcertificate	No dirigida	218			9,0
#roaming	#meluxina	No dirigida	213			9,0
#roaming	#westernbalkans	No dirigida	214			9,0
#roaming	#euocovidcertificate	No dirigida	215			9,0

Figura 4.20. Captura de la tabla de datos (aristas) de Gephi

Fuente: Captura de pantalla de la herramienta Gephi.

4.2.2.3 Previsualización

En esta ventana del programa se da el último paso donde se aplican los últimos detalles al grafo antes de exportarlo. En este grafo se han aplicado los últimos “Parámetros” más importantes de la siguiente manera:

- Aristas rectas.
- No se muestran las etiquetas en los nodos.
- Se muestran las aristas con un grosor de 0.5
- No se muestren las etiquetas de las aristas.

El resultado final es exportado en formato PNG y se muestra así:

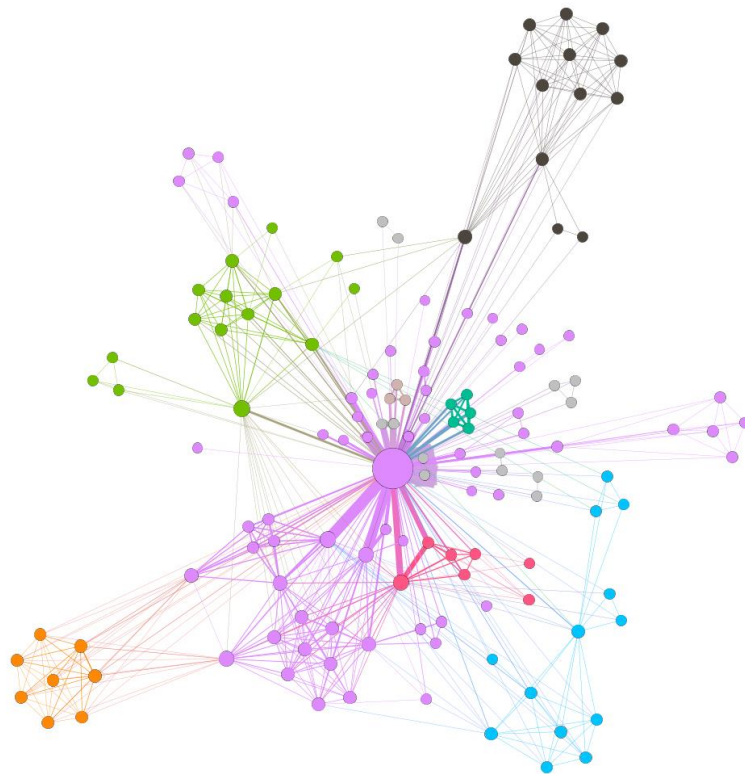


Figura 4.21. Red acabada

Fuente: Red o grafo importado desde Gephi.

Por último, se le ha aplicado un filtro a la red para que muestre sólo los nodos que tienen un valor de grado mayor que 16. Se ha elegido este filtro porque es representativo de distintas variables como closeness centrality o centralidad de vector propio. Estos nodos que aparecen en el siguiente grafo tienen un valor mayor en las dos características antes mencionadas porque pertenecen a un grupo o comunidad fuerte y cercana al nodo central. Además, se les ha añadido unas etiquetas para identificar el hashtag de cada nodo.

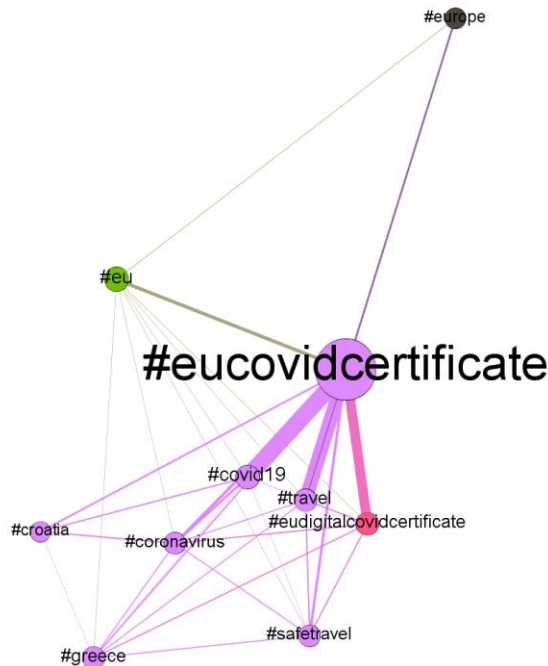


Figura 4.22. Red de nodos más importantes según valor de grado.

Fuente: Red o grafo importado desde Gephi.

En resumen, nos encontramos con un grafo que tiene un nodo (#euCOVIDcertificate) con una influencia y una importancia sobre los demás fuertemente marcada como muestran los resultados de valor de intermediación, de closeness centrality, de excentricidad y de grado medio, estos dos últimos indicándonos que tiene una conexión directa con todos los demás nodos de la red, por lo que su influencia dentro de la red es total. Esto es lógico, ya que se trata del hashtag que se usó como patrón de búsqueda en la recolección de datos

Por otra parte, encontramos que el grafo está dividido en 15 grupos o comunidades, existiendo grupos de un tamaño más o menos considerable pero que se encuentran lejos del centro del grafo o del nodo central, por lo que tienen un valor o importancia menor, es el caso de los grupos representados en color naranja, celeste, aunque este si está algo más cerca del centro, y del grupo representado en color negro, pero cabe decir de este último grupo que cuenta con nodos con un buen valor de vector propio ya que hay un nodo integrante del grupo, denominado #europe, que ejerce de intermediario, ya que conecta a todos los nodos de su grupo o comunidad con nodos cercanos al centro de la red, como podemos ver en la siguiente imagen. La flecha negra señala el grupo o comunidad a la que pertenece el hashtag #europe. La flecha naranja señala el nodo #europe. Y las flechas moradas señalan los nodos centrales y las aristas que los conectan con #europe.

