UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

LA ESCALA DE RUPTURA: MEDICIÓN DEL IMPACTO DE LAS OPERACIONES DE DESINFORMACION

Trabajo Fin de Máster

Máster Universitario en Ingeniería Informática

AUTOR/A: Corral Sastre, Antonio

Tutor/a: Rebollo Pedruelo, Miguel

CURSO ACADÉMICO: 2022/2023

Resumen

En este trabajo se van a analizar las técnicas más utilizadas (teoría de grafos, procesamiento del lenguaje natural, aprendizaje automático, entre otras) en el análisis de información tanto en redes sociales como en medios tradicionales para poder clasificar casos de operación de influencia dentro de la escala de ruptura y analizar su impacto.

Palabras clave: noticias falsas, operaciones de influencia, escala de ruptura, grafo, detección comunidades, modelo de tópicos, modelo SEIZ

Resum

En aquest treball van a analitzarse les tècniques més utilitzades (teoria de grafs, processament del llenguatge natural, aprenentatge automàtic, entre altres) en l'análisis d'informació tant en xarxes socials com en medis tradicionals per a poder clasificar casos d'operació d'influencia dintre de l'escala de ruptura i analitzar el seu impacte.

Paraules clau: noticies falses, operacions d'influencia, escala de ruptura, graf, detecció comunitats, model de tópics, model SEIZ

Abstract

In this paper we analyze the most used techniques (graph theory, natural language processing, machine learning, among others) in the analysis of information both in social networks and in traditional media in order to classify cases of influence operation within the breakout scale and analyze his impact.

Keywords: fake news, influence operations, the breakout scale, graph, community detection, topic model, SEIZ model

Dedicado a mi familia
y a mi tutor

Lista de algoritmos

Lista de figuras

Lista de tablas

Índice

Pág.

1. Introducción

La propaganda y la difusión de información falsa con el objetivo de manipular la opinión pública no es algo nuevo. Sin embargo, el auge de las comunicaciones digitales y las redes sociales ha puesto a disposición de cualquier persona u organización una herramienta de bajo coste y gran escalabilidad y con posibilidades de automatización con la que diseñar y lanzar campañas de desinformación. La gran rapidez con la que se transmite la información por las redes y el hecho de que las opiniones negativas tienen más facilidad de difusión que las positivas hacen de las redes sociales un medio idóneo para el inicio de estos procesos.

La difusión de mensajes falsos se vale de la facilidad para manipular imágenes y vídeos, de la confianza que tenemos en nuestros contactos, la dificultad de comprobar toda la información que nos llega por su sobreabundancia y el afán por obtener el reconocimiento de ser la primera persona en destapar ciertos asuntos o conseguir visibilidad al convertir los mensajes en virales.

Por eso es importante detectar este tipo de mensajes lo antes posible para neutralizar su efecto, dificultando su transmisión, o utilizando la misma red que los ha propagado para difundir la información veraz. Pero para estimar el alcance y determinar su impacto es necesario disponer de alguna medida con la que cuantificarlo. Para cuantificarlo, nos vamos a apoyar en la escala de ruptura.

1.1. Motivación

La motivación de este trabajo es la de analizar noticias o temáticas susceptibles a difundirse como noticias/mensajes/información falsa (*fake news*), además de medir el impacto de estas campañas de influencia en la sociedad. El acceso a cada vez más cantidad de información (sobre todo a través de Internet) ha hecho que sea un caldo de cultivo para las noticias falsas, seguramente uno de los mayores problemas en el día de hoy a la hora de informarse correctamente.

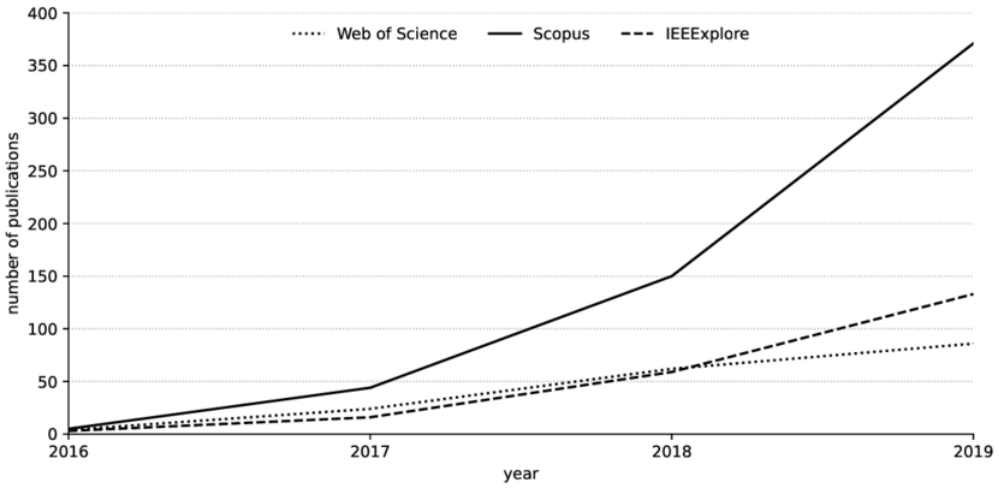


Figura 1: Estudios de noticias falsas por año en Web of Science, Scopus y IEEE Xplore [CHO20]

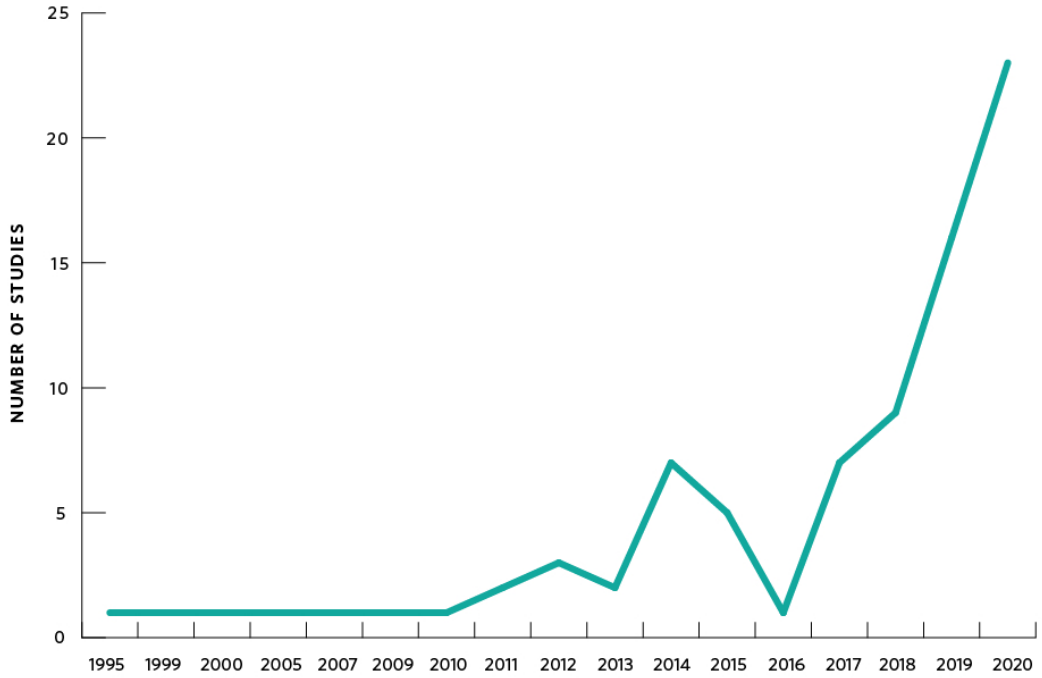


Figura 2: Estudios de operaciones de influencia por año [BAT21]

1.2. Objetivos

El objetivo principal del trabajo es analizar los pasos que realizaría una aplicación que, dada la búsqueda de una temática, realice un análisis del mismo mediante técnicas del estado del arte tanto en las redes sociales como en los medios de información tradicionales, para que podamos deter-

minar en que categoría de la escala de ruptura se encuentra.

Los problemas que nos vamos a encontrar a la hora de clasificar las operaciones de influencia en la escala de ruptura son identificar comunidades, identificar actores influyentes y como se difunde de la información en un grafo que represente una red social. Además, vamos a tener que identificar de que se está hablando (tanto en las redes sociales como en los medios tradicionales) para saber si se está hablando de la temática que nos interesa analizar y si esa temática que estamos analizando ha desencadenado en violencia o en una acción política.

Para buscar solución a los problemas anteriores, se han definido una serie de subobjetivos a cumplir, los cuales son:

1. Construir un repositorio con los mensajes y textos sobre un tema.
2. Clasificar una operación de influencia siguiendo las seis categorías de la escala de ruptura.
3. Emplear técnicas de análisis de redes sociales para construir la red de la operación de influencia y analizar su estructura.
4. Identificar los actores de la red claves en la transmisión de la operación de influencia.
5. Implementar un prototipo para validar la metodología propuesta con casos reales.
6. Predecir como va a comportarse una operación de influencia a través del tiempo.

1.3. Estructura de la memoria

La memoria empieza con la introducción a la situación actual de las noticias falsas y operaciones de influencia en el que se comentaran los hechos históricos a destacar relacionados con las noticias falsas y operaciones de influencia para saber como se ha llegado a la situación actual, los tipos de herramientas que hay hoy en día para combatir las noticias falsas y como

identificar los tipos de información que se pueden difundir, sean verídicas o falsas.

A continuación se explica la escala de ruptura y como identificar si una operación de influencia o noticia falsa va cumpliendo las categorías de la escala de forma progresiva hasta conocer la máxima categoría a la cual pertenece.

En el siguiente punto se expone el modelo epidemiológico SEIZ para analizar la difusión de la información y como detectar si en la temática que estamos analizando puede haber mucha o poca presencia de noticias falsas.

Por último, se realiza un análisis en la escala de ruptura de un caso de estudio del sabotaje del gasoducto Nord Stream. En él se analizan los resultados que se han ido encontrando a la hora de aplicar la metodología descrita en los puntos anteriores y como interpretarlos.

2. Estado del arte

Como introducción al estado del arte vamos a comentar los hechos más significativos a lo largo de la historia relacionados con las noticias falsas y campañas de influencia para conocer como se ha llegado al contexto actual. Un ejemplo de las primeras campañas de noticias falsas u operaciones de influencia conocidas se puede remontar hasta el antiguo Egipto con Ramsés II en la batalla de Kadesh [WFN23].

El primer salto importante a la hora de transmitir información sucedió con la invención de la imprenta a mediados del siglo XV. Anteriormente solo se podía transmitir la información con el boca a boca, manuscritos o libros mediante escritura y dibujos. Con la imprenta se hace más fácil realizar copias en papel, por lo que el papel se convierte en el medio de transmisión de información predominante. Poco después, en el siglo XVII se empieza a citar las fuentes a pie de página. A finales de siglo XIX se inventa la radio, que permite la transmisión de información prácticamente en tiempo real, además de acuñarse el término *fake news*.

El siguiente salto importante ocurre con la aparición de la televisión a mediados del siglo XX. Se transmite tanto sonido como imagen a tiempo real, juntando las ventajas de la imprenta y la radio. Al mismo tiempo surge la teoría de los seis grados de separación, que dice que una persona está conectada con cualquier otra persona en el mundo a través de una cadena de máximo cinco personas intermediarias.

El último salto ha sido la universalización de Internet en la última década del siglo XX donde cualquier persona puede dar su opinión de forma "libre" y compartirla con otra gente. A partir de aquí, Internet será el medio de transmisión de información más utilizado. Una década después se popularizarán las redes sociales. En 2016, Facebook estimó que la media de grados de separación era de 3.5 [1]. Como cada vez la gente utiliza más las

[1] https://cronicaglobal.elespanol.com/vida/estas-a-seis-grados-de-cualquier-persona-en-facebook-solo-a-3-57_33245_102.html

redes sociales, no es extraño suponer que a día de hoy el número sea aún más bajo.

El presente y futuro de las noticias falsas, como muchos otros campos, será guiado por la inteligencia artificial. En los próximos años la gente ya no podrá distinguir *deepfakes* y cualquiera podrá imitar la voz de alguien con *fake voice*. Además, a día de hoy cualquiera puede generar noticias falsas en forma de texto completamente creíbles con herramientas de dominio público como ChatGPT. La inteligencia artificial está más avanzada con las herramientas que favorecen las noticias falsas que las que lo combaten. Un ejemplo lo encontramos en el mismo ChatGPT, donde OpenAI, la empresa que lo creo, ha creado una herramienta de inteligencia artificial para detectar textos generados con ChatGPT con una eficacia del 26 % [2].

Figura 3: Ejemplos de noticias de operaciones de influencia. Fuente: elEconomista.es[3]y ElMundo[4]

Una de las cosas más curiosas que podemos observar analizando la historia es que a lo largo del tiempo la gente ha ido ganando acceso a mucha más información, cosa que tiene ventajas como el progreso y la universalización de conocimiento. Por ejemplo, anteriormente si querías ver como era el Taj Mahal tenías que ir a la India o ver unas pocas fotos a las que podías tener acceso. Hoy en día es muy fácil hacer un tour virtual por el Taj Mahal a

[2]https://openai.com/blog/new-ai-classifier-for-indicating-ai-written-text/

[3]https://www.eleconomista.es/deporte-negocio/noticias/10361795/02/20/El-Bara-paga-para-mejorar-la-imagen-del-club-y-desprestigiar-a-jugadores-y-rivales.html

[4]https://www.elmundo.es/cataluna/2021/11/09/61896962e4d4d840138b458a.html

través de Internet o en un documental de la televisión.

Aunque también tiene sus inconvenientes como la saturación de información. El ser humano no puede asimilar y ni mucho menos verificar si es o no correcta toda la información a la cual es expuesto. Por otra parte, hay gente que directamente tergiversa tanto la información que básicamente miente y difunde bulos en beneficio de sus intereses.

Por último, estaría el acceso a información que es explicada de forma sesgada o difusa. Se puede decir lo mismo (sea verdad o mentira) de muchas formas distintas y según en la forma con la cual lo expresemos, puede influir en la gente de una forma u otra. Un ejemplo sería que no es lo mismo decir que hay un 87 % de empleados en España que hay un 13 % de desempleados en España y que es el porcentaje de desempleo más alto de la eurozona. Esto ha proliferado en los últimos años en las redes sociales para ganar seguidores o que aparezcan más veces en el algoritmo de búsqueda. Otra técnica para mantener a usuarios en su comunidad es exponer al usuario información sesgada con la que se sienta a gusto para que no abandone la comunidad o la red social, conocido como el filtro de la burbuja.

Aunque la gente sea más accesible debido a internet y haya surgido la figura del *influencer*, eso no implica que sea más fácil hacer que cambien de opinión y puede llegar a ser una tarea costosa. Sin embargo, ¿para qué necesitas dedicar tiempo de tu vida en hacerlo cuando puedes programar un programa (bot) que simule ser una persona en Internet? Por poner un ejemplo, en Twitter se estima que entre un 5 y 20 % de cuentas activas son bots [5].

Anteriormente, se decía el dicho de "quien tiene la información tiene el poder". A día de hoy este dicho ha perdido el sentido, ya que en la práctica todo el mundo tiene acceso a la misma información. Aquel que sabe interpretar y utilizar esa información en su beneficio es el que a día de hoy tendrá el poder. Este ha sido el motivo por el cual la última década ha habido un incremento en el interés por el *big data* para la recolección de datos,

[5] https://www.businessinsider.com/twitter-bots-comprise-less-than-5-but-tweet-more-2022-9

ciencia de datos (*data science*) y aprendizaje automático (*machine learning*) para interpretar los datos.

Actualmente dos tipos de herramientas hay para combatir las noticias falsas. La primera consiste en la combinación de herramientas de aprendizaje automático en conjunto con herramientas de procesado del lenguaje natural (*natural language processing*). En caso de tratar con redes sociales se pueden añadir análisis de redes sociales (*social network analysis*) [LEE21]. La segunda consiste en la verificación de datos (*fact-check*), es decir, un conjunto de hechos contrastados de forma manual que se pueden consultar y están relacionados con una temática.

Algunos ejemplos de herramientas *fact-check* son el Fact Check Explorer de Google o el Fact Check de Newtral, entre otros. Otras herramientas que combaten las noticias falsas y no se basan únicamente en el *fact-check* son las listadas por la RAND Corporation[6] o la metodología DebunkEU[7] (una herramienta que se basa en la escala de ruptura).

Por último, comentar algunos de los tipos de información que se pueden difundir hoy en día para identificar a qué tipo pertenecen una vez contrastada la información y conocido el contexto. Se suelen diferenciar entre sí por si la información es verídica y la intencionalidad de la misma, como se puede observar en la tabla 1.

	Autenticidad	Intención	Noticias?
Noticias falsas	Falso	Desconocido	Sí
Satira	Desconocido	No es mala	Sí
Desinformación	Falso	Mala	Desconocido
Malinformación	Falso	Desconocido	Desconocido
Rumor	Desconocido	Desconocido	Desconocido
Clickbait	Desconocido	Mala	Desconocido

Tabla 1: Características de tipos de información [ZHO21]

[6] https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html
[7] https://www.debunkeu.org/methodology

La desinformación [CSI23] es la difusión deliberada de información falsa o inexacta con el fin de desacreditar a una persona u organización, mientras que la malinformación *misinformation* es el intercambio de información inexacta y engañosa de manera no intencional. Por último, el mal uso de información se le conoce como *malinformation*.

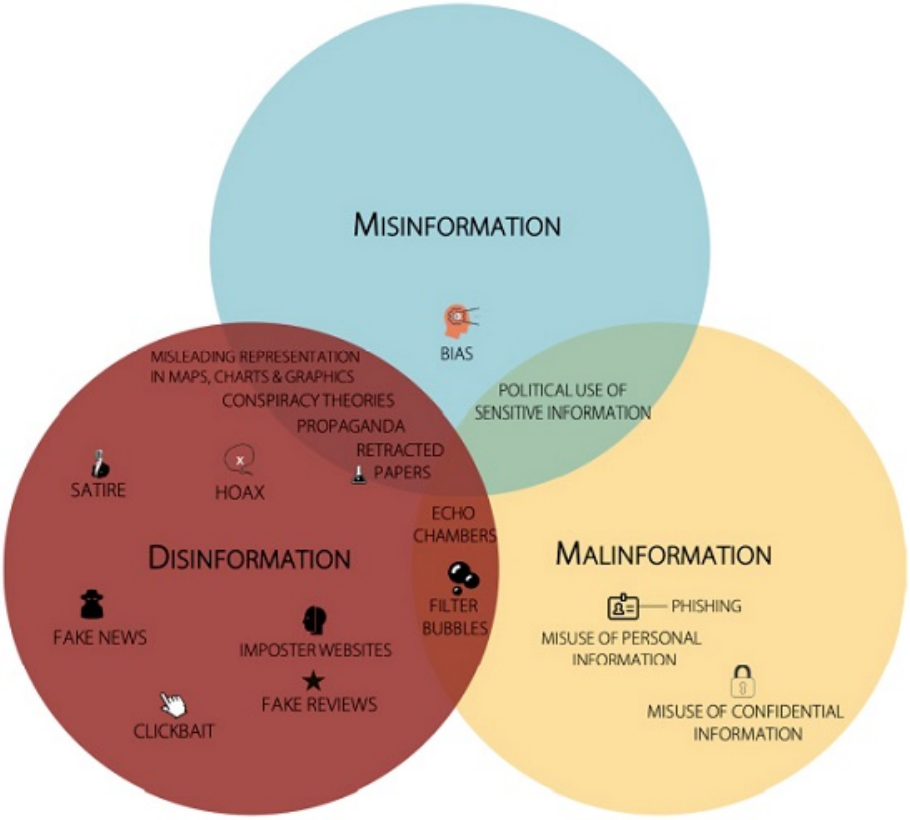


Figura 4: Clasificación de tipos de información según [KSD21]

3. La escala de ruptura

Las operaciones de influencia (*influence operations*) se definen como los esfuerzos de influenciar en el debate público/político o los procesos de toma de decisión que confían en parte o en su totalidad en actividad encubierta. En resumen, acciones que pretenden conseguir un efecto concreto sobre una audiencia.

Esto implica que las operaciones de influencia no tienen por qué ser siempre acciones donde se propagan noticias falsas, aunque es así en la mayoría de casos.

La escala de ruptura (*the breakout scale*) se define como un modelo comparativo para medir operaciones de influencia basadas en información que es observable, replicable, puede ser verificada y está disponible desde el momento en el que la información ha sido publicada. Fue desarrollada por Ben Nimmo [NIM20].

La escala de ruptura divide las operaciones de influencia en las siguientes seis categorías, donde una categoría cumple todas las categorías anteriores:

1. Las operaciones solo se extienden a través de una comunidad y en una plataforma.
2. Las operaciones se extienden en varias comunidades en la misma plataforma o en una comunidad en varias plataformas.
3. Las operaciones se extienden en varias plataformas y en varias comunidades.
4. Las operaciones se salen de las redes sociales y llegan a los medios tradicionales.
5. Las operaciones son difundidas por individuos muy influyentes como políticos o celebridades.
6. Desencadena una respuesta política u otra forma de acción concreta o si hace una llamada a la violencia.

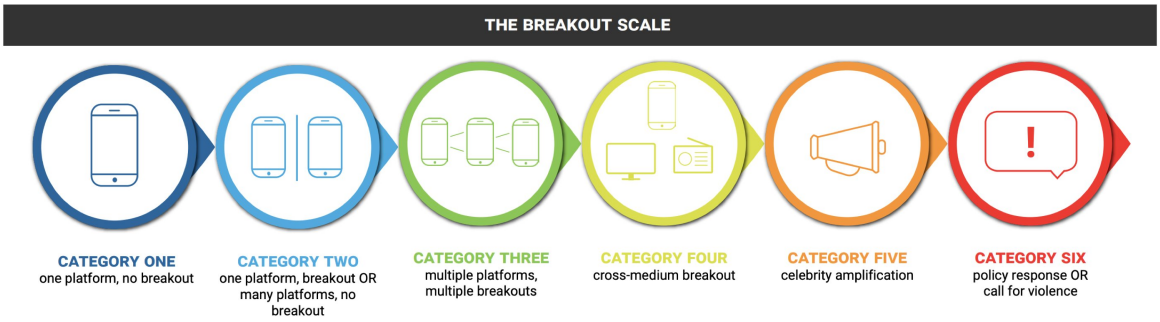


Figura 5: Categorías de la escala de ruptura

Dicho esto, hay que tener en cuenta que para poder realizar la clasificación, necesitaremos información de mínimo dos redes sociales y un medio tradicional.

3.1. Primera categoría: una comunidad en una plataforma

Para saber si una temática estaría dentro de esta categoría, bastaría con una búsqueda acotada en un tiempo determinado en una red social para asegurarnos de que nos devuelve información del tema que nos interesa analizar. Habría que contrastar que la búsqueda devuelve resultados específicos y no generales. Por ejemplo, si quiero saber como ha ido el partido de esta semana del Real Madrid no voy a realizar una búsqueda solo del Real Madrid esta semana, ya que del Real Madrid se hablan de más cosas aparte de su partido de la semana. Ese problema se abordará en la tercera categoría de la escala de ruptura. Por este motivo se recomienda tener un mínimo de conocimiento sobre la temática que vamos a analizar a la hora de realizar la búsqueda.

La gran mayoría de plataformas online tienen características o componentes típicos de una red social. De hecho, las que analizaremos permiten comentarios de la comunidad con el fin de fomentar la interacción entre los usuarios y crear comunidad (seguramente los dos aspectos más importantes en una red social).

Para recolectar información de las redes sociales se pueden utilizar tanto API o *web scraping* (ya sea interaccionando con los elementos del DOM de

la página como por API en el *frontend*). Para hacer *web scraping* es recomendable leer los términos y condiciones del servicio, además del archivo robots.txt del servidor, puesto que en muchas de ellas su uso es ilegal para evitar sobrecarga en los servidores que prestan el servicio. *Web scraping* puede ser útil si, por ejemplo, se quieren recuperar *tweets* más antiguos de 7 días, ya que mediante la API de Twitter eso no se puede realizar. Si una página web se permite *web scraping*, sería interesante utilizar, por ejemplo, Google Search Console API para realizar búsquedas en Google sobre esa noticia y hacer *web scraping* las páginas que salgan como resultado.

No en todas se puede extraer la información necesaria para efectuar un estudio de la misma. Incluso en algunas, gran parte de la información que se puede sacar son imágenes/videos. Para este trabajo se han estudiado como candidatas las redes sociales reflejadas en la tabla 2.

Redes sociales	Descripción	API oficial
Twitter	Post	Sí
YouTube	Vídeos	Sí
Menéame	Post	Sí
Telegram	Chats, grupos y canales	Sí
Facebook	Post, páginas y grupos	Sí
Reddit	Post	Sí
Quora	Post	No
VKontakte	Post, páginas y grupos	Sí
Instagram	Imagen/vídeo	Sí
TikTok	Imagen/vídeo	No
Forocoches	Post	No
4chan	Post de imagen	No

Tabla 2: Redes sociales analizadas

Al final se han escogido para recoger información Twitter y YouTube. Respecto a los motivos de porque se han descartado el resto, por ejemplo, Facebook e Instagram han sido descartadas debido a la complejidad de uso de su API. En Reddit se ha descartado debido a que la comunidad hispana no es aún muy activa y el análisis de las temáticas será en castellano debido a que el medio tradicional escogido es en español. En Telegram hay que

buscar manualmente grupos específicos donde ya se sepa que se difunden noticias del tipo de temática que vamos a estudiar, etc.

A la hora de comprobar las redes sociales, se ha observado que en redes sociales como ForoCoches o Telegram es más fácil que se propaguen más noticias falsas que en Reddit por ser redes menos "formales" y menos moderadas. Cuanta menos moderación, más fácil que se difundan bulos, en sacrificio de "libertad" de expresión.

3.1.1. Twitter

A día de hoy se trata de la red social más utilizada, además de que con sus datos es fácil crear grafos que representen esta red social, motivos por los cuales se ha escogido como red social referente en este trabajo a la hora de realizar análisis. Se trata de una red social basada en el microblog (*microblogging*), cuyas características son: un usuario puede seguir a otro usuario, se pueden publicar tuits (*tweets*) con una etiqueta (*hashtag*) para agruparlos en una misma temática, visualizar listas y categorías de tuits y consultar el *trending topic* (lo más comentado en ese instante). Además, en un tuit puede haber menciones a un usuario, lo cual significa que se ha:

- Retuiteado el tuit: Consiste en copiar un tuit de otro usuario. Puedes mencionar a cualquier usuario más, pero al hacer un retuit automáticamente mencionas al autor del tuit.
- Respondido el tuit: Puedes mencionar a cualquier usuario en la respuesta, pero al responder a un tuit automáticamente mencionas al autor del tuit respondido.
- Citado el tuit: Un tuit con una referencia a otro tuit. Puedes mencionar a cualquier usuario, pero al hacer una cita automáticamente mencionas al autor de tuit citado.
- Mencionado a uno o varios usuarios: Mencionas a usuarios en un tuit. La diferencia con los tres anteriores es que no hace referencia a algún tuit.

3.1.2. YouTube

Se ha escogido YouTube como segunda red social debido a que se pueden identificar fácilmente las comunidades donde cada canal se correspondería con una comunidad, además de que vamos a extraer los comentarios de videos que se pueden equiparar a tuits de usuarios. Se trata de una plataforma donde se suben videos y la gente puede comentarlos, además de que cada usuario al crearse una cuenta tiene asociado un canal, lugar donde ese usuario sube sus videos y sobre el que se crea su comunidad.

3.2. Segunda categoría: varias plataformas o varias comunidades

Para comprobar si estamos en esta categoría, por una parte, hay que comprobar que en varias redes sociales están hablando de lo mismo (esto se comentará como hacerlo en la tercera categoría). Por otra parte, en esta sección se va a comentar como comprobar cuantas comunidades tiene un grafo. Para ello, primero comentaremos la tipología de los grafos sociales y luego como se pueden detectar sus comunidades.

3.2.1. Tipología de grafos sociales

Los grafos de las redes sociales reales (*real-world networks*) cumplen una o ambas tipografías de grafo que vamos a presentar a continuación.

Las redes de mundo pequeño (*small-world*) se basan en la característica de que cualquier nodo del grafo está conectado a otro mediante un camino corto, a pesar de que algunos nodos pueden incluso no tener muchos vecinos. Este tipo de grafo se puede encontrar en aeropuertos o el routing de la electricidad a través del tráfico web, además de estar asociado a la teoría de los seis grados de separación.

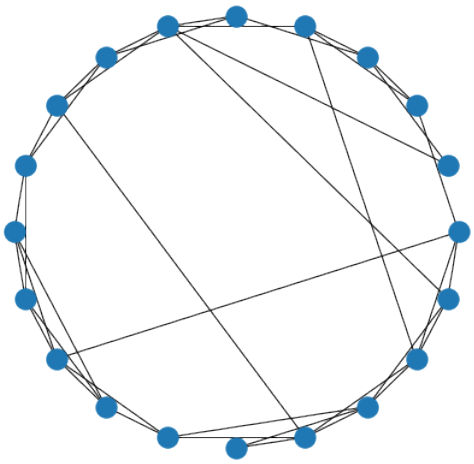


Figura 6: Red de mundo pequeño

Algunas de las características de este tipo de grafo son que tienen un coeficiente de agrupación (*clustering*) muy elevado. Para identificarlos se puede utilizar el componente sigma[8] del grafo.

Los grafos *scale-free* [LUN05] son grafos que cumplen la ley de la potencia, es decir, un grafo donde hay muchos nodos con una única conexión y pocos nodos que tienen muchas conexiones.

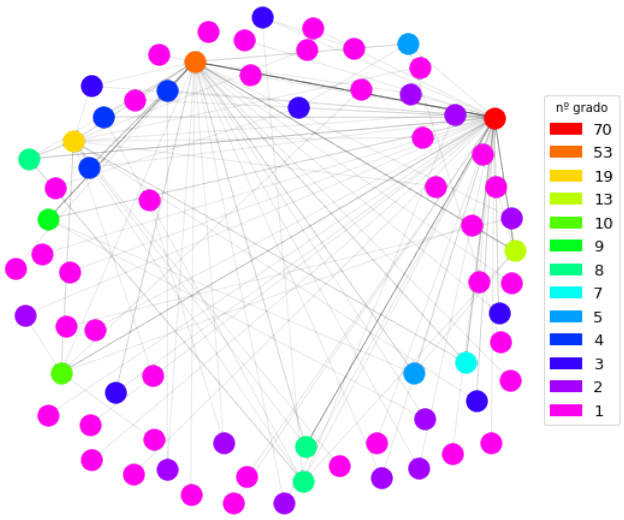


Figura 7: Grafo *scale-free*

[8] https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.smallworld.sigma.html

Para finalizar, otro tipo de grafo que se puede encontrar en las redes sociales son las redes egocéntricas (*ego-network*), grafos donde solo hay un nodo con un grado muy superior al resto de nodos del grafo.

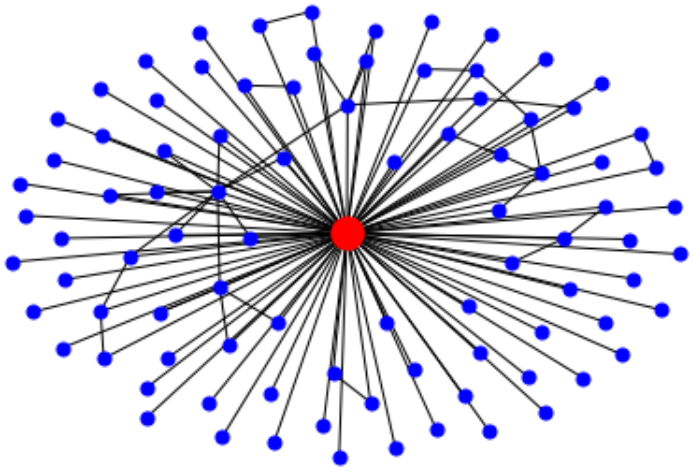


Figura 8: Red egocéntrica

Una de las cosas más características de los grafos de redes sociales es que están bajo una atmosfera homofílica, es decir, la tendencia de los nodos a estar asociados con otros de propiedades similares. Por ejemplo, si hacemos un grafo de las interacciones entre los usuarios de Twitter, seguramente un usuario que comente partidos de fútbol esté conectado con un camino más corto al usuario que represente la cuenta de Twitter oficial del Real Madrid que un usuario que represente a una tienda de alimentación para promocionarla.

Normalmente, los análisis de grafos de redes sociales se crean mediante las relaciones o interacciones entre usuarios (sociograma), enfoque que utilizaremos nosotros en Twitter. Pero si hay texto en las redes sociales, también se puede crear un grafo con las relaciones de palabras (por ejemplo, un nodo representa las palabras “es guapo”) o la relación de una palabra con la siguiente (la palabra “es” es un nodo y está relacionada con el nodo “guapo”, “feo”, “tonto”).

3.2.2. Comunidades en grafos

Las comunidades es una propiedad de los grafos que clasifica en grupos a aquellos nodos que estén más conectados entre sí. Cuanto más cerca esté el número de aristas con el número de nodos, será más fácil distinguir comunidades. Los grafos donde se intentan detectar comunidades a través de tópicos son conocidos como *topic network graph*.

Por ejemplo, en Twitter podríamos diferenciar tres tipos de comunidades: la comunidad fiel del usuario (seguidores y siguiendo), la comunidad formada a partir de la temática que se trata en la misma (*topic modeling*) y la comunidad formada a partir de la interacción entre sus usuarios (grafo).

En YouTube una comunidad se puede considerar los usuarios que interactúan en un canal, es decir, que cada canal se corresponde a una comunidad. En Reddit una comunidad sería un subreddit. En los casos de YouTube o Reddit se puede considerar un número n de comunidades al cual se haya extendido para considerar si la información ha cruzado otra comunidad debido a que hay muchos canales/subreddits que hablan de las mismas cosas.