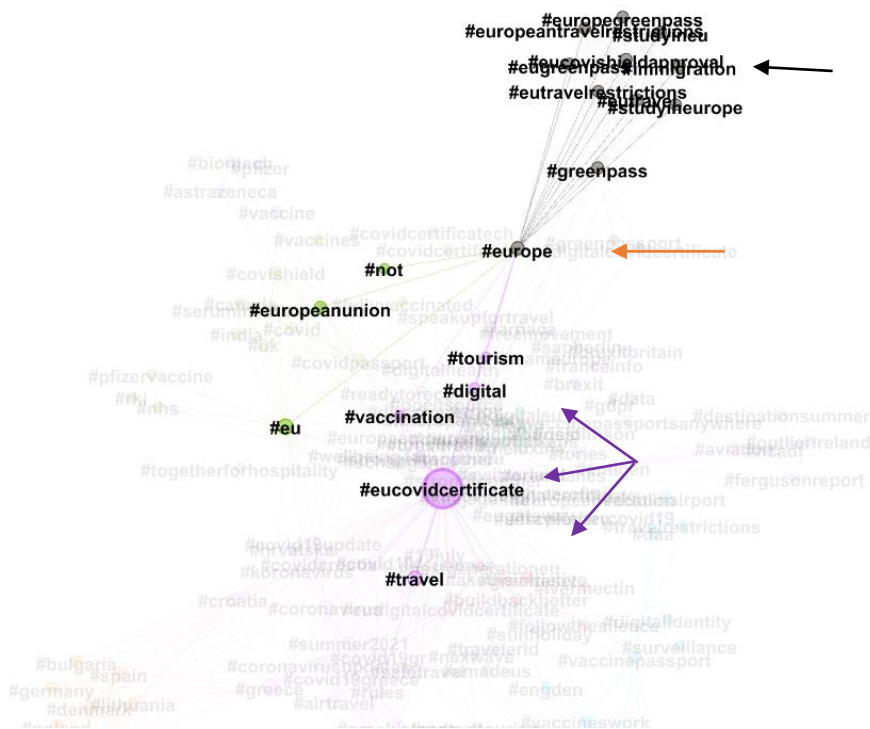


Figura 4.23. Valor de intermediación de #europe

Fuente: Captura de pantalla de la herramienta Gephi.

Por otra parte, encontramos el grupo representado por el color rosa, el cual cuenta con pocos integrantes, pero tiene una buena posición en la red con una conexión pesada con el nodo central y un valor de grado alto en uno de sus nodos #eudigitalcertificate.

4.3 MOZDEH

4.3.1 Definición

Para detectar la polaridad y la frecuencia de los tweets obtenidos con Socioviz hemos utilizado el software denominado Mozdeh. Mozdeh es un programa gratuito de Windows especializado en el análisis del Big Data. Es capaz de realizar el proceso que conforma el Análisis de sentimientos, es decir, desde la recolección de datos y creación del grafo hasta el análisis de sus métricas y de la polaridad e intensidad de los textos (Wolverhampton, 2021). En este trabajo, se va a utilizar mozdeh para:

- La detección de la polaridad de los tweets obtenidos.
- Detectar la intensidad del sentimiento obtenido, tanto negativo como positivo.

4.3.2 ¿Cómo funciona?

Para empezar, se debe tener una matriz de datos que analizar. Esta matriz de datos se puede recolectar con Mozdeh o se puede importar. En este caso se han importado los mismos datos que se importaron con anterioridad a la herramienta Gephi desde Socioviz para poder continuar con el Análisis de Sentimientos. Una vez están los datos que se quieren analizar dentro de Mozdeh, nos encontraremos con la página principal la cual está dividida en diferentes partes:

- I. Pestaña de búsqueda: en esta parte de la página principal se pueden realizar las búsquedas en base a diferentes y variados filtros entre los tweets que están importados.

The screenshot shows the Mozdeh search interface. At the top, there is a search bar and a 'Search' button. To the right, there are options for 'Next Page', 'Previous', and 'Show original text (slow)'. Below the search bar, there are several filter options: 'Report a maximum of one match per user per interval', 'ID of first item to show', 'First date to show', 'User country', 'Only return results from this topic/label authors (case sensitive)', 'Results only from this user ID', 'Only posted by label user', 'Select alternative subproject', 'User gender', 'Text must be in username / URL', 'Exclude matching URL/source instead', and 'Exclude users with this many posts collected'. There is also a 'Results per page' dropdown set to 38. On the right side, there are 'Advanced Search Tips' and a sentiment filter section with a range selector for positive and negative sentiment.

Figura 4.24. Pestaña de búsqueda de Mozdeh

Fuente: Captura de pantalla de la herramienta Mozdeh.

De entre los buscadores que aparecen en la imagen anterior a nosotros nos interesa el que aparece a la derecha de la pestaña cuya función es delimitar los resultados en base a la intensidad del sentimiento. Esto significa que, si se elige la opción “+” se filtrarán los tweets que tengan una intensidad de sentimiento positivo entre 3 y 5, siendo 1 el mínimo y 5 el máximo y una intensidad de sentimiento negativo entre 1 y 2. Mientras que, si se elige la opción “-” se filtrarán los tweets con intensidad de sentimiento positivo comprendido entre 1 y 2 y sentimiento negativo comprendido entre 3 y 5.

The close-up shows the sentiment filter section. It features a range selector for positive sentiment (smiley face icon) with values 3 and 5, and a range selector for negative sentiment (frowny face icon) with values 1 and 2. The filters are connected by 'AND'. To the right, there are buttons for '+', '-', and '0' to adjust the sentiment range.

Figura 4.25. Filtro de búsqueda por frecuencia de sentimiento de Mozdeh

Fuente: Captura de pantalla de la herramienta Mozdeh

- II. Biblioteca de tuits: aquí aparece una lista con todos los tuits según los requisitos de búsqueda que se hayan ejecutado, o si por el contrario no se ha ejecutado ningún requisito de búsqueda de los anteriormente mencionados, aparecerán todos los tuits sin distinción.

ID	Day	Text (abbreviated)
000001 1 2	2021.7.2	@dwnnews hi dw the #eucovidcertificate isn't available in all eu countrie from july ireland is delaying it until july difficultie a
000002 1 1	2021.7.2	croatia covid-19 update july new case active via @dailycroatia over the past day #croatia have reported new #covid19 case #koronav
000003 1 3	2021.7.2	wtf is wrong with the eu with all this favouritism on vaccine all the az vaccine are the same #astrazeneca #biontech #vaccine #pfi
000004 1 1	2021.7.2	@drivetimere @sarahamcinerney i used the #digitalcovidcertificate today flying from vienna-dublin it was not accepted in dublin h
000005 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000006 3 1	2021.7.2	wonderful new for #europe #travel #eucovidcertificate
000007 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000008 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000009 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000010 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000011 2 1	2021.7.2	still european can you help me and a few of my friend #eucovidcertificate #speakupfortravel
000012 2 1	2021.7.2	with @cypriusineu tonight the country joined the #eucovidcertificate ahead of deadline only two day to go and many dozen of million
000013 3 1	2021.7.2	taking it time but is ready as promised at start of #eucovidcertificate all eu eea countrie are live connected to eu gateway and a
000014 3 1	2021.7.2	taking it time but is ready as promised at start of #eucovidcertificate all eu eea countrie are live connected to eu gateway and a
000015 1 2	2021.7.2	@dwnnews hi dw the #eucovidcertificate isn't available in all eu countrie from july ireland is delaying it until july difficultie a
000016 1 1	2021.7.2	croatia covid-19 update july new case active via @dailycroatia over the past day #croatia have reported new #covid19 case #koronav
000017 1 3	2021.7.2	wtf is wrong with the eu with all this favouritism on vaccine all the az vaccine are the same #astrazeneca #biontech #vaccine #pfi
000018 1 1	2021.7.2	@drivetimere @sarahamcinerney i used the #digitalcovidcertificate today flying from vienna-dublin it was not accepted in dublin h
000019 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000020 3 1	2021.7.2	wonderful new for #europe #travel #eucovidcertificate
000021 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000022 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000023 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000024 1 1	2021.7.2	all hand on deck when it come to getting the #eucovidcertificate in the hand of all eu traveller
000025 2 1	2021.7.2	still european can you help me and a few of my friend #eucovidcertificate #speakupfortravel
000026 2 1	2021.7.2	with @cypriusineu tonight the country joined the #eucovidcertificate ahead of deadline only two day to go and many dozen of million
000027 3 1	2021.7.2	taking it but is ready as promised at start of #eucovidcertificate all eu eea countrie are live connected to eu gateway and all ex
000028 3 1	2021.7.2	taking it but is ready as promised at start of #eucovidcertificate all eu eea countrie are live connected to eu gateway and all ex
000029 2 1	2021.7.3	@seankellymep @europarl_en @eu_commission @epppgroup @finegael hi sean can you do a quick on the process for obtaining the #eucovid
000030 1 1	2021.7.3	@newstalkfm no option for antibody test either in recovered know we know these last a lifetime #eucovidcertificate
000031 1 2	2021.7.3	#greece will no longer accept negative self-test result for island air #travel #greekislands #summer2021 #airtravel #rules #safetr
000032 1 1	2021.7.3	#greece open to more non-eu tourist extend #travel #rules #summer2021 #airtravel #safetravel #covid19 #coronavirus #coronavirusupd
000033 2 1	2021.7.3	#eucovidcertificate what about citizen when might we be able to obtain it thank
000034 1 2	2021.7.3	#greece will no longer accept negative self-test result for island air #travel #greekislands #summer2021 #airtravel #rules #safetr
000035 2 2	2021.7.3	etc @santander_edu welcome the launch of #eucovidcertificate on @bbcworld this is a tangible tool for people to #travel again acro
000036 2 1	2021.7.3	@seankellymep @europarl_en @eu_commission @epppgroup @finegael hi sean can you do a quick on the process for obtaining the #eucovid

Figura 4.26. Biblioteca de tuits de Mozdeh

Fuente: Captura de pantalla de la herramienta Mozdeh

Como se puede ver en la imagen, la lista está compuesta por una serie de columnas donde aparecen, en orden de izquierda a derecha, el ID de cada tuit, la puntuación de sentimiento del 1 al 5 según la intensidad tanto positiva como negativa, el día en que fue publicado el tweet y el texto que contiene el tweet.

4.3.3 Análisis de resultados con Mozdeh

En este subapartado se van a exponer todos los resultados por tablas y gráficas que la página principal de Mozdeh ofrece en relación a la polaridad y frecuencia de sentimientos, teniendo en cuenta que el grado mínimo de intensidad es 1 y el máximo es 5.

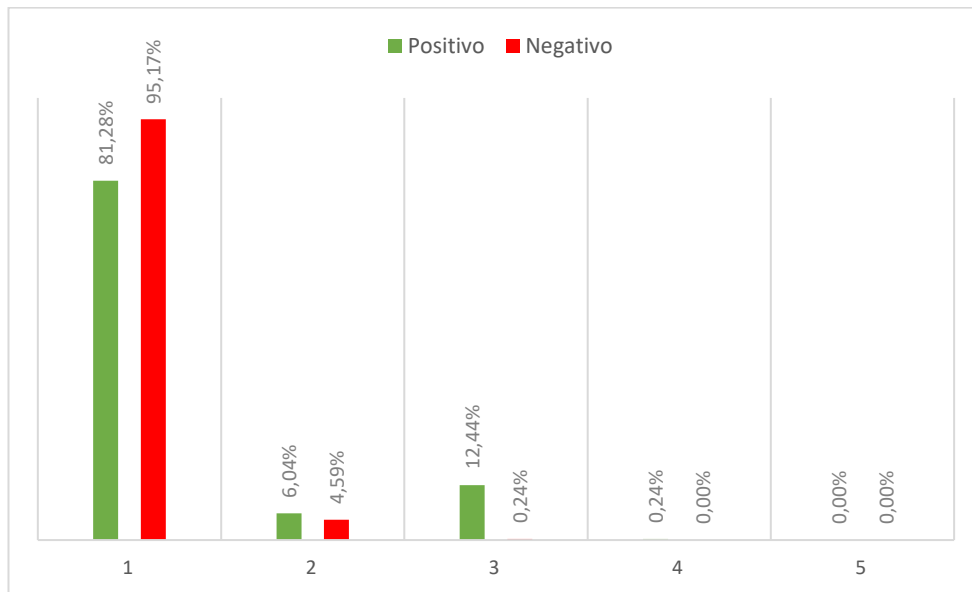
- a) Como hemos comentado anteriormente, cada tuit tiene asignado una puntuación tanto de sentimiento positivo como negativo, en esta tabla se recoge el % de tuits que hay en cada puntuación de intensidad de ambos sentimientos.