En un grafo se pueden tener en cuenta muchos aspectos para realizar detección de comunidades. Por simplicidad solo vamos a analizar los métodos donde solo se consideran las aristas. Los métodos que utilizan las aristas para detectar comunidades se pueden dividir en métodos aglomerativos y divisorios.

En los métodos aglomerativos las aristas son añadidas una por una al grafo que únicamente tiene los nodos. Las aristas son añadidas desde la arista más fuerte a la arista más débil. Por otra parte, los métodos divisorios se basan en realizar lo contrario a los métodos aglomerativos. En ellos las aristas son eliminadas una a una del grafo.

Algunos de los algoritmos más conocidos en el estado del arte en la de-

tección de comunidades en grafos mediante aristas son Louvain, Leiden, Girvan-Newman, Kernighan-Lin, PageRank, etc [FOR10].

En los últimos años se están aplicando redes neuronales (*graph neural network*) [MEN22] para clasificación de nodos en un grafo, clasificación de texto y sobre todo en la predicción de enlaces en las redes sociales (por ejemplo, recomendación de nuevos amigos).

```
Data: G el grafo inicial
repeat
    poner cada nodo de G en su comunidad;
    while algun nodo es movido do
        foreach nodo n de G do
            coloca n en su comunidad vecina incluyendo la suya propia
             que maximiza la ganancia de modularidad;
        end
    end
    if la nueva modularidad es más grande que la inicial then
        G = la red entre las comunidades de G;
    else
        termina;
    end
until;
```

Algoritmo 1: Pseudocódigo del algoritmo de Louvain

Se ha escogido el algoritmo Leiden [TRA19] (una mejora del algoritmo de Louvain), ya que en ocasiones Louvain detecta comunidades que no están bien conectadas. Esto se consigue dividiendo de forma periódica al azar las comunidades ya detectadas en comunidades más pequeñas y bien conectadas.

Con el algoritmo de Leiden no le tenemos que especificar el número de comunidades que debe encontrar en el grafo, ya lo hace él solo. En otros algoritmos si hay que hacerlo. Esto plantea un problema, y es que si nos salen

por ejemplo seis comunidades, ¿puedo estar seguro de que se corresponde exactamente con seis comunidades de Twitter? Para responder a esta pregunta podemos tomar varios enfoques:

- Se puede considerar que efectivamente las comunidades que salen del grafo son las que hay y si el algoritmo saca más de una comunidad, suponer que hay varias comunidades hablando sobre la temática.
- Otro método que podríamos utilizar es que si nos salen cinco comunidades, cogemos a los diez nodos más influyentes del grafo (los métodos para detectar los nodos más influyentes se comentan en la quinta categoría) y si hay varios de ellos que caen en comunidades diferentes, suponer que hay varias comunidades.

Para realizar detección de comunidades se pueden utilizar otros atributos del grafo aparte de las aristas como el grado (número de conexiones que tiene un nodo), *broker* (nodo que conectan varios grupos), intermediación (link crítico para llegar a otros nodos) o cercanía (cuan fácil puede un nodo crear conexiones) como se puede ver en la figura 9. También se puede utilizar la posición del nodo en el grafo o los atributos de los nodos para detectar comunidades, por ejemplo, con k vecinos más cercanos.

Por último, comentar que se puede efectuar un análisis de sentimiento (*sentiment analysis*) en aquellas operaciones de influencia que traten de temas dualmente polarizados (negativo/positivo) para detectar comunidades. Se pueden utilizar métodos supervisados debido a que hay muchos conjuntos de datos *datasets* públicos etiquetados según su sentimiento. A fin de cuentas, un método supervisado bien entrenado y con datos de entrenamiento bien etiquetados suele dar mejores resultados que un método no supervisado. Para ello tenemos que conocer el contexto de lo que se va a hablar y si la temática que vamos a tratar está polarizada.

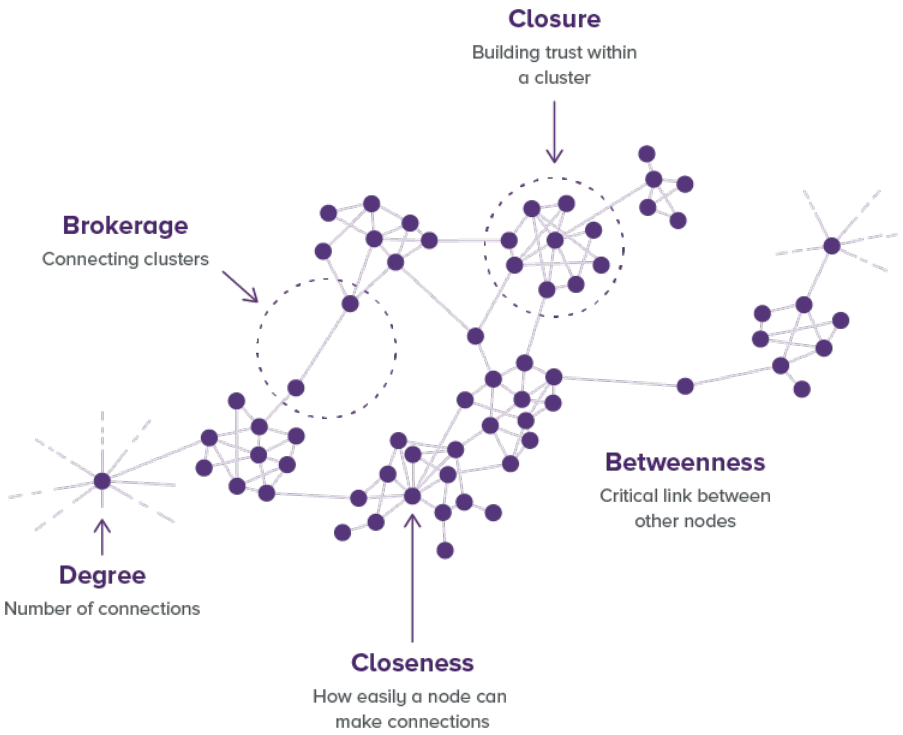


Figura 9: Atributos útiles de un grafo para detectar comunidades

3.3. Tercera categoría: varias plataformas y varias comunidades

Para comprobar si una temática cae en esta categoría vamos a tener que comparar la información de varias redes sociales para ver si están hablando de lo mismo. Para ello se realizará modelado de tópicos, aunque antes de ello habrá que procesar la información recolectada mediante procesado del lenguaje natural (*natural language processing*).

3.3.1. Procesado del lenguaje natural

Como se van a procesar tuits (Twitter) y comentarios de videos (YouTube) que suelen estar compuestas de una o pocas frases, se han decidido realizar los siguientes procesados:

- Poner todo el texto en letra mayúscula o minúscula.
- Tokenizacion: Elemento básico sobre el que vamos a trabajar. En nuestro caso se ha considerado una palabra como un token, aunque un to-

ken puede estar formado de varias palabras contiguas (n-grama).

- Eliminar elementos que no son útiles como *stopwords*, etiquetas HTML, enlaces web, los RT del principio de los retuits, varios espacios contiguos, menciones de usuarios, caracteres especiales en el token (por ejemplo, en la palabra "*state-of-the-art*" eliminar los '-').
- Lematización de las palabras.
- *Parts-of-speech* (POS): Si la palabra es un sustantivo, verbo, adverbio, adjetivo, etc.

Otros pasos que se pueden realizar son:

- Separar frases: Esto se suele hacer en caso de textos muy extensos y con muchas frases, ya que al cambiar de una frase a otra puede cambiar el contexto de lo que se está hablando.
- *Stemming*: Se suele utilizar en vez de la lematización. La librería que utilizaremos para el procesado de texto (spaCy) no acepta esta opción, al contrario que NLTK.
- *Expanding contractions*: Se suele utilizar solo en inglés, por ejemplo: *aren't* $\rightarrow$ *are not*.
- *Spell-correcting/spell-checking*: Corregir una palabra mal escrita, a veces por error del usuario y otras para evitar censura.
- Tratar/eliminar emojis: Puede ser interesante tratar los emojis en caso de realizar un análisis de sentimiento, ya que estos normalmente suelen reflejar sentimientos.

3.3.2. Aprendizaje automático

Se utiliza el aprendizaje automático para resolver el problema de clasificación de texto, donde en cada grupo se hablará de una temática distinta. Dentro del aprendizaje automático hay dos métodos a escoger, los métodos supervisados y los no supervisados. Los supervisados son métodos los cuales entrenamos con datos que ya tenemos clasificados para así generar

el modelo, mientras que los no supervisados generamos el modelo con datos que no están clasificados.

Como no siempre se va a conocer el contexto de la noticia ni dispondremos de datos clasificados de la noticia para entrenar, nos vamos a enfocar en métodos no supervisados, concretamente en modelado de tópicos (*topic modeling*), una técnica que además de ser no supervisada realiza una clasificación de textos en grupos donde cada grupo está compuesto de sus palabras más significativas que conformarán un tópico. Gracias a ello vamos a poder comprobar que se está hablando de la temática que queremos analizar (en caso de comprobar que no es así, podemos coger las palabras de la búsqueda más las palabras de los tópicos que nos interese y volvemos a realizar la búsqueda) y compararlo con el resto de redes sociales o medios tradicionales. Nos vamos a centrar en tres métodos.

Latent Dirichlet allocation (LDA) [BLE03] es el método más conocido del modelado de tópicos no supervisado. Se trata de un modelo generativo donde los documentos del corpus se representan como mezclas aleatorias sobre varios temas, donde cada tema es caracterizado mediante una distribución sobre todas las palabras.

Data: Dataset, K tópicos, hiperparámetros α y β
Output: tópicos, modelo
Elegir $\phi_i \sim Dirichlet(\alpha)$ donde $i \in$ M;
Elegir $\theta_k \sim Dirichlet(\beta)$ donde $k \in$ K;
foreach *palabra j de cada documento i* **do**
  $Z_{i,j} \sim Multinomial(\theta_i)$;
  $W_{i,j} \sim Multinomial(\phi Z_{i,j})$;
end

Algoritmo 2: Pseudocódigo de LDA

Non-Negative Matrix Factorization (NMF) [PAA94] se basa en reducir las dimensiones de un vector cuyo resultado será un vector reducido compuesto por valores no negativos. Dada una matriz A (los documentos por

palabras), es el resultado de la multiplicación de dos matrices H (los documentos por tópicos) y W (tópicos por palabras).

Los dos métodos mencionados anteriormente dan buenos resultados en textos extensos [AMR19]. Para tratar textos más cortos como tuits también hay métodos de modelado de tópicos como Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) [YIN14], basado en Movie Group Process. En caso de querer un texto más extenso, la mayoría de las redes sociales permiten conversaciones, por lo que se podrían unir varios comentarios de una conversación.

En todos los métodos mencionados anteriormente y en la mayoría de métodos de modelado de tópicos hay que especificar el número de tópicos que hay que buscar. En cada tipo de modelo hay que especificar otros argumentos, pero por temas de simplificación solo nos vamos a centrar en el de número de tópicos. En caso de no saberlo, se puede crear el modelo con 1, 2, 3, ... tópicos y sacar métricas que nos permitan evaluar la efectividad del modelo como perplejidad, *log likelihood* o el parámetro de coherencia del tópico. Este último, dependiendo de su valor, nos puede indicar si hemos generado un buen modelo, donde 0.3 es malo, 0.5-0.6 está bien y 0.8 empieza a ser malo.

Dependiendo de como sean los datos que trataremos, convendrá evaluar una métrica o varias de estas al mismo tiempo. Las métricas entre sí no tienen por qué ser proporcionales y, por ejemplo, una de ellas puede decirte que el modelo generado es bueno, mientras que otra métrica te diga lo contrario. En nuestro caso, el parámetro de coherencia del tópico seguramente sea el más idóneo, ya que en Twitter se suele retuitear mucho y esta métrica se basa en la repetición de palabras en cada documento. Esto en aprendizaje automático es conocido como ajustar los parámetros (*hyperparameter tuning*), es decir, encontrar los argumentos que optimizan el valor de una de las métricas.

La mayoría de métricas para analizar el modelo tenderán a mostrar ma-

yor efectividad conforme haya más tópicos (si cogemos 300 noticias relacionadas con una temática y decimos que tiene 300 tópicos, mostrará una efectividad del modelo muy alta) a pesar de que en estos modelos hay más argumentos a definir aparte del número de tópicos. Sin embargo, hay que evitar coger un número muy elevado de tópicos por dos razones.

La primera es que si escogemos muchos tópicos, llegará un punto en que muchos de los tópicos tendrán palabras comunes y será muy difícil distinguir de que están hablando en cada tópico.

La segunda es para evitar un concepto conocido a la hora de trabajar con modelos en aprendizaje automático llamado sobreajuste (*overfitting*) en las métricas. De igual forma, existe la contraparte del sobreajuste, conocido como subajuste (*underfitting*). Un ejemplo sería considerar que todas las noticias se engloban en un solo tópico o que la puntuación de la métrica de la coherencia del tópico tenga un valor bajo (subajuste) o un valor alto (sobreajuste).

Otra cosa aconsejable es la intervención humana para comprobar que los tópicos detectados tengan sentido y la persona pueda saber de qué se está hablando en ese tópico. Puede ocurrir que aparezca algún tópico el cual la persona no puede clasificar o que dos tópicos en realidad sean el mismo. Lo que no sería normal es que la persona no identificase de que se está hablando en ningún tópico encontrado por el modelo. En ese caso habría que mirar otro modelo para contrastar tópicos, por lo que la coherencia no nos basta para determinar si se ha hecho bien o no el modelado de tópicos.

Otros métodos del aprendizaje automático que se pueden utilizar son k vecinos (vectorizando el texto), Tfidf, Top2Vec, BERTopic, Hierarchical Dirichlet process, *neural topic modeling*, etc.

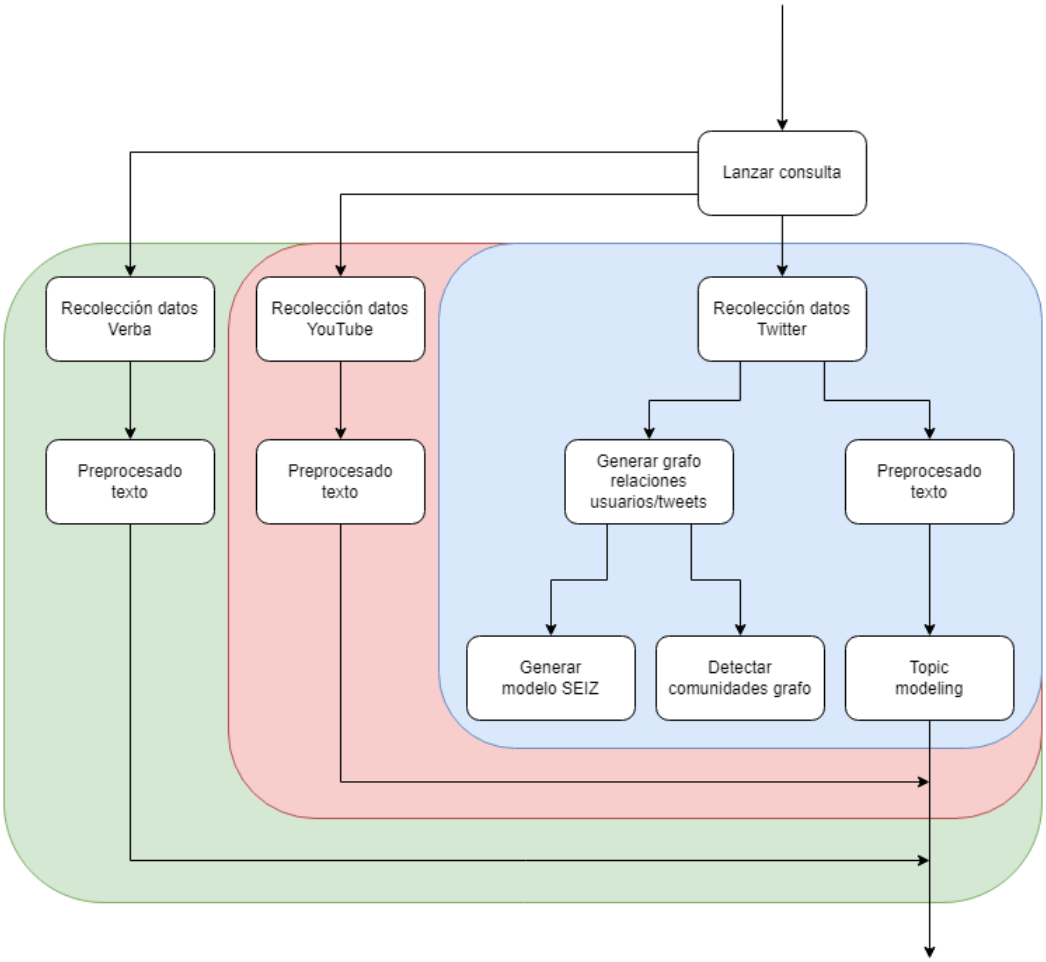


Figura 10: Diagrama de flujo de las tres primeras categorías

3.4. Cuarta categoría: medios tradicionales

A día de hoy los medios tradicionales no solo informan a través de la televisión, radio o papel, sino que utilizan Internet como medio para informar. Los medios tradicionales se han dado cuenta de que a día de hoy el no estar en Internet es quedarse atrás.

La ventaja que suele aportar Internet respecto a los medios tradicionales es que si ocurre una noticia, a los pocos minutos ya está colgada en la red y dependiendo de su viralidad se extenderá más o menos, mientras que los medios tradicionales había que esperarse horas (e incluso días en el caso del papel) para que se publicase la noticia. Los métodos para hacerlo por Internet son a través de su página web, perfiles en varias redes sociales, boletines enviados a un correo electrónico (*newsletters*), mediante suscrip-

ción a su portal web (suele ser de pago) o a través de un *feed* RSS. Solo las agencias de noticias internacionales más importantes (Reuters, Agencia EFE, Bloomberg, BBC) tienen API para consultar sus noticias.

El medio tradicional que se utilizara es la transcripción de los informativos de RTVE Verba/Civio[9]. Como alternativa para abarcar más medios tradicionales se podría utilizar Google News API.

Para considerar que la información ha pasado a los medios tradicionales y, por tanto, entrar en esta categoría, se puede realizar el procesado de lenguaje natural en la noticia (como se ha hecho en la tercera categoría) y si aparecen más de x palabras en común en un documento con uno de los tópicos encontrados en Twitter o en y porcentaje de documentos aparecen todas las palabras de un tópico, consideramos que la noticia también se está hablando en los medios tradicionales.

Otra aproximación sería realizar el modelado de tópicos sobre las noticias y si algún tópico tiene más de x palabras en común con uno de los tópicos encontrados en Twitter, también considerar que se está hablando de la noticia en los medios tradicionales.

3.5. Quinta categoría: individuos influyentes

Hay varios atributos que se pueden utilizar para saber si cierto usuario es influyente en todo Twitter como el número de seguidores, cuantas veces ha sido citado o si tiene el verificado de Twitter. En YouTube se puede utilizar el número de suscriptores o de 'me gusta' de un comentario para determinar la importancia de ese usuario.

Por otra parte, se pueden utilizar algoritmos de centralidad como Betweeness centrality, PageRank, VoteRank, entre otros [AMA21], para identificar a los actores influyentes en un grafo. En los grafos se conoce como centralidad a la importancia que adquiere un nodo en un grafo.

[9] https://verba.civio.es/

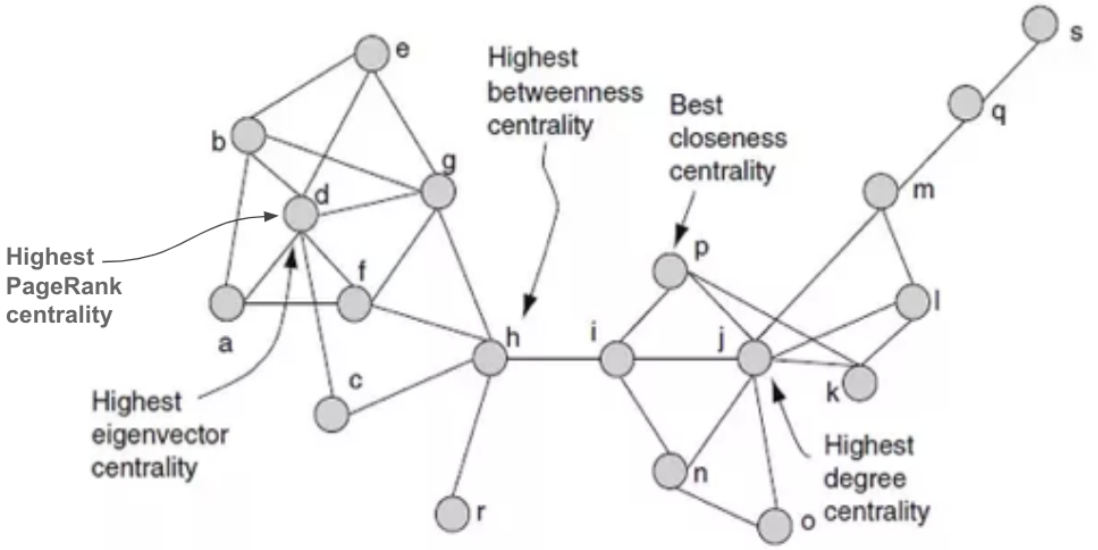


Figura 11: Medidas de centralidad en nodos

3.6. Sexta categoría: acción política o violencia

Como es muy difícil deducir si algún político ha tomado alguna acción al respecto, lo que se propone es una modificación de esta categoría, donde se puede hacer una o ambas de las siguientes cosas.

Se podría hacer un análisis de sentimiento de los tuits y si la puntuación de todos los tuits (incluyendo los repetidos) supera cierto umbral o entra dentro de una clasificación muy negativa, suponer que ha desencadenado una respuesta violenta. Para ello se pueden utilizar librerías como VADER, TextBlob.

Otra vía sería comprobar las ocurrencias de las palabras ley, orden o violencia y si consideramos que se repiten mucho suponer que se entra en esta categoría.

Como hemos comentado, es muy difícil comprobar si algún político ha tomado alguna acción a respecto, pero se puede tener en cuenta una lista de cuentas de políticos en Twitter y comprobar si alguno de ellos ha tuiteado sobre la temática, y si ese es el caso, considerar que está en esta categoría.

Por último, se podría realizar un modelado de tópicos a boletines oficiales de estados como el BOE o comunidades autónomas.

4. Modelo SEIZ

Hemos analizado todos los objetivos del trabajo, excepto el de predecir el comportamiento y difusión de la operación de influencia. Algunos de los métodos más conocidos para hacerlo se pueden encontrar en [ZHA14] como grafos dinámicos (analizar el grafo a través del tiempo) o métodos de cascada independiente.

En este trabajo se ha optado por un modelo epidemiológico, un modelo que permite analizar el comportamiento de agentes infecciosos en una población. Los más conocidos son el modelo susceptible-infectado (SI) donde la población al principio es susceptible y con el paso del tiempo individuos susceptibles se infectarán pasando al grupo de los infectados; susceptible-infectado-susceptible (SIS), como el anterior solo que los infectados al cabo de un tiempo volverán a ser susceptibles y se pueden infectar múltiples veces; susceptible-infectado-recuperado (SIR), igual que el SI solo que los infectados se recuperan con el paso del tiempo pasando al grupo de recuperados, sin posibilidad de volverse a infectar. Por último está el modelo susceptible-expuesto-infectado-recuperado, parecido al SIR solo que hay un nuevo grupo llamado expuestos entre los susceptibles y los infectados, donde está la gente que se ha expuesto a infectados y que algunos de ellos pasarán a ser infectados y otros puede que nunca dejen ese grupo.

En nuestro caso vamos a utilizar un modelo epidemiologico adaptado a la difusión de información llamado modelo SEIZ [JIN13]. Se trata de un modelo epidémico que clasifica la población en cuatro clases (susceptible, expuesto, infectado, escéptico) que utiliza las probabilidades de transición de un estado a otro para caracterizar la información errónea a partir de la información real. Esto se realiza comparando el ratio de infectados respecto al ratio de escépticos.

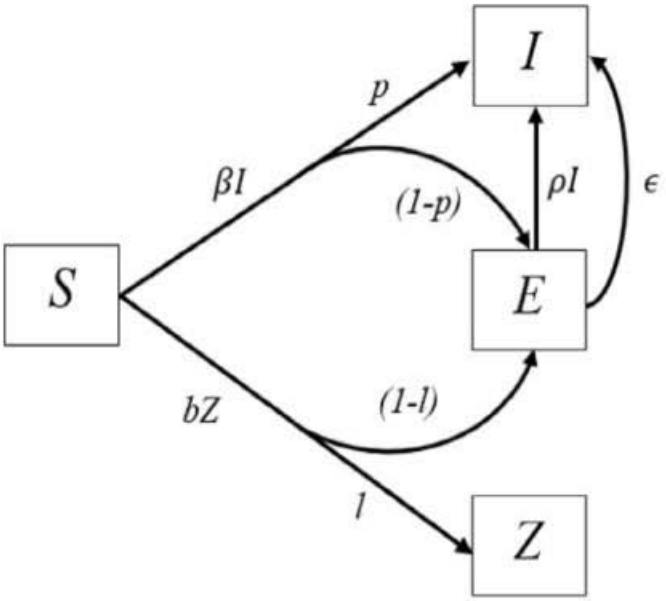


Figura 12: Diagrama de flujo SEIZ

El modelo está inicialmente compuesto de N individuos, donde $N = S$ y S son los susceptibles, es decir, individuos propensos a ser contagiados. Como hay mínimo un contagiado, conforme se avanza en el tiempo expondrá a otros individuos y estos pasarán al grupo de expuestos E. Los expuestos, en este modelo, pueden contagiarse propagando la información pasando al grupo de infectados I o pueden decidir no propagar la información, pasando al grupo de escépticos Z. También hay que comentar que un susceptible puede pasar directamente a estar infectado sin pasar por el grupo de expuesto (en nuestro caso será comentando la noticia por su cuenta sin exponerse) como se puede ver en la figura 12.

Como el componente más importante en una red social es la interacción entre usuarios, cuantos más escépticos hay, suele ser un indicador de que lo que se está difundiendo son noticias falsas. Que el crecimiento de infectados no sea acentuado, es decir, que pasen de pocos a muchos en poco tiempo, puede ser síntoma de que se trate de una noticia falsa, ya que tanto las noticias reales como las noticias falsas suelen tener un crecimiento de infectados acentuado, pero las noticias reales no suelen tener un crecimiento poco acentuado.

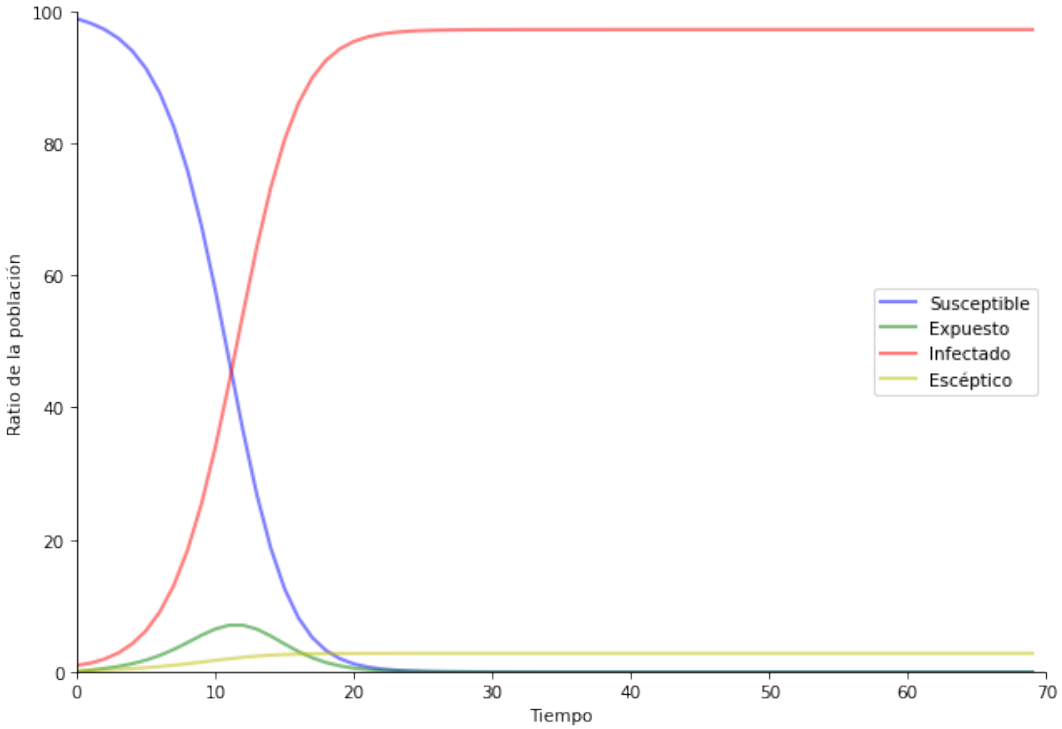


Figura 13: Modelo SEIZ común

La variación de individuos entre los diferentes grupos están definidos a través de este sistema de ecuaciones diferenciales:

$$\begin{aligned}
\frac{dS}{dt} &= -\beta S\frac{I}{N} - bS\frac{Z}{N} \\
\frac{dE}{dt} &= (1-p)\beta S\frac{I}{N} + (1-l)bS\frac{Z}{N} - \rho E\frac{I}{N} - \epsilon E \\
\frac{dI}{dt} &= p\beta S\frac{I}{N} + \rho E\frac{I}{N} + \epsilon E \\
\frac{dZ}{dt} &= lbS\frac{Z}{N}
\end{aligned} \tag{1}$$

Para comprobar cuando se está tratando con información verdadera o noticias falsas, se estudia el flujo y variación de los usuarios entre los distintos grupos en un tiempo determinado con las siguientes ecuaciones. Por ejem-

plo, la información verdadera suele mostrar un R_{SI} mayor que una noticia falsa.

$$\frac{bl}{\beta p} \tag{2}$$

$$\frac{b(1-l)}{\beta(1-p)} \tag{3}$$

$$\frac{\epsilon}{\rho} \tag{4}$$

$$R_{SI} = \frac{(1-p)\beta + (1-l)b}{\rho + \epsilon} \tag{5}$$

Algunos de los problemas que podemos encontrar en este modelo es que tenemos que realizar una estimación de los valores de N (población total), S_0 (susceptibles iniciales), E_0 (expuestos iniciales), I_0 (infectados iniciales), Z_0 (escépticos iniciales).

Parámetro	Definición
β	Ratio de contacto S-I
b	Ratio de contacto S-Z
ρ	Ratio de contacto E-I
ϵ	Ratio de incubación
1 / ϵ	Media de tiempo de incubación
bl	Ratio de efectividad de S $\rightarrow$ Z
$\beta\rho$	Ratio de efectividad de S $\rightarrow$ I
b(1 - l)	Ratio de efectividad de S $\rightarrow$ E vía contacto con Z
β(1 - p)	Ratio de efectividad de S $\rightarrow$ E vía contacto con I
l	Probabilidad S $\rightarrow$ Z dado el contacto con escépticos
1 - l	Probabilidad S $\rightarrow$ E dado el contacto con escépticos
p	Probabilidad S $\rightarrow$ I dado el contacto con infectados
1 - p	Probabilidad S $\rightarrow$ E dado el contacto con infectados

Tabla 3: Definición de parámetros en el modelo SEIZ

Incubación es el tiempo que está un individuo expuesto hasta que es infectado

El ratio de efectividad se mide en contactos efectivos por unidad de tiempo mientras que el ratio de contacto es el porcentaje de individuos de un grupo que vienen de otro grupo.

5. Caso de estudio

Se ha analizado el caso del sabotaje del oleoducto Nord Stream, por lo que ha hecho una búsqueda entre el 24 de septiembre del 2022 hasta el 9 de octubre del 2022. Se puede hacer un seguimiento de los acontecimientos del caso en el siguiente artículo[10].

Para recolectar los tuits se ha utilizado t-hoarder_kit[11] (que por debajo utiliza una librería de Python llamada tweepy). También se puede utilizar la librería tweepy directamente y en ese caso se recomienda recoger los atributos de los siguientes campos:

- tweet.fields: created_at, lang, public_metrics, referenced_tweets
- expansions: referenced_tweets.id.author_id
- user.fields: description, public_metrics, verified

Para poder utilizar la API de Twitter hay que crearse una cuenta y generar una aplicación en Twitter Developer Portal junto con los credenciales de la aplicación para poder hacer llamadas a la API de Twitter.

Si se quieren recoger los datos mediante una petición HTTP GET a falta de especificar la ventana de tiempo y especificar las credenciales de acceso a la API, el enlace sería: `https://api.twitter.com/2/tweets/search/recent?query=nord%20AND%20stream&tweet.fields=created_at,lang,public_metrics,referenced_tweets&expansions=referenced_tweets.id.author_id&user.fields=description,public_metrics,verified`

Para visualizar los grafos se ha utilizado Gephi. Para el resto se ha utilizado Google Colab mediante librerías en Python donde para el procesamiento del lenguaje natural se ha utilizado la librería spaCy, ya que es conocida por tener un *pipeline* (conjunto de operaciones para procesar texto) más rápido que la librería NLTK además de permitir lematización. Para el

[10] https://en.wikipedia.org/wiki/2022_Nord_Stream_pipeline_sabotage
[11] https://github.com/congosto/t-hoarder_kit

análisis de grafos se ha utilizado la librería NetworkX y por último, para el modelado de tópicos se ha utilizado la librería Gensim.