Puntuación	Sent. Positivo	Sent. Negativo
1	81,28%	95,17%
2	6,04%	4,59%
3	12,44%	0,24%
4	0,24%	0,00%
5	0,00%	0,00%

Tabla 4.5. % de tuits negativos y positivos en cada intensidad

Fuente: Elaboración propia a partir de los datos aportados por Mozdeh.

La siguiente gráfica es una representación de los datos expresados en la anterior imagen.



Gráfica 4.1. % de tuits negativos y positivos en cada intensidad

Fuente: Elaboración propia a partir de los datos aportados por Mozdeh.

- b) En esta sección se van a mostrar tres resultados relacionados con el número medio de tuits en cada sentimiento e intensidad:
- La puntuación positiva media es 1,3164 y la negativa es 1,0507.
 - Además, que la mayoría de los tweets que tienen sentimiento positivo se encuentran en el intervalo (1,2693, 1,3636), mientras que los que tienen sentimiento negativo se encuentran en el intervalo (1,0350, 1,0664).
 - Por último, calcula la diferencia que hay entre la puntuación media de los dos sentimientos (0,2657). Esto se calcula restandole a la media de valor de sentimiento positivo la media de valor de sentimiento negativo.
- c) La siguiente gráfica representa como están divididos los tuits teniendo en consideración la puntuación conjunta de sentimiento positivo y de sentimiento negativo. En el eje horizontal encontramos los diferentes grados de intensidad que existen del sentimiento positivo y en el eje vertical encontramos los grados de intensidad del sentimiento negativo.
- Como se puede observar, el cuadrado de mayor tamaño expresa que la mayoría de los tuits recolectados tienen una puntuación positiva de 1 y una puntuación negativa de 1 en cuanto a su intensidad, por lo que la puntuación conjunta de la mayoría de nuestros tuits es (+1,-1).
- La segunda combinación que más se repite es (+3,-1).

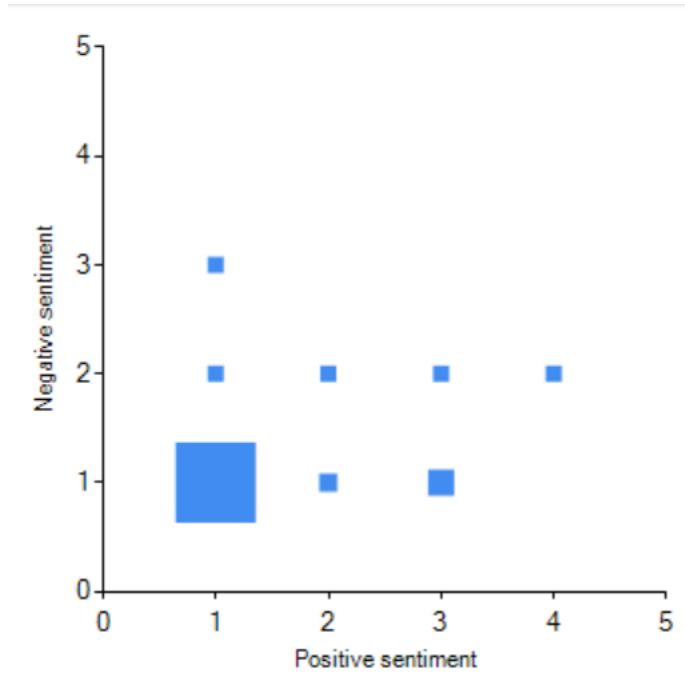


Figura 4.27. Captura de gráfica de la puntuación conjunta de los tuits

Fuente: Captura de pantalla de la herramienta Mozdeh

En resumen, podemos comprobar que hay un sentimiento positivo mayoritario hacia la implementación del Certificado COVID Digital de la UE.

CAPÍTULO 5: CONCLUSIONES

Como hemos comentado a lo largo de los capítulos anteriores, sobre todo en el apartado “Justificación del tema”, la implantación del Certificado COVID digital para la UE podía ser un punto de inflexión para la industria turística. Pasados unos meses desde el 1 de julio, día en el que se empezó a aplicar el Reglamento sobre el Certificado COVID de manera oficial en Europa, se ha podido observar de qué manera ha afectado al sector turístico y qué previsiones hay y, sobre todo, ha servido para conocer cuál ha sido la reacción en las calles que ha habido por parte de la población europea ante la obligación de portar este certificado para viajar, e incluso para asistir a algunos establecimientos y eventos. Además, hemos podido descubrir si esta reacción popular ha sido parecida a lo que hemos hallado en nuestra investigación.

Se trata de un estudio que ha servido para averiguar y prever la reacción de la opinión pública ante la implantación del Certificado COVID y la obligatoriedad de portarlo para viajar, e incluso en algunos territorios europeos para realizar actos cotidianos. Esto se traduce en un trabajo que puede ser útil para aquellos investigadores del sector turístico que quieran conocer cuál es el nivel de influencia que tiene la implantación del Certificado COVID sobre el mercado, y así poder intuir cómo se puede comportar el mercado en un futuro inmediato. Pero, además, se trata de un trabajo que ha explicado paso por paso como realizar un correcto uso de la técnica de análisis de sentimientos, aportando las herramientas necesarias, además de cómo usarlas, por lo que es útil para todas aquellas personas que quieran aprender a usar esta técnica y no tengan los conocimientos necesarios.

La investigación ha arrojado los siguientes resultados: la mayoría de las puntuaciones positivas (81,28%) se encuentran en la puntuación 1 dentro de la escala de intensidad. De igual manera, aunque en mayor medida, la mayoría de las puntuaciones negativas (95,17%) se encuentran en la puntuación 1, siendo la combinación que más repite (+1,-1). La gran diferencia entre ambos sentimientos se encuentra en la puntuación 3, la cual representa un 12,44% para el sentimiento positivo y sólo un 0,24% para el pensamiento negativo, algo residual. Podemos ver que donde más se iguala la intensidad de ambos sentimientos es en la puntuación 2 con un 6,04% de las opiniones positivas y un 4,59% de las opiniones negativas.