Para recoger datos de YouTube hay que crearse una cuenta de Google, crear una aplicación en Google Cloud Console para poder llamar a la API de YouTube y generar credenciales de la aplicación. Una vez realizado esto se ha utilizado las librerías de Python google-api-python-client y la de langdetect para cribar los videos con títulos en español y así no coger comentarios que no sean en español.

5.1. Primera categoría

Los operadores de búsqueda de cada sitio de donde vamos a extraer información (sea Twitter, YouTube o Civio/Verba) tiene sus propios operadores de búsqueda, aunque nosotros solo vamos a usar los básicos (AND, OR) y estos suelen ser comunes en todos los sitios.

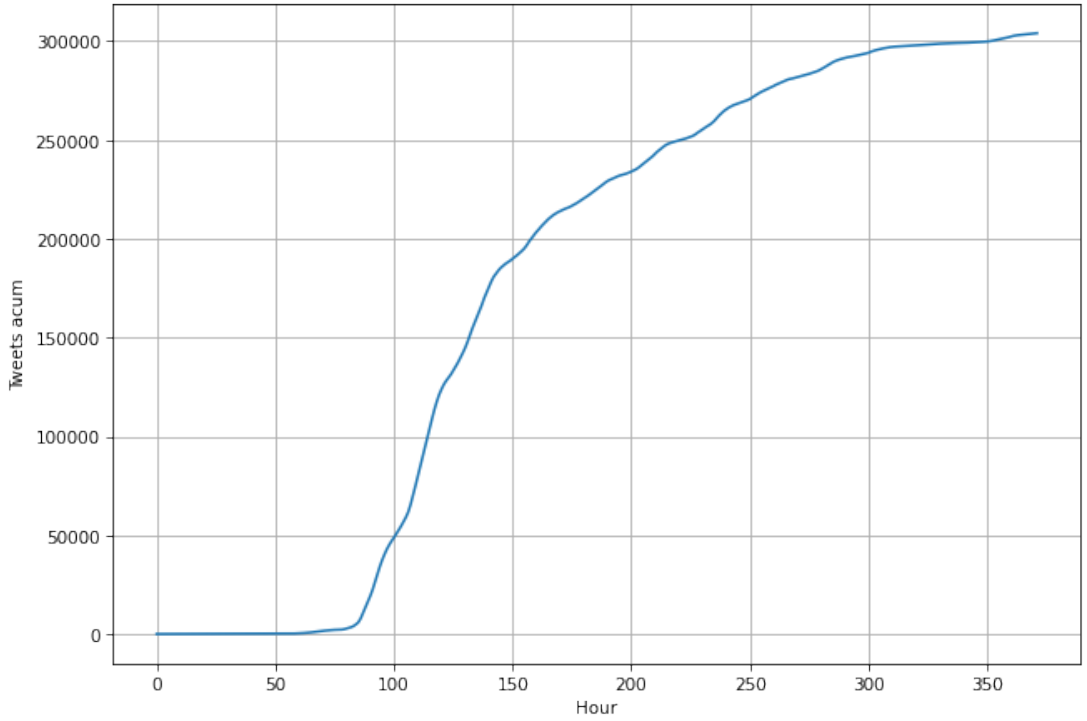


Figura 14: Número de tuits acumulados por hora Nord Stream

Se ha realizado la búsqueda en los tres lugares y primero nos hemos centrado en los tuits, que observando la figura 14 se ha decidido descartar los tuits de las 50 primeras horas, puesto que aún no había sucedido el suceso.

5.2. Segunda categoría

Debido a que la detección de comunidades a partir de 50000 nodos tarda alrededor de 1 hora (el procesado de texto suele tardar demasiado a partir de 10000 frases), se ha realizado un bucle haciendo k-core al grafo para eliminar los nodos menos conectados de forma iterativa hasta que en el grafo haya menos de 2000 nodos.

A continuación se le ha aplicado la detección de comunidades cuyo resultado se puede visualizar en la figura 15.

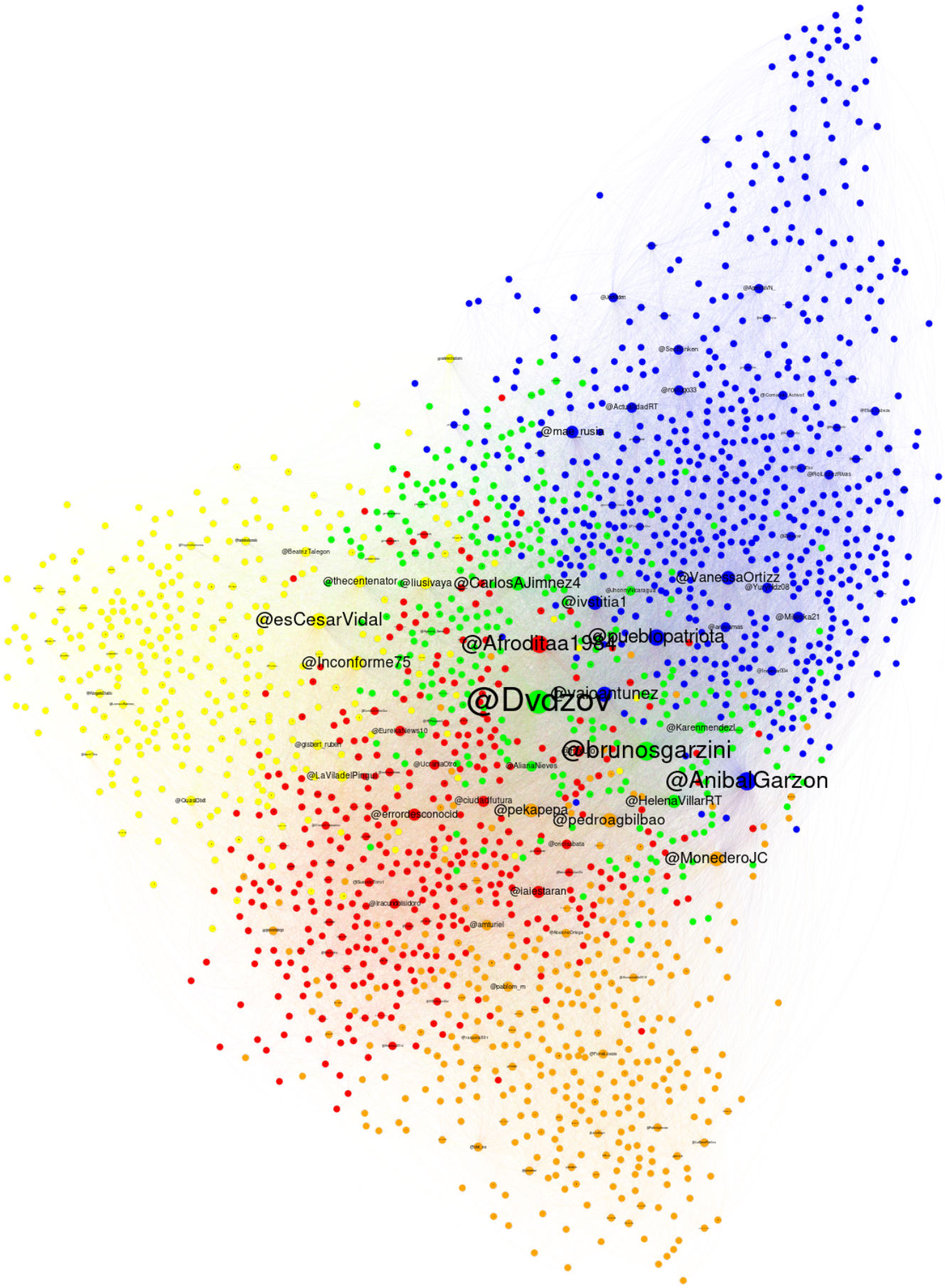


Figura 15: Detección de comunidades mediante Leiden

5.3. Tercera categoría

A partir de este punto se han aplicado las técnicas de procesado del lenguaje natural en los tuits comentadas anteriormente, cogiendo solo los sustantivos, verbos y adjetivos, cuyo resultado se puede visualizar en la tabla 4.

Tuits	Procesado del tuit
Nueva fuga en el Nord Stream. Jugada redonda: EEUU sabotea, Rusia es sancionado. Es una suposición. La guardia costera de Suecia detecta una cuarta fuga de gas en el Nord Stream Suecia, Dinamarca y la UE atribuyen los escapes a un sabotaje. https://t.co/nTjwZPdkHu	nuevo, fuga, nord, jugada, redondo, sancionado, suecia, detectar, cuarto, fuga, gas, nord, atribuir, escape, sabotaje
Es cómico cómo el sabotaje de Nord Stream será investigado por Suecia, un solicitante de la OTAN. La parte más divertida de esto es que Rusia, el país que diseñó y construyó las tuberías, no puede participar. #TerroristUSA https://t.co/e25cN20TAL	comico, sabotaje, nord, sero, investigar, suecia, solicitante, parte, divertido, rusiar, pais, diseno, construyo, tuberia, participar, terroristusa
Para hacer más análisis objetivos y menos apología sobre conflicto de Ucrania, se debe hablar más del Golpe Euromaidan, Guerra de Donbass, 14.000 muertos, Crimea, Batallón Azov, Cuarteto de Normandía, Nord Stream, hijo de Biden, Ucraniagate, OTAN,... y menos Putin, Putin, Putin	hacer, analisis, objetivo, apologia, conflicto, hablar, golpe, euromaidar, guerra, donbass, 14000, muerto, crimea, normandia, nord, stream, hijo, bidir, ucraniagate, putin

Tabla 4: Ejemplo tuits procesados

Una buena forma de comprobar si se está hablando de lo que queremos de una forma visual es haciendo una nube de palabras (*wordcloud*) del resultado del procesamiento del texto y ver si las palabras más resaltadas se corresponden con lo que queremos analizar, como se puede observar en la

figura 16.

Figura 16: Nube de palabras lematizadas

A continuación hemos empezado a aplicar LDA, NMF y GSDMM de 5 a 15 tópicos (más o menos tópicos llevaría a subajuste, sobreajuste o a no diferenciar ningún tópico) para quedarnos con el mejor de cada tipo, cuyos resultados se pueden visualizar en la figura 17.

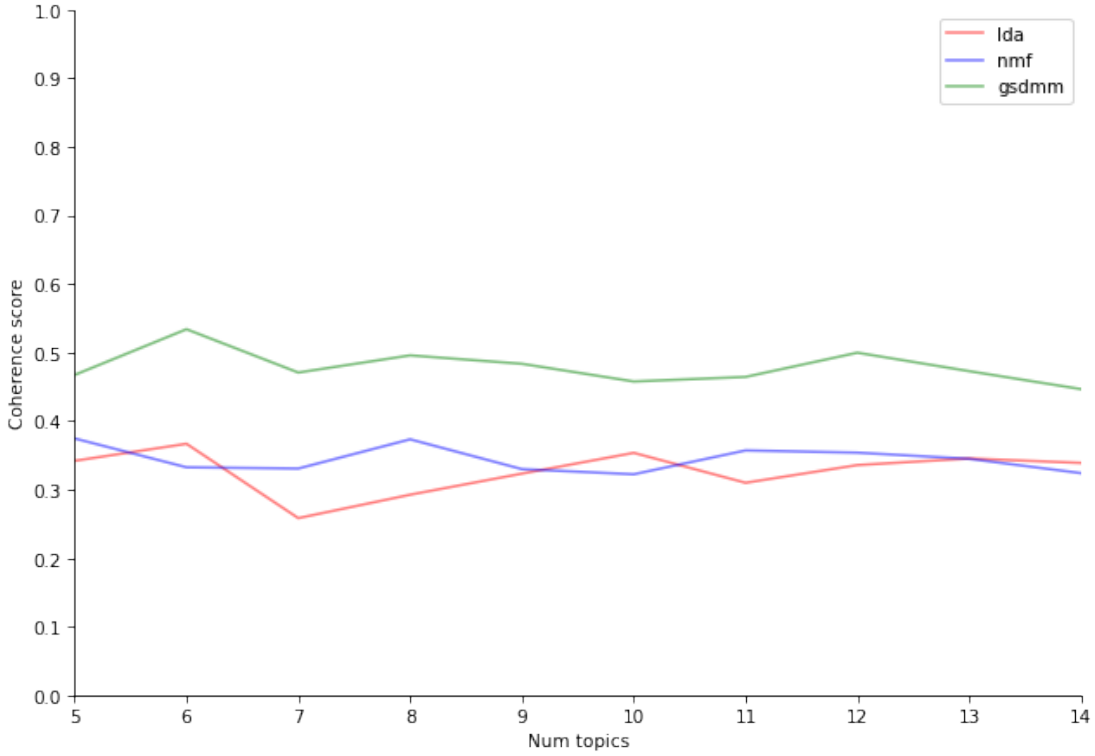


Figura 17: Comparación métodos de modelado de tópicos

En todas las pruebas que se han realizado en este trabajo nunca se ha superado el 0.6 de coherencia en cualquier modelo.

Uno de los problemas que nos hemos encontrado a la hora de aplicar modelado de tópicos es que la mayoría de tópicos tenían como palabra más repetida "nordstream" o "nord" y "stream", por lo que se han excluido del diccionario para montar el modelo las cinco palabras más repetidas de los documentos.

El número de palabras por tópico se ha escogido 7 palabras. Se recomienda escoger un número entre 5 y 10, ya que si son pocas o muchas palabras puede hacer imposible reconocer de que están hablando.

Respecto al análisis de los tópicos, en caso de que no se conozca el tema, se recomienda realizar una búsqueda avanzada en Google con las palabras de cada tópico más con las de la búsqueda principal realizada al principio para saber el contexto del mismo. Una vez descubiertos los tópicos puede ser de ayuda dar importancia a palabras de un tópico que no se encuentran en el resto, ya que suele ayudar a diferenciar el tópico respecto al resto.

Por ejemplo, para el tópico 2 la búsqueda en Google sería: internacional seguridad oleoducto terrorismo energetico kremlin objetivo nordstream OR nord stream. Esto buscaría:

- internacional seguridad oleoducto terrorismo energetico kremlin objetivo nordstream

y

- internacional seguridad oleoducto terrorismo energetico kremlin objetivo nord stream

Los tópicos de LDA se pueden visualizar en la figura 18, donde las barras representan el número de documentos que tiene ese tópico.

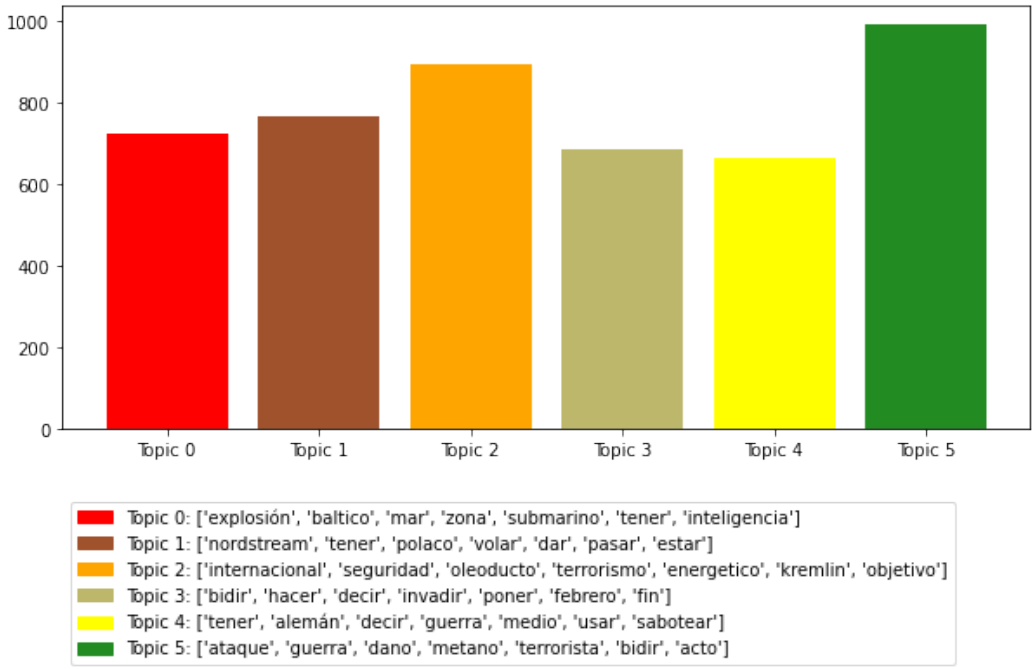


Figura 18: Tópicos LDA

Buscando en Google se ha deducido que los tópicos detectados en LDA son:

1. El primer tópico es de noticias de que Suecia detectó 2 explosiones en el mar Báltico antes de que se produjesen las fugas.

2. El segundo tópico trata de las declaraciones de Biden el 7 de febrero del 2022.

3. El tercer tópico es "Rusia pedirá una reunión del Consejo de Seguridad de la ONU por el sabotaje del Nord Stream".

4. El cuarto tópico son de las declaraciones de Biden en febrero diciendo que acabaría con el Nord Stream si Rusia invadía Ucrania.