De estos resultados podemos concluir que la mayoría de las opiniones se encuentran en una situación que podríamos denominar de neutralidad, pero por lo general, con ligera tendencia positiva como muestra la puntuación 3 en ambos sentimientos. Aunque no hay que dejar pasar por alto a ese segmento de población que, aunque es bastante minoritario (un 4,59% en la puntuación 2 y un 0,24% en la puntuación 3) tienen una opinión negativa con respecto al Certificado. Comparando los resultados del estudio con los acontecimientos que han ido sucediendo en forma de reacciones populares ante el anuncio de nuevas medidas relacionadas con el Certificado, podemos observar que la población se ha comportado de una manera similar a lo que arrojaron los resultados, esto es; se han dado manifestaciones en contra de la obligación de portar el Certificado COVID en distintos países europeos, como por ejemplo, en Francia con movilizaciones que reunieron a unas 200.000 personas a mediados de julio y que se han venido repitiendo cada sábado hasta que a finales de octubre sólo participaron unas 40.000 personas (EFE, 2021), igual en Italia, donde también existieron movimientos populares pero no consiguieron los bloqueos económicos esperados (Pacho, 2021). Estas cifras de personas que están en

desacuerdo con la implantación del Certificado COVID han representado un porcentaje muy bajo en comparación con el total de la población de sus respectivos países, por lo que entendemos que la gran mayoría de los ciudadanos están a favor del Certificado COVID, o al menos, no están en contra.

Por otra parte, aconsejamos que se hagan estudios con cierta frecuencia a lo largo del tiempo, ya que la reacción pública puede sufrir variaciones en función del uso que le den las autoridades pertinentes al Certificado COVID. Por poner ejemplos, en Francia es imprescindible que cada persona lo porte para ir al cine, a un bar, a un concierto o a un estadio de fútbol, o como Italia, que fue el primer país occidental en imponer la vacunación a todos los trabajadores (Verdú, 2021), obligándoles así, a tener en regla el Certificado COVID para poder trabajar. Sin embargo, en España por lo general, las exigencias de portar el certificado son más suaves. Por lo que entendemos que los resultados podrían variar si se endurecen las medidas.

Pero ¿De qué manera y en qué proporción está afectando al turismo la implantación del Certificado COVID Digital de la UE? La aplicación del certificado está sufriendo inconveniencias estructurales que tienen como consecuencia el no poder aprovechar al máximo sus ventajas. Aunque es obvio que ha ayudado a relanzar al sector, también es cierto que la cuota de mercado de las empresas se va a ver reducida y va a ser cambiante, ya que el segmento de ciudadanos que están en contra del Certificado COVID dejará de viajar y de consumir en establecimientos de aquellas regiones que lo soliciten, así como, aquellos ciudadanos que estén vacunados y porten el Certificado se sentirán más seguros en sitios donde lo pidan, lo que puede generar desigualdades en la demanda por factores ajenos y poco controlables por las empresas turísticas. Por lo que, es importante que una vez que se ha instaurado este certificado, se respete al máximo el principio por el que se creó; estandarizar las regulaciones para aprovechar al máximo las ventajas que ofrece, no solo en las distintas comunidades autónomas españolas, también en los distintos países de la Unión Europea, hecho que hasta ahora no se está cumpliendo según la Asociación de Transporte Aéreo Internacional (IATA) (Vialarasau, 2021).

Visto lo anterior, cabe hacerse unas preguntas. ¿Atenta el Certificado COVID Digital de la UE contra la intimidad de las personas? ¿Ejerce un trato discriminatorio? ¿Vale la pena realmente su implementación? ¿Hasta cuándo permanecerá con nosotros?

Para empezar, cuando se dio a conocer esta medida al mundo, se habló de “libre circulación de los ciudadanos entre los países miembros”, no se especificó que iba a ser necesario para realizar acciones cotidianas como trabajar, como hemos visto que ocurre en algunos países europeos, lo que ha podido generar de entrada una controversia. En cuanto si atenta contra la intimidad de las personas, es cierto que, se baraja el escenario de tener que mostrar el Certificado COVID junto con una serie de datos personales cada vez que se quieran ejercer ciertas actividades cotidianas, dando la posibilidad a las autoridades de ejercer un mayor control sobre los ciudadanos, y teniendo una serie de datos que podrían llegar a terceros interesados. Por lo que debería existir un gran control en cuanto a la protección de datos y a la divulgación de estos. Aunque es cierto que, hoy día le damos la potestad -casi inconscientemente- de recolectar una gran cantidad de datos y movimientos personales a compañías e instituciones a través de los aparatos electrónicos que nos rodean, sobre todo a través del teléfono móvil. (Cordero, 2021)

Por otra parte, pienso que es una medida discriminatoria, ya que aún hay muchos países cuyas poblaciones no han tenido la posibilidad de acceder completamente a las vacunas y que tienen que pagar por test de detección del Covid-19, o que han sido vacunados con vacunas de marcas que no han sido aceptadas a nivel internacional.

Con todo esto, estoy convencido de que merece la pena su implementación como primer paso para intentar conseguir una actividad similar a la época prepandemia, siempre y cuando se mejore la unificación de criterios entre países, se use éticamente (dentro de lo posible) y no rompa derechos fundamentales. Los datos proporcionados por Hosteltur (Ortega, 2021) indican que Europa es líder en cuanto al número de viajes en el tercer trimestre de 2021, gracias a que tiene el mayor porcentaje de vacunación y por la creación del Certificado COVID Digital de la UE.

A corto plazo, se sabe que cada persona poseerá el certificado al menos 12 meses desde la inoculación de la segunda dosis de la vacuna, momento en el que tendrá que inocularse otra dosis de refuerzo, aunque en algunos países como Austria se está barajando rebajar el número de meses hasta los 9. A largo plazo, da la sensación de que esto sólo es el principio, porque aún los países están incorporándolo a nuevas medidas y su uso no para de crecer. Además, aún prevalece una parte significativa de la población que no está vacunada en Europa (casi el 40% no lo está) y que parece que por el momento no lo va a hacer. (Almería, 2021) Y en el mundo hay muchos países con pocos recursos y con tasas de vacunación muy bajas.

Limitaciones del estudio

El análisis de sentimientos es un fenómeno relativamente nuevo, por lo que, como es lógico, posee sus limitaciones. En este apartado vamos a ver las complicaciones e inconvenientes más frecuentes que se han encontrado los investigadores a la hora de aplicar la técnica de análisis de sentimientos en sus diferentes fases.