5. El quinto tópico son las declaraciones de Alemania tras el sabotaje.

6. El sexto tópico es el mismo que el cuarto.

Tanto el cuarto como el sexto tópico son muy susceptibles a fomentar las noticias falsas. Los tópicos de NMF se pueden visualizar en la figura 19.

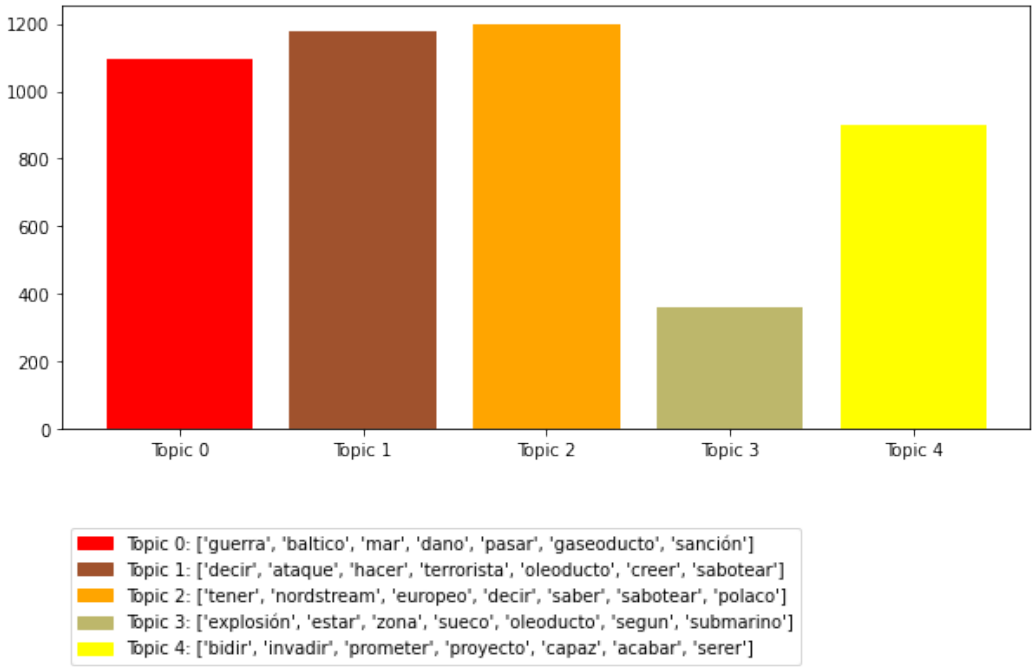


Figura 19: Tópicos NMF

Respecto al análisis de los tópicos del NMF nos encontramos que:

1. El primer tópico es noticia del suceso.
2. El segundo tópico pueden ser noticias de declaraciones cruzadas del suceso.
3. El tercer tópico es la noticia de las declaraciones del ministro polaco.
4. El cuarto tópico son las declaraciones suecas del suceso.
5. El quinto tópico son las declaraciones de febrero de 2022 de Biden.

Por último tenemos los tópicos del GSDMM, que se pueden visualizar en la figura 20.

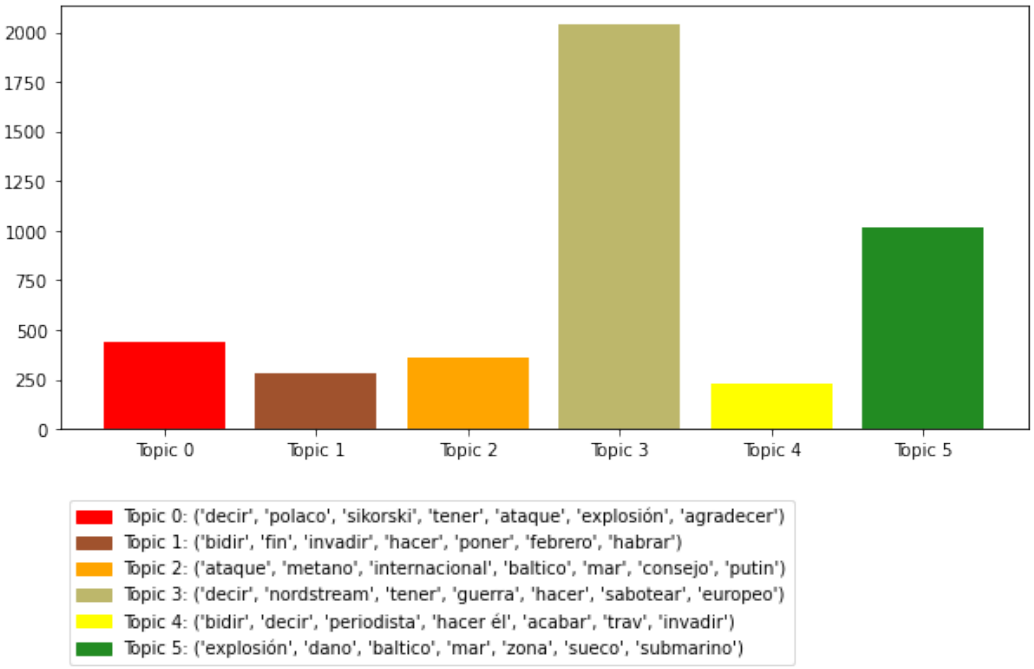


Figura 20: Tópicos GSDMM

Observando el modelo GSDMM, se observa que los tópicos son muy parecidos a los que refleja LDA solo que añade las declaraciones del ministro polaco, a pesar de tener una distribución de documentos más descompensada. Los tópicos con más documentos son noticias del suceso y declaraciones (tópicos 3 y 5).

A continuación veremos los grafos con los nodos (usuarios) coloreados con el color del tópico que se corresponda de las tres figuras anteriores. Como alguien puede haber publicado varios tuits que hayan caído en varios tópicos, lo hemos coloreado con el tópico donde han caído más tuits de ese usuario.

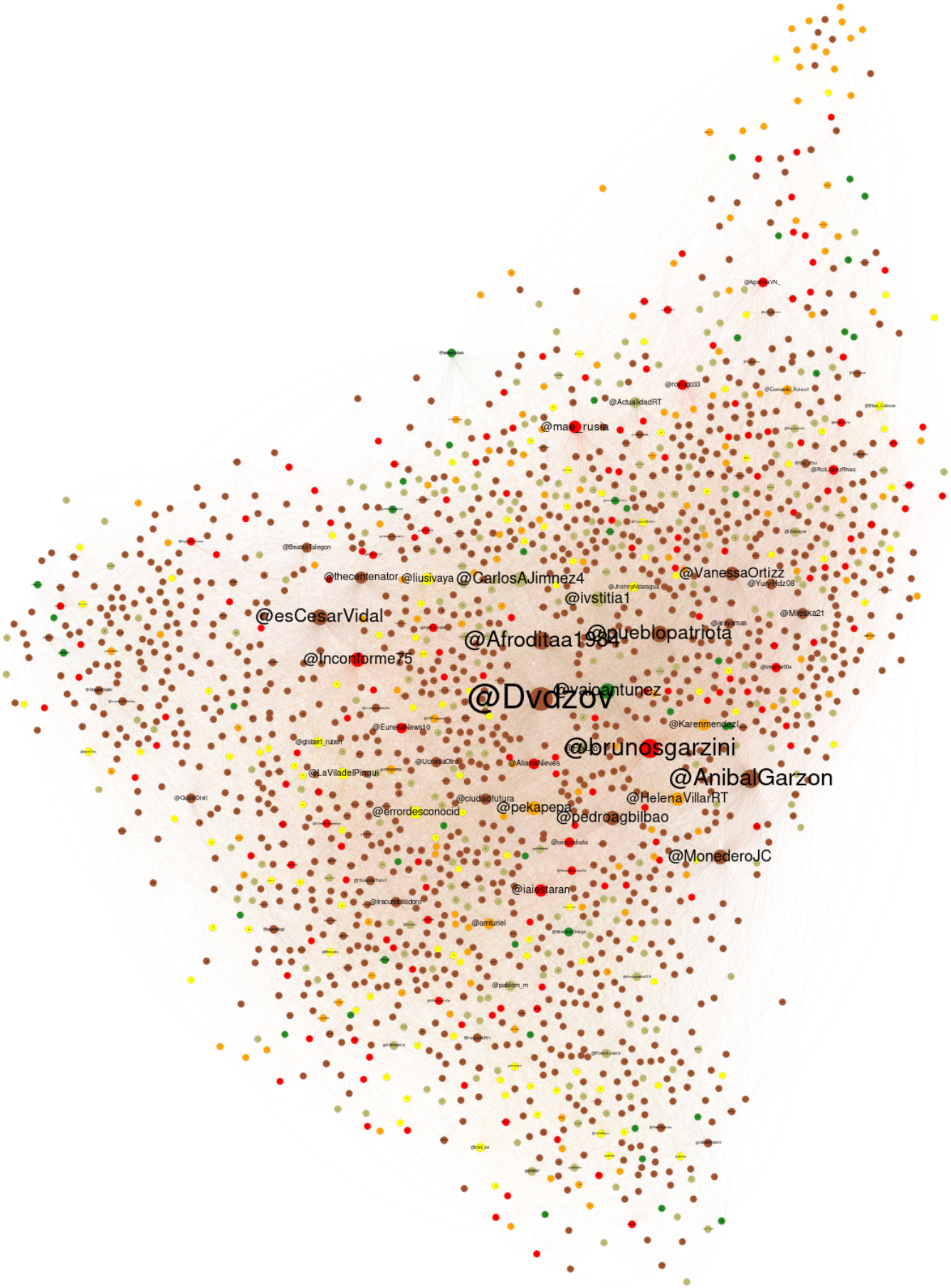


Figura 21: Grafo LDA

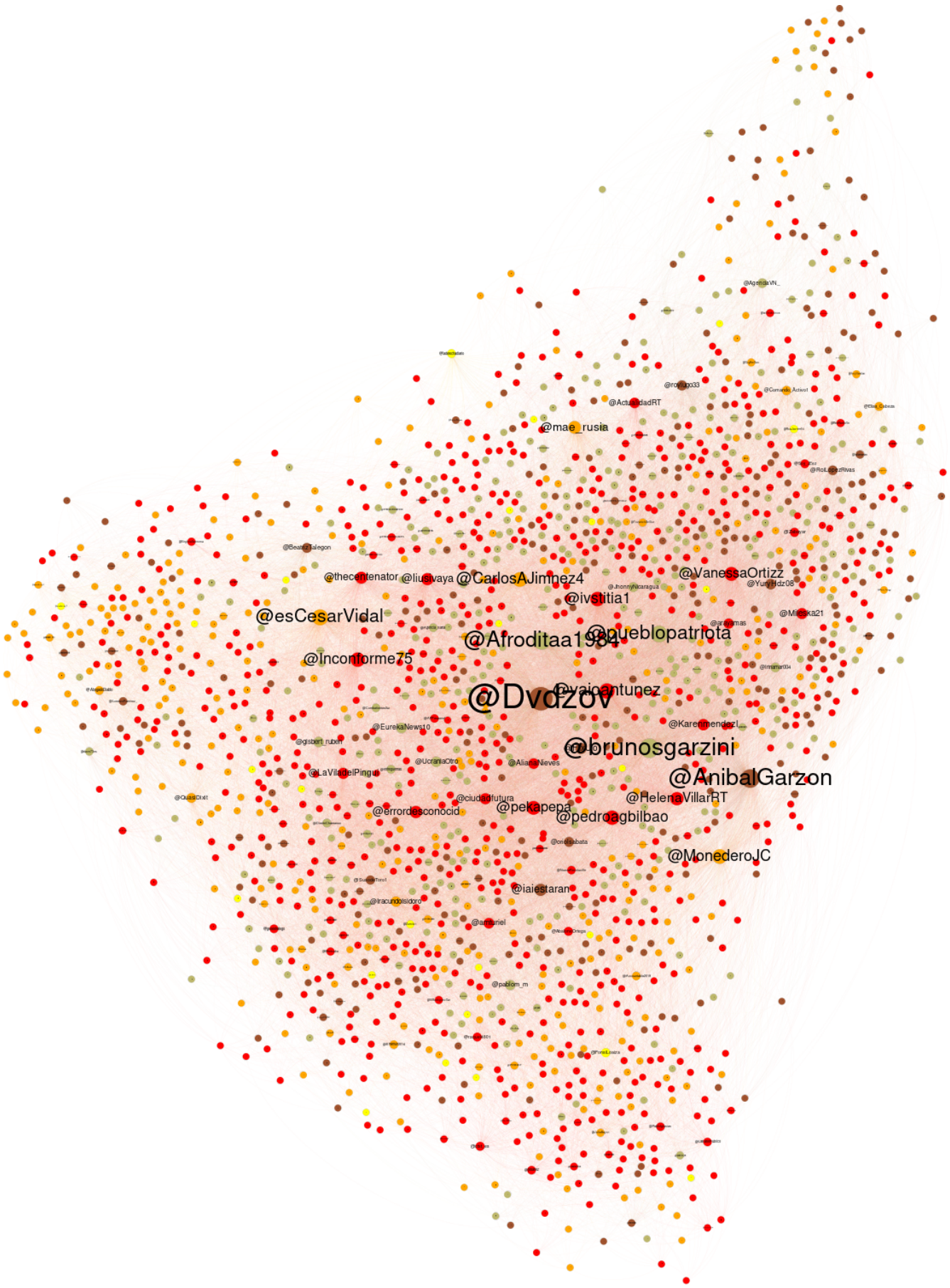


Figura 22: Grafo NMF

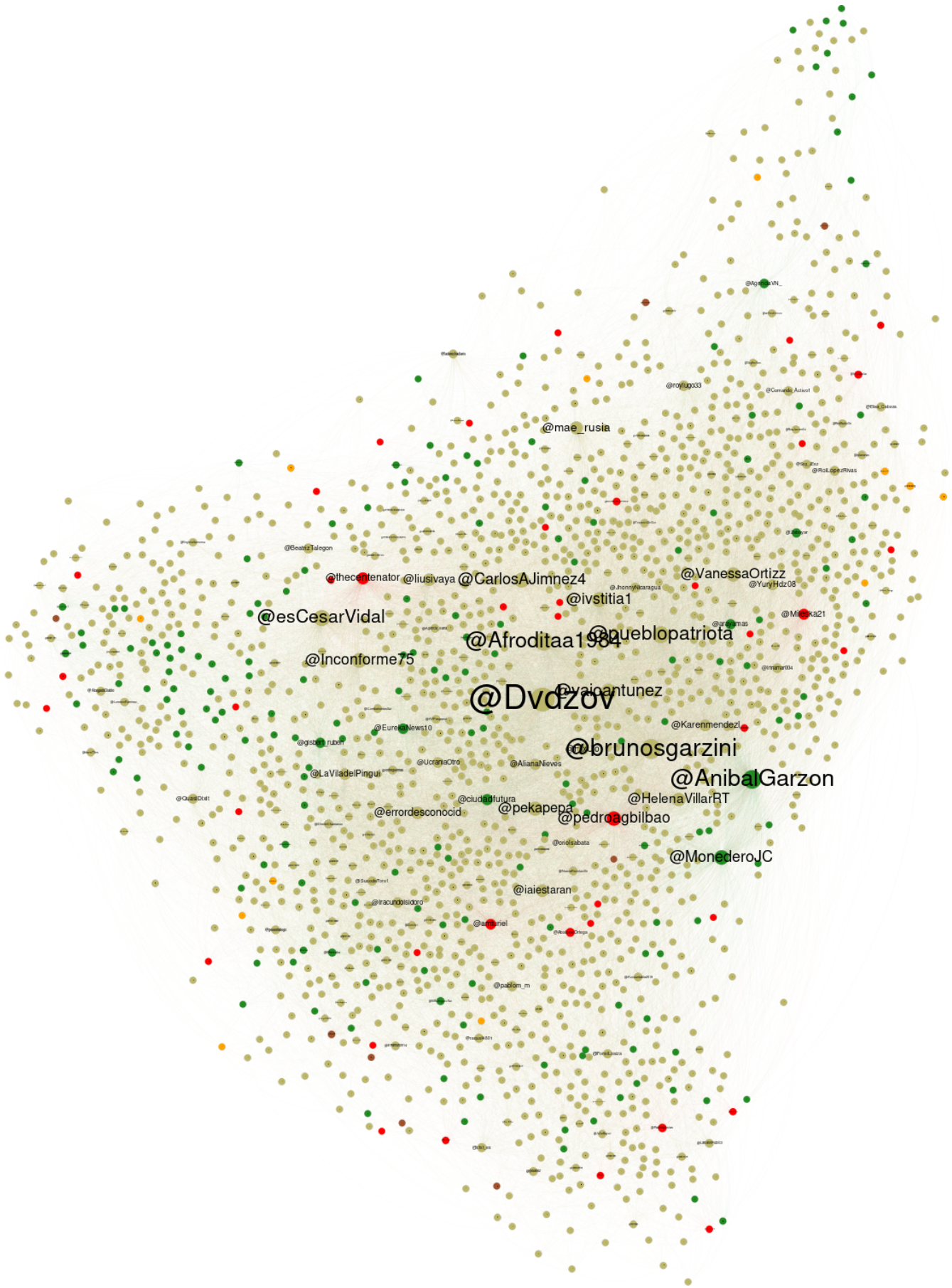


Figura 23: Grafo GSDMM

Como se puede observar, modelado de tópicos no tiene por qué coincidir con la detección de comunidades. La mayoría de usuarios tenían varios tuits publicados a lo largo de esas 300 horas, por lo que la mayoría de usuarios solía tener un número considerable de tuits en cada tópico detectado.