- Limitaciones en la recolección de datos: aunque haya una gran cantidad de datos, a veces resulta complicado acceder a ellos, ya que por una parte existe una industria que hace negocio y, porque aún existe una cierta falta de competencia entre los investigadores para extraer datos desde las plataformas. La suma de estos dos puntos anteriores deriva a que en muchos casos sea necesaria la ayuda de un intermediario que facilite los datos precisos, pero estos, en muchas ocasiones son insuficientes o cuentan con una falta de transparencia en su proceso.
- Limitaciones de Procesamiento y Análisis: la mayoría de las complicaciones que desembocan en la limitación del correcto procesamiento y análisis del método tiene que ver con la mala interpretación de textos y palabras. Las limitaciones más frecuentes suelen ser:
 - La correcta detección de la ironía y/o el sarcasmo. (Son, y otros, 2019) (Deepak, Kumar, & Garg, 2020) (Karthik & Anandhakumar, 2020).
 - La existencia de algoritmos mejor preparados para detectar oraciones positivas que negativas o neutras, lo que puede desembocar en estudios no del todo precisos. (Reza, Becken, & Stantic, Sentiment Analysis in Tourism: Capitalizing on Big Data, 2019). Además, resulta particularmente complicado detectar la neutralidad en comparación con la negatividad y la positividad.
 - La mayoría de los recursos del análisis de sentimientos están creados para el idioma inglés, discriminando otros idiomas muy hablados como español, francés o chino.
 - Dificultad para detectar la jerga, la cual es muy utilizada en las redes sociales, así como, la tendencia a acortar palabras o cometer errores ortográficos.

Estudios futuros

En los siguientes puntos se han recopilado algunas ideas que podrían ser interesantes para ser objeto de estudio y de aplicación en el análisis de sentimientos:

- La mayoría de los estudios que hemos visto al realizar este trabajo (este incluido) sólo utilizan una plataforma social como fuente de datos. Sería interesante combinar varias plataformas como Facebook y Twitter para recolectar la información.
- Tratar de descubrir a los denominados usuarios falsos que publican reseñas falsas para engañar o persuadir a otros usuarios.
- integrar varios tipos de datos para generar resultados de análisis de sentimientos a otra escala, esto sería, combinando puntuaciones de polaridad con otros datos como el clima, el transporte o el servicio prestado para comprender la trascendencia de cada uno de estos aspectos en la satisfacción del cliente. (Reza, Becken, & Stantic, 2019)

En cuanto al horizonte a largo plazo; se prevé que la computación cuántica protagonice la próxima revolución tecnológica. Este tipo de tecnología tendrá la virtud de analizar de una forma mucho más efectiva y rápida grandes bases de datos, lo que seguro, brindará más y mejores posibilidades al fenómeno del análisis de sentimientos. (G. Rodríguez, 2021).

Glosario de términos

Algoritmo: *“Un algoritmo informático es un conjunto de instrucciones definidas, ordenadas y acotadas para resolver un problema o realizar una tarea”.* (Maluenda De Vega, 2021)

Big Data: se trata de un conjunto de datos que por sus tres “V” (volumen, variabilidad y velocidad) se hace casi imposible de capturar, gestionar y analizar a partir de los procedimientos o tecnologías convencionales.

Hashtag: se trata de una palabra clave precedida por el símbolo de la almohadilla (#) que los usuarios utilizan para marcar o etiquetar el contenido que están compartiendo en las redes sociales. Estos hashtags son hipervínculos que conectan a todos los usuarios que están compartiendo el mismo hashtag, por lo que sirve como cadena o red para identificar a grupos de usuarios que estén compartiendo información semejante.

Hiperónimo: *“Palabra cuyo significado está incluido en el de otras. Pájaro es hiperónimo de jilguero y de gorrión”.* (Española, RAE, 2021)

Inteligencia artificial: se trata de aquellos procesos que realizan las máquinas y que simulan a los hechos por los humanos. Está compuesta por distintos niveles según su aplicación y alcance, pudiendo, desde resolver tareas bien acotadas y simples para el cerebro humano hasta obtener la capacidad de ser consciente de sí misma.

Lingüística computacional: Se trata de una disciplina híbrida que combina el estudio de las lenguas y la computación, mediante técnicas matemáticas y estadísticas y con el apoyo generalmente de equipos informáticos.

Minería de datos o Data mining: se trata del proceso de extraer datos como patrones, tendencias o reglas de una inmensa cantidad de datos heterogéneos o Big data que resultan de interés para la investigación que se quiera realizar a través de técnicas de aprendizaje automático. La información que se extrae de las bases de datos está estructurada.

Minería de textos o text mining: Es una variante de la minería de datos. Trata de extraer información a partir de grandes cantidades de texto donde los datos no están estructurados.

Retuit: Dentro de la red social Twitter, consiste en compartir un tuit de otro usuario pudiéndole añadir contenido propio.

Software de código abierto: se trata de aquel software informático cuyo código fuente, así como, otros de sus derechos, son de dominio público en lugar de estar reservados exclusivamente para quienes poseen los derechos de autor. Esto significa que cualquier usuario puede utilizar, cambiar, mejorar y redistribuir el software, lo que tiene consecuencias positivas, ya que muchas veces se consigue una gran mejora del software gracias a la colaboración de diferentes usuarios, pero también tiene sus aspectos negativos, puesto que resulta más complicado de mantener y de proteger.

Técnicas de procesamiento del lenguaje natural: trata de comunicar a las personas con las máquinas, consiguiendo que éstas entiendan el lenguaje natural de los humanos. Esto se consigue a través del análisis sintáctico, semántico, pragmático y

morfológico de las interacciones que se dan entre las computadoras y el lenguaje humano.

Tuits o tweets: *“Mensaje digital que se envía a través de la red social Twitter y que no puede rebasar un número limitado de caracteres”* (280). (Española, RAE, 2021)

Bibliografía

- Almería, C. d. (15 de Noviembre de 2021). *Centro de Documentación Europea de Almería*. Obtenido de Centro de Documentación Europea de Almería: <https://www.cde.ual.es/evolucion-vacunacion-covid-19-en-europa/>
- Alonso Almeida, M. d., Borrajo Millán, F., & Yi, L. (2019). Are Social Media Data Pushing Overtourism? The Case of Barcelona and Chinese Tourists. *MDPI*.
- Arimetrics. (10 de Noviembre de 2021). *Arimetrics*. Obtenido de Arimetrics: <https://www.arimetrics.com/glosario-digital/gephi>
- Baviera, T. (2016). Techniques for sentiment analysis in Twitter: Supervised Learning and SentiStrength. *DIGITOS*.
- Bello, E. (17 de Mayo de 2021). *IEBS*. Obtenido de IEBS: <https://www.iebschool.com/blog/historia-de-internet-innovacion/>
- Borrajo Millán, F., Alonso Almeida, M. d., & Escat Cortes, M. (2021). Sentiment Analysis to Measure Quality and Build Sustainability in Tourism Destinations. *MDPI*.
- Brzustewicz, P., & Singh, A. (2021). Sustainable Consumption in Consumer Behavior in the Time of COVID-19: Topic Modeling on Twitter Data Using LDA. *MDPI*.
- Cambria, E., & Amir, H. (2012). *Sentic Computing: Techniques, Tools, and Applications*. Springer Link.
- Casanueva, C., Gallego, Á., & García-Sánchez, M.-R. (2014). Análisis de redes sociales en turismo. *Temas de actualidad en turismo*, 1190-1209.
- Chen, M., & Jin, X. (2019). What do Airbnb users care about? An Analysis of online review comments. *International Journal of Hospitality Management*, 58-70.
- Cordero, P. (15 de Noviembre de 2021). *ConSalud.es*. Obtenido de ConSalud.es: https://www.consalud.es/pacientes/especial-coronavirus/pasaporte-covid-19-eficaz-frente-confinamiento-amenaza-libertades_104872_102.html
- Cordón García, Ó. (17 de marzo de 2014). *Redes y Sistemas Complejos*. Obtenido de <https://sci2s.ugr.es/sites/default/files/files/Teaching/GraduatesCourses/RedesSistemasComplejos/Tema03-RedesSociales-13-14.pdf>
- Cortizo Pérez, J. C. (2017). *BrainSINS*. Obtenido de <http://www.brainsins.com/es/blog/mineria-opiniones/3555#>
- Deepak, J., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. *Applied Soft Computing Journal*, Volumen 91, 106198.
- Echeverri Torres, M. M., & Manjarrés-Betancur, R. (2020). ASISTENTE VIRTUAL ACADÉMICO UTILIZANDO TECNOLOGÍAS COGNITIVAS DE PROCESAMIENTO DE LENGUAJE NATURAL. *Revista Politécnica ISSN 2256-5353*, 85-96.

- EFE. (15 de Noviembre de 2021). *Agencia EFE*. Obtenido de Agencia EFE: <https://www.efe.com/efe/espana/sociedad/las-protestas-contra-el-certificado-covid-reunen-a-40-000-personas-en-francia/10004-4659195>
- ElDiario. (8 de Junio de 2021). *elDiario.es*. Obtenido de El "pasaporte" COVID ya es una realidad: para qué sirve, cómo pedirlo y dónde: https://www.eldiario.es/sociedad/pasaporte-covid-realidad-sirve-pedirlo_1_8015958.html
- Española, R. A. (15 de Julio de 2021). *RAE*. Obtenido de RAE: <https://dle.rae.es/tuit>
- Española, R. A. (14 de julio de 2021). *RAE*. Obtenido de RAE: <https://dle.rae.es/hiper%C3%B3nimo>
- Estadística, I. N. (2021). *Cuenta Satélite del Turismo de España*. Obtenido de Cuenta Satélite del Turismo de España: <https://www.dataestur.es/economia/cst/>
- Europea, C. (7 de Julio de 2021). *Web Oficial de la Unión Europea*. Obtenido de Web Oficial de la Unión Europea: https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/safe-covid-19-vaccines-europeans/eu-digital-covid-certificate_es
- Europea, U. (19 de Julio de 2021). *Web Oficial de la Unión Europea*. Obtenido de Reabrir la UE: <https://reopen.europa.eu/es>
- Europeo, P. (15 de Julio de 2021). *Centro Multimedia*. Obtenido de Centro Multimedia: https://multimedia.europarl.europa.eu/en/eu-digital-covid-certificate_18602_pk?_package_detail_portlet_PackageDetailPortlet_selectedDate=07%2F07%2F2021
- Fernández, H. (25 de Junio de 2019). *EconomíaTIC*. Obtenido de EconomíaTIC: <https://economiatic.com/historia-de-internet/>
- Forero, T. (17 de Agosto de 2019). *rockcontent*. Obtenido de rockcontent: <https://rockcontent.com/es/blog/historia-del-internet/>
- G. Rodríguez, A. (18 de Julio de 2021). *El Orden Mundial*. Obtenido de La carrera por dominar la computación cuántica, la próxima revolución tecnológica: <https://elordenmundial.com/carrera-dominar-computacion-cuantica-revolucion-tecnologica/>
- Gour, A., Aggarwal, S., & Erdem, M. (2021). Lectura entre líneas: análisis de reseñas en línea mediante un enfoque de análisis web de múltiples métodos. *Revista internacional de gestión hotelera contemporánea*, Volumen 33, Número 2.
- Green, A. (2021). *Varonis*. Obtenido de <https://www.varonis.com/blog/what-is-user-behavior-analytics/>
- Gutierrez, H. (21 de Febrero de 2021). *El País*. Obtenido de El País: <https://elpais.com/economia/2021-02-03/el-turismo-cifra-su-debacle-espana-recibio-189-millones-de-viajeros-extranjeros-en-2020-un-77-menos.html>
- Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 1495-1545.