5.4. Cuarta categoría

Para recoger los datos de las transcripciones de los telediarios se puede llamar a la API de Civio/Verba sin falta de ninguna credencial con el siguiente enlace: `https://verba.civio.es/api/search.csv?q=nord%20stream&size=10000&date_from=2022-09-27&date_to=2022-10-04`. Para que la petición nos devuelva un JSON en el enlace anterior hay que sustituir "search.csv" por "search".

index	video_id	title	type	hash	text
0	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	comment	0	Saben de sobra quien es el responsable del sabotaje, pero no van a dar noticias negativas de sus colegas. Así opera la libertad de prensa y de expresión.
1	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	comment	1	Al momento que ese "accidente" paso el gas gringo subio un 20% mas su costo. Quien es el mayor beneficio con esto?
2	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	reply	1	Del mismo modo subió el precio del gas ruso... La diferencia entre ambos es que estados unidos no depende de las ventas de gas como si depende Rusia.
3	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	reply	1	Sería para quién no tuviera contrato de compra de gas.
4	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	reply	1	Paola has comprago gas gringo para saberlo? A mi me ha cortado menos la botella que antes..... Qué cosas....
5	eGFGG5eTKY4	El Nord Stream 1 podría quedar inutilizable para siempre	reply	1	EEUU

Figura 24: Comentarios de videos en YouTube

Esto nos devolverá transcripciones compuestas por una frase y el intervalo de tiempo en la que se dijo esa frase en el telediario (ideal ya que se pueden equiparar a tuits o a comentarios de videos).

5.5. Modelo SEIZ

Para generar el modelo SEIZ se ha ido haciendo un barrido de usuarios expuestos e infectados cada hora. Al principio, todos son susceptibles y conforme pasa el tiempo van pasando al grupo de expuestos o al de infectados.

Si alguien es infectado y está conectado en el grafo a alguien que es sus-

ceptible, este pasa a estar expuesto. Esto ya puede dar inexactitud en el modelo, ya que a lo mejor alguien en Twitter menciona a un usuario y el mencionado no tiene por qué leer el tuit al momento, sino al cabo de 3 horas, por lo que ese usuario no pasaría al grupo de expuestos hasta luego de 3 horas. Incluso puede que no lo lea nunca (por lo tanto no pasaría a ser expuesto ni infectado y sería siempre susceptible) o que haya otros usuarios que se expongan a esos tuits porque Twitter les ha recomendado ese tuit. Como los casos comentados anteriormente no se pueden controlar, vamos a suponer que todos los que han sido mencionados leen el tuit justo cuando se publica y que no hay usuarios expuestos mediante recomendaciones de Twitter, por lo que toda la población al final va a estar infectada o escéptica. De lo único que vamos completamente seguros es de cuantos infectados habrá en un determinado tiempo.

Al final en el grupo de susceptibles no quedará nadie, la mayoría estarán en el grupo de infectados y unos pocos estarán en el grupo de expuestos. En el grupo de escépticos no pasaremos a nadie, ya que no sabemos quién es escéptico. Pero justo se hace este barrido para saber quien es escéptico, porque todos los que se queden en el grupo de expuestos son los escépticos y consideraremos que son usuarios que han sido expuestos a la información, pero que por cualquier razón al final no han tuiteado.

Una vez sabemos qué usuarios son escépticos, hacemos otro barrido desde cero, con la diferencia de que ahora, en caso de que un escéptico es expuesto, pasa directamente al grupo de escépticos. En los modelos SEIZ al final en el grupo de expuestos no debería quedar nadie. Los resultados se pueden visualizar en la figura 25a.

A partir de aquí podemos analizar el flujo de usuarios a los distintos grupos en un determinado tiempo para ver el comportamiento. Por ejemplo, se pueden sacar los usuarios que han tuiteado después de que les hayan mencionado explícitamente en un tuit, que en nuestro caso han sido 5406 de 8308 menciones, es decir, un ratio muy elevado, señal de que la información parece ser real. Además, 8662 usuarios han tuiteado directamente

sin estar expuestos (han pasado de susceptible a infectado directamente), con lo cual podemos deducir que han habido muchos retuits (exactamente el 91 % de los 300000 tuits que hay en total), respuestas o menciones.

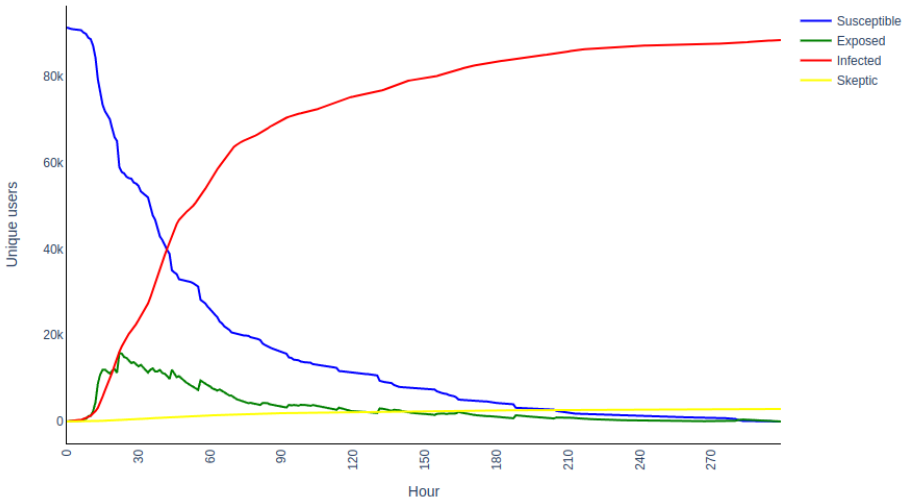
Observando el ratio de infectados en comparación con el ratio de escépticos, parece que la gran parte de información que ha circulado sobre la temática del Nord Stream parece ser información verídica.

Por último, para estimar los parámetros del modelo SEIZ se ha realizado un ajuste de mínimos cuadrados con el número de expuestos e infectados (es el que mejor ajuste nos ha dado, aunque como en este modelo solo sabemos con certeza el número de infectados sería más conveniente realizar el ajuste solo con los infectados), donde los parámetros a estimar del modelo oscilan entre un valor de mínimo 0.01 y máximo de 5. Los valores estimados se pueden ver en la tabla 5.

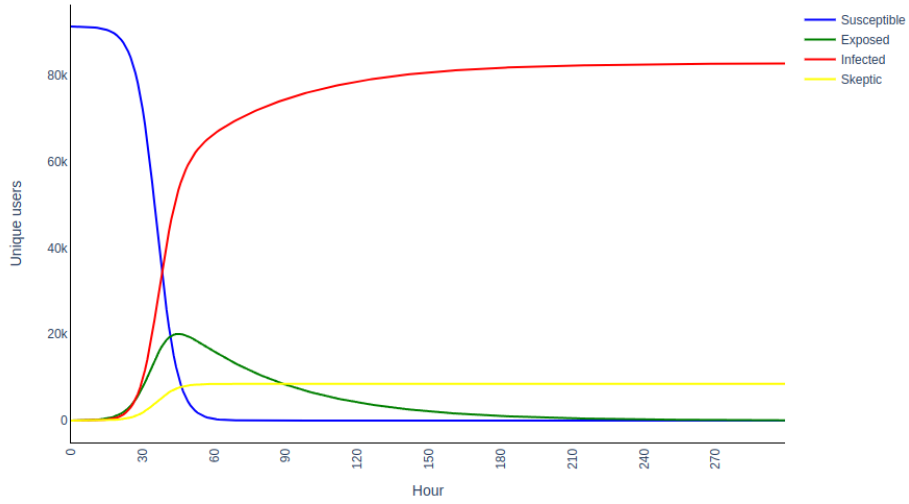
Variable	Valor
β	0.04949989
b	2.08352032
ρ	0.01
p	5.0
l	0.10641760
ϵ	0.01418473

Tabla 5: Valores de los parámetros estimados del modelo SEIZ

Una vez obtenidos los valores, hemos realizado una simulación del modelo SEIZ con esos parámetros, dando como resultado la figura 25b.



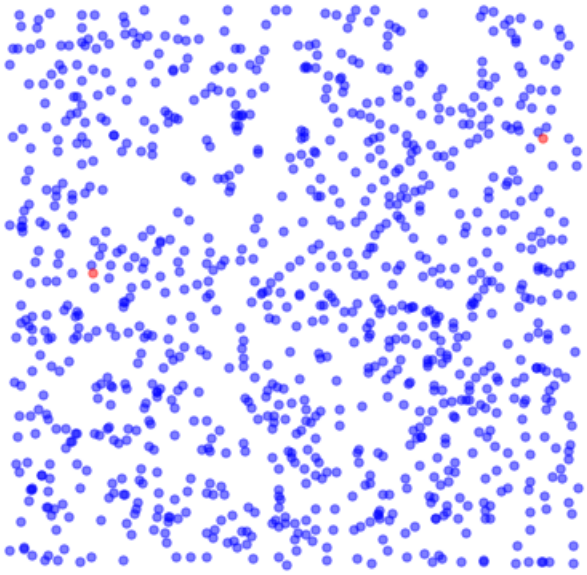
(a) Modelo SEIZ de la búsqueda Nord Stream



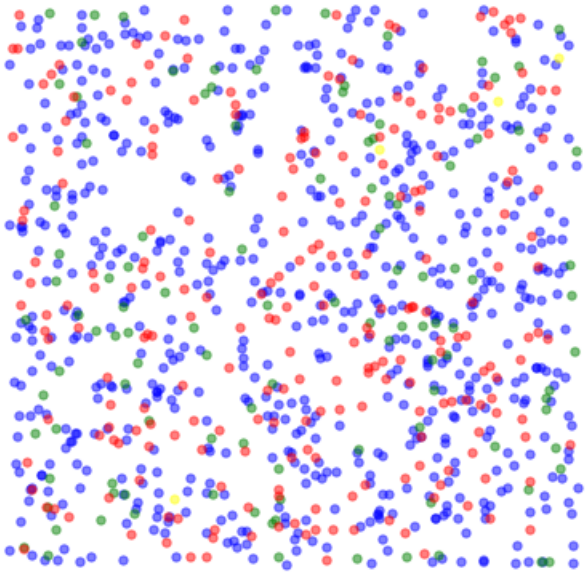
(b) Simulación con parámetros estimados

Figura 25: Modelo SEIZ de datos reales y simulado

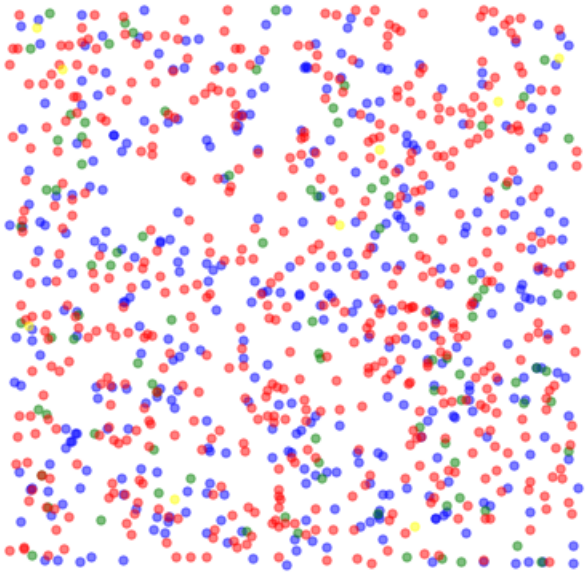
Para finalizar el caso de estudio se ha visualizado una línea del tiempo cada 25 horas de mil usuarios elegidos aleatoriamente mostrando en que grupo se encuentran. Los usuarios azules son los susceptibles, los verdes son expuestos, los rojos son los infectados y los amarillos son los escépticos. Sobre la hora 101 prácticamente todos los usuarios ya están infectados. Si se observa la figura 14 entre las horas 80-150 es cuando se da el crecimiento más acentuado, luego hasta la hora 300 el crecimiento es menos acentuado y así cada vez se va atenuando más.



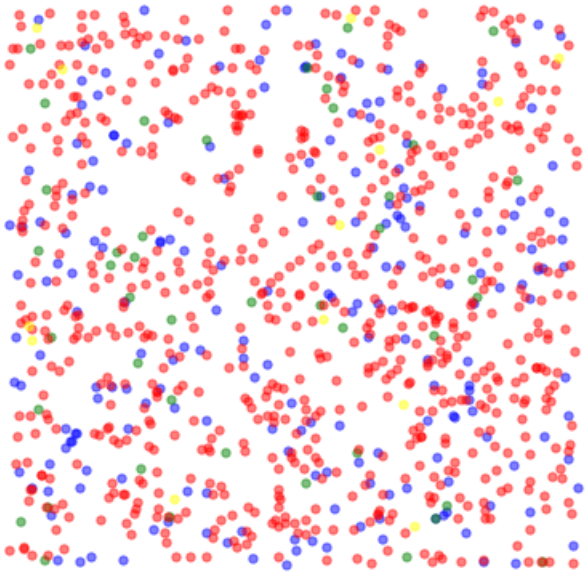
(a) Hora 1



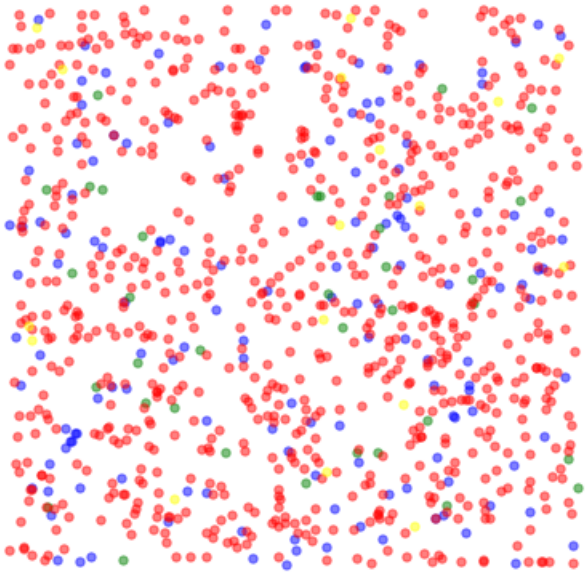
(b) Hora 26



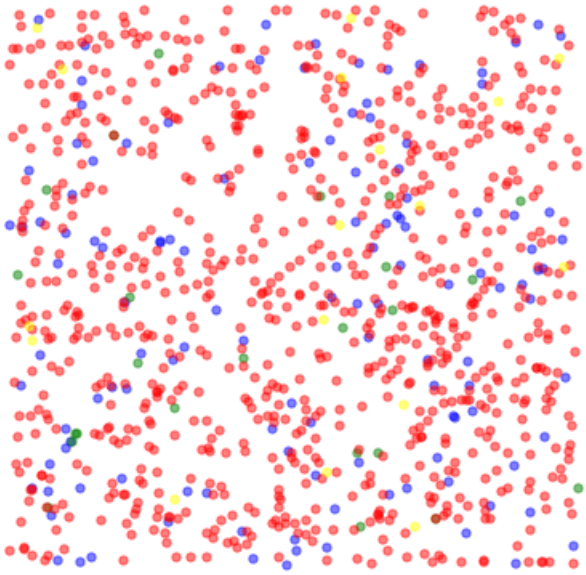
(c) Hora 51



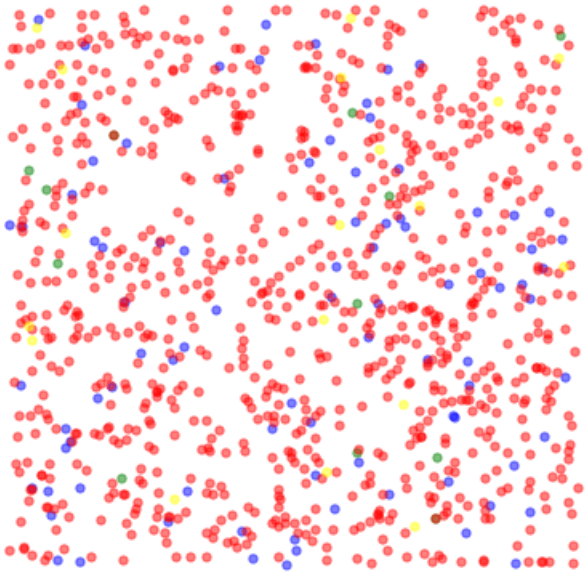
(d) Hora 76



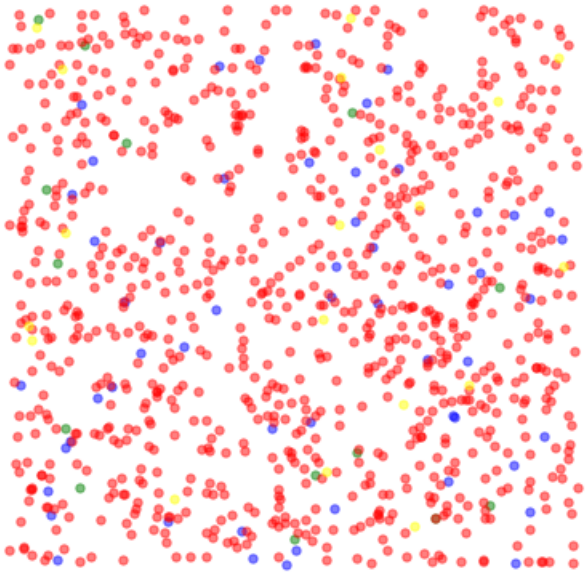
(e) Hora 101



(f) Hora 126



(g) Hora 151



(h) Hora 176

Figura 26: Timeline modelo SEIZ de 1000 individuos aleatorios

6. Conclusiones

Para concluir el trabajo, vamos a hacer repaso de sí se han cumplido los subobjetivos marcados al inicio de la memoria:

1. El primer subobjetivo de construir un repositorio con los mensajes y textos sobre un tema se ha realizado, ya que al realizar una búsqueda genérica, encuentra información en varios lugares (en nuestro caso Twitter, YouTube y Civio/Verba).