- Herrera Bruselas, E. (6 de Julio de 2021). Bruselas recomienda usar el pasaporte covid también en eventos masivos. *La Vanguardia*.
- HOSTELTUR. (12 de Diciembre de 2020). *HOSTELTUR*. Obtenido de HOSTELTUR: https://www.hosteltur.com/141109_el-turismo-aumento-su-aportacion-al-pib-en-2019-hasta-el-124.html
- Jiang, W., Zhengan, X., Su, Q., Long, Y., Song, X., & Sung, P. (2021). Using Geotagged Social Media Data to Explore Sentiment Changes in Tourist Flow: A Spatiotemporal Analytical Framework. *MDPI*.
- Kammenov, Y., & Sychev, O. A. (2020). Eligibility of English Hypernymy Resources for Extracting Knowledge from Natural-Language Texts. *11th Annual International Conference on Brain-Inspired Cognitive Architectures for Artificial Intelligence* (págs. 501-507). Natañ, Brasil: Springer Science and Business Media Deutschland GmbH.
- Karthik, S., & Anandhakumar, P. (2020). Multi-Rule Based Ensemble Feature Selection Model for Sarcasm Type Detection in Twitter. *Computational Intelligence and Neuroscience*, Article ID 2860479.
- Kim, K., Park, O.-J., Yun, S., & Yun, H. (2017). What makes tourists feel negatively about tourism destinations? Application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, 362-369.
- Kumar, J., Raymond, K.-K., Kumar, A., Bakshi, S., Kumar, S., & Williams, K. (2017). A model for sentiment and emotion analysis of unstructured social media text. *Electronic Commerce Research*, 181-199.
- López Avellaneda, D. (s.f.). *Matemáticas IES*. Obtenido de Matemáticas IES: <https://matematicasies.com/Matriz-de-adyacencia-de-un-grafo>
- Malasia, D. d. (7 de Julio de 2021). *Delegación de la Unión Europea en Malasia*. Obtenido de Comunicado de prensa de la Delegación de la Unión Europea en Malasia: https://eeas.europa.eu/delegations/malaysia/101413/press-statement-delegation-european-union-malaysia_en
- Maluenda De Vega, R. (22 de Julio de 2021). *Profile*. Obtenido de Profile: <https://profile.es/blog/que-es-un-algoritmo-informatico/>
- Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *Journal of Destination Marketing and Management*, Volumen 20, Artículo nº 100598.
- Molina, C. (30 de Agosto de 2019). *El País*. Obtenido de El País: https://cincodias.elpais.com/cincodias/2019/08/29/companias/1567086634_731386.html
- Moreno, A. (20 de Julio de 2021). *Instituto de Ingeniería del conocimiento*. Obtenido de Instituto de Ingeniería del conocimiento: <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>
- OroyFinanzas.com. (17 de Julio de 2021). *OroyFinanzas.com*. Obtenido de OroyFinanzas.com: <https://www.royfinanzas.com/2014/11/bitcoin-que-es-software-codigo-abierto-open-source/>

- Ortega, J. L. (15 de Noviembre de 2021). *Hosteltur*. Obtenido de Hosteltur: https://www.hosteltur.com/147866_los-viajes-en-europa-no-recuperaran-los-niveles-prepandemia-hasta-2024.html
- Pacho, L. (15 de Noviembre de 2021). *El País*. Obtenido de El País: <https://elpais.com/internacional/2021-10-15/las-protestas-contras-el-certificado-covid-se-desinflan-en-italia.html>
- Pang, B., & Lee, L. (2008). *Opinion Mining and sentiment analysis*. Boston: NOW the essence of knowledge.
- Patrarca, D., & Beatriz. (20 de Septiembre de 2015). *PRÁCTICA DE REDES SOCIALES GEPHI*. Obtenido de PRÁCTICA DE REDES SOCIALES GEPHI: <https://pagines.uab.cat/joseluismolina/sites/pagines.uab.cat/joseluismolina/files/PR%C3%81CTICA%20DE%20REDES%20SOCIALES%20GEPHI.pdf>
- Pérez, J., & Gardey, A. (2010). *Definición.DE*. Obtenido de <https://definicion.de/web-2-0/>
- Peter, S. D., E. M., Suma, D., Morgan, R. F., & Reagan, A. J. (2015). Human language reveals a universal positivity bias. *PNAS*, 2389-2394.
- Ramanathan, V., & Meyyappan, T. (2019). *IEEE. Twitter Text Mining for Sentiment Analysis on People's Feedback about Oman Tourism*. ieee.
- Ramiro, R. (30 de Agosto de 2017). *ciberseguridad.blog*. Obtenido de <https://ciberseguridad.blog/ueba-user-and-entity-behavior-analytics-deteccion-por-comportamiento/>
- Recuero De los Santos, P. (19 de Julio de 2021). *Thing Big*. Obtenido de Telefónica: <https://empresas.blogthinkbig.com/que-algoritmo-elegir-en-ml-aprendizaje/>
- Reuters, T. (8 de Julio de 2021). *EURONEWS*. Obtenido de EURONEWS: <https://es.euronews.com/2021/07/08/salud-coronavirus-ue-suiza>
- Reyes-Menéndez, A., Ramón Saura, J., & Filipe, F. (2019). The importance of behavioral data to identify online fake reviews for tourism businesses: a systematic review. *PeerJ Computer Science*.
- Reza, A., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism: Capitalizing on Big Data. *SAGE*, 175-191.
- Reza, A., Becken, S., & Stantic, B. (2019). Sentiment Analysis in Tourism; Capitalizing on Big Data. *SAGE journals*, 175-191.
- Ricaurte Quijano, P., & Ramos-Vidal, I. (2015). Investigación en redes sociales digitales: consideraciones metodológicas desde el paradigma estructural. *Revista Virtualis*, 165-194.
- Salas Narváez, L., Jiménez Cercado, M. E., & Acosta Vélez, M. (2016). Aproximación a los comentarios on-line de los consumidores de los hoteles de Guayaquil. *RICIT*, Número 10.
- Smith, M., Rainie, L., Himelboin, I., & Shneiderman, B. (20 de Febrero de 2014). Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. *Pew Research Center*.

- Sobrino Sande, J. C. (Junio de 2018). Análisis de Sentimientos en Twitter. *Análisis de Sentimientos en Twitter*.
- Son, L. H., Kumar, A., Sangwan, S. R., Arora, A., Nayyar, A., & Abdel-Basset, M. (2019). Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE*, 23319-23328.
- Stuart, O., Ben, F., Linden, P., Aurina, A., Chao, S., & Alex, F. (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS ONE*.
- University of Illinois at Chicago, U. S. (2012). Sentiment Analysis and opinion mining. pág. 1.184.
- Verdú, D. (15 de Noviembre de 2021). *El País*. Obtenido de El País: <https://elpais.com/sociedad/2021-09-16/italia-se-convierte-en-el-primer-pais-occidental-en-imponer-la-vacunacion-a-todos-los-trabajadores.html>
- Vialarasau, D. R. (15 de Noviembre de 2021). *Hosteltur*. Obtenido de Hosteltur: https://www.hosteltur.com/146445_tantas-reglas-de-viaje-europeas-anulan-las-ventajas-del-certificado-covid.html
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wolverhampton, G. d. (11 de Noviembre de 2021). *Mozdeh*. Obtenido de Mozdeh: <http://mozdeh.wlv.ac.uk/>
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 51-65.