2. El segundo subobjetivo de clasificar una operación de influencia siguiendo las seis categorías de la escala de ruptura no se ha realizado a pesar de que sí que se ha explicado como se puede realizar con cualquier temática.

3. El tercer subobjetivo de emplear técnicas de análisis de redes sociales para construir la red de la operación de influencia y analizar su estructura también se ha realizado, puesto que hemos profundizado en la tipología de los grafos de redes sociales, hemos visto algoritmos que detectan comunidades en el grafo.

4. El cuarto subobjetivo de identificar los actores de la red claves en la transmisión de la operación de influencia también se ha realizado al proponer métodos para identificarlos en la quinta categoría.

5. El quinto subobjetivo de implementar un prototipo para validar la metodología propuesta con casos reales no se ha realizado, ya que el prototipo realizado no implementa la quinta y sexta categoría.

6. El sexto subobjetivo de predecir como va a comportarse una operación de influencia a través del tiempo se ha realizado al estimar los parámetros de un modelo SEIZ a partir de los datos reales recogidos en Twitter.

A continuación vamos a comentar las conclusiones extraídas de este trabajo. Analizando los tipos de tuits a la hora de recolectar los tuits, se ha observado que la mayoría de ellos son retuits y la mayoría de los mencionados suelen responder al tuit.

Respecto al modelado de tópicos, se ha observado que GSDMM muestra resultados mejores que LDA y NMF, pero suele tener una distribución descompensada de los documentos en los tópicos. Además, el modelado de tópicos no es bueno para hacer agrupación en Twitter debido a que los usuarios suelen tuitear varios tuits sobre la temática pero de distinto tópico según el modelado de tópicos.

Por último, reafirmar que a día de hoy aún falta intervención humana en cualquier herramienta de reconocimiento de noticias falsas.

Como posibles vías de futuro estudio se podría hacer k vecinos más cercanos para detectar comunidades con los atributos de los usuarios de Twitter como seguidores, retuits, ... para detectar comunidades como las agencias de noticias (publican muchos tuits, no suelen retuitear) entre otras. También sería interesante analizar los *graph neural network* para detección de comunidades.

Otra vía de interés consistiría en realizar un caso de estudio de una temática que sepamos de antemano que difunde noticias falsas y compararlo con el caso de estudio de este trabajo, puesto que el caso de estudio de este trabajo hemos concluido que no se han difundido muchas noticias falsas. Por otra parte, sería interesante profundizar sobre la relación entre infectados y escépticos en el modelo SEIZ y su relación con las noticias falsas.

Además, como objetivo no finalizado, convendría terminar el caso de estudio para la categoría 5 y 6 para determinar los umbrales que hay que definir en cada una de esas categorías para determinar si entran o no en esa categoría.

Apéndice

A. Bosquejo aplicación

A continuación mostraremos un bosquejo de lo que podría ser una aplicación web funcional. Se recomienda utilizar algún *framework* de *backend* de Python como Django o Flask con Sigma.js para que los resultados con grafos puedan ser interactivos.

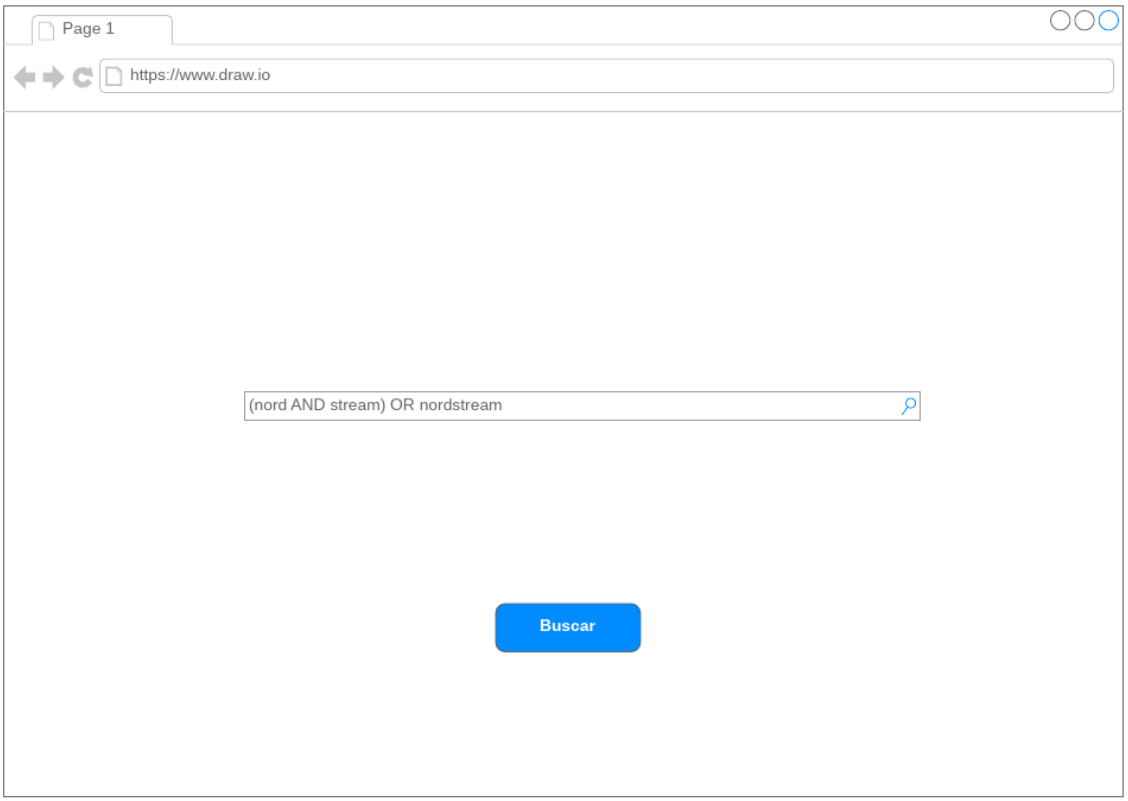


Figura 27: 1ª pantalla

Estaría constituida por la búsqueda que sería la misma para Twitter, YouTube y Civio/Verba.

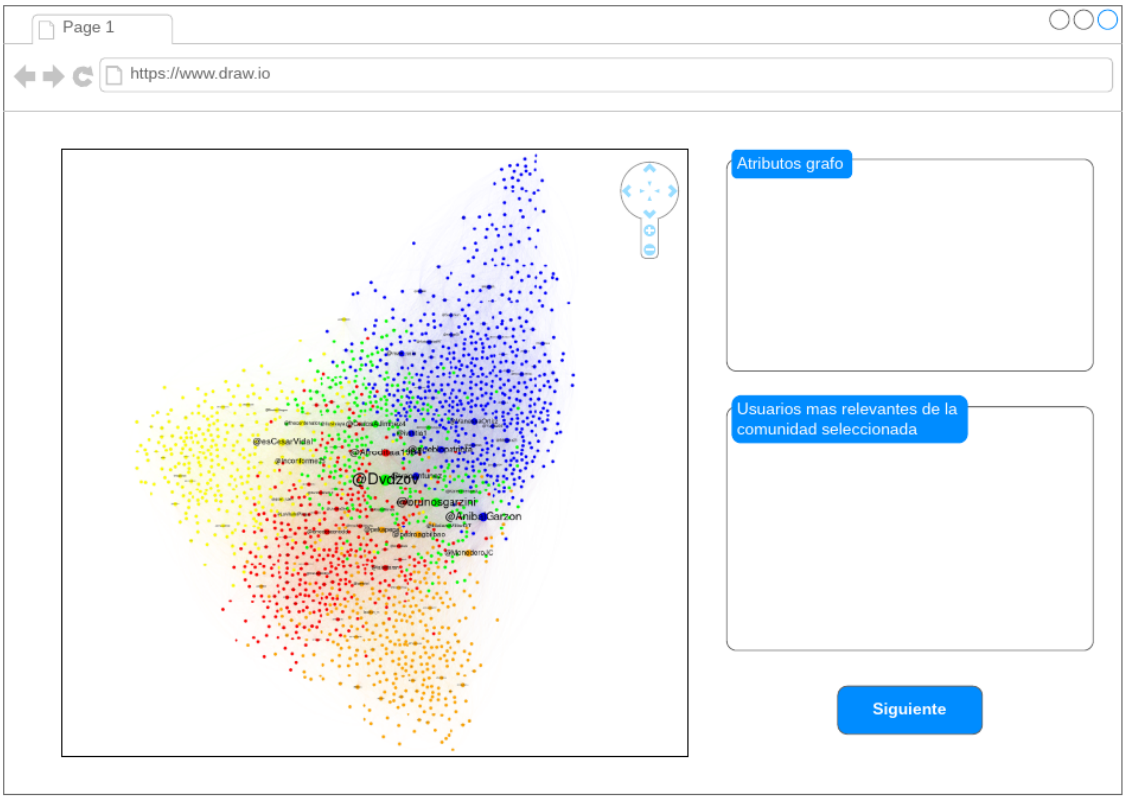


Figura 28: 2ª pantalla

La 2ª pantalla estaría formada por el grafo con la detección de comunidades más medidas genéricas del grafo y un listado de los usuarios más influyentes. Entre esta y la 1ª pantalla se podría realizar una pantalla para seleccionar el intervalo de tiempo que nos interese a través de una gráfica como la de la figura 14 o que se haga automáticamente.

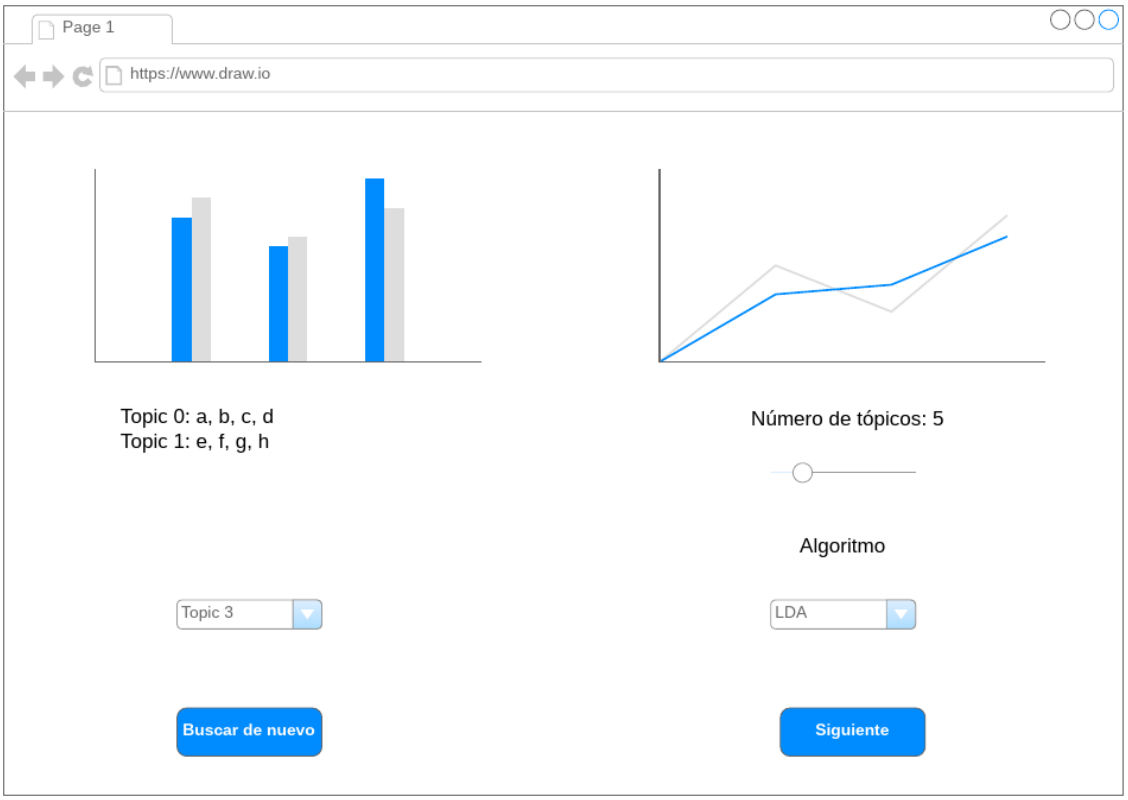


Figura 29: 3ª pantalla

La 3ª pantalla estaría constituida de la elección del método de modelado de tópicos, con la opción de elegir el número de tópicos y un listado de los tópicos. Además, dando la opción de reiniciar la búsqueda en caso de que veamos en el modelado de tópicos que no se está hablando de la temática que nos interesa.

Figura 30: 4ª pantalla

La última pantalla estaría conformada con indicadores de si cumple la cuarta, quinta y sexta categoría de la escala de ruptura y un indicador de en que categoría se encuentra.

B. Objetivos de desarrollo sostenible

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible:

Nº	Objetivos de Desarrollo Sostenible	Alto	Medio	Bajo	No procede
1	Fin de la pobreza			x	
2	Hambre cero				x
3	Salud y bienestar			x	
4	Educación de calidad	x			
5	Igualdad de género				x
6	Agua limpia y saneamiento				x
7	Energía asequible y no contaminante				x
8	Trabajo decente y crecimiento económico			x	
9	Industria, innovación e infraestructuras				x
10	Reducción de las desigualdades		x		
11	Ciudades y comunidades sostenibles	x			
12	Producción y consumo responsables			x	
13	Acción por el clima				x
14	Vida submarina				x
15	Vida de ecosistemas terrestres				x
16	Paz, justicia e instituciones sólidas	x			
17	Alianzas para lograr objetivos		x		

A continuación vamos a justificar los grados de relación de cada punto.

No procede: Para temas de industria, energía, recursos, igualdad y naturaleza, la detección de operaciones de influencia o noticias falsas no debería influir o en caso de que lo hiciese sería de forma muy leve.

Bajo: Los puntos 1, 3 y 8 están mínimamente relacionados debido a que el consumo de información verídica lleva a una sociedad más difícil de engañar por parte de agentes externos o las propias instituciones, además de tener más herramientas para hacer crítica, lo que lleva a una mejora de la sociedad en aspectos tanto económicos como de salud. Respecto al punto 12 es debido a que el identificar fuentes de información veraces hace que la gente no esté gastando tanto tiempo suyo en verificar si la información es verdadera o no.

Medio: El detectar información falsa hace que la gente sea propensa a te-

ner la misma información y que esta sea lo más verídica posible, por lo que puede ayudar a reducir las desigualdades sociales y fomentar alianzas con otras comunidades/países para colaborar en la difusión de información veraz junto a que puedan colaborar conjuntamente en nuevas vías de investigación.

Alto: El enseñar a detectar información veraz permitirá alcanzar una mejor calidad en educación, al igual que permitirá que las comunidades en redes sociales puedan ser moderadas de una forma más justa y permitiendo la crítica constructiva. Además, no solo consolidaría las instituciones, sino que permitiría comprobar si realmente las instituciones lo están haciendo bien.

Bibliografia

[WFN23] Fake news. Consultado a https://en.wikipedia.org/wiki/Fake_news

[CHO20] Michal Choras, Konstantinos Demestichas, Agata Gielczyk, Álvaro Herrero, Pawel Ksieniewicz, Konstantina Remoundou, Daniel Urda, Michal Wozniak. Advanced Machine Learning Techniques for Fake News (Online Disinformation) Detection: A Systematic Mapping Study. Consultado a https://www.researchgate.net/figure/Evolution-of-the-number-of-publications-per-year-retrieved-from-the-keyword-fake-news_fig2_348212162

[BAT21] Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research. Consultado a https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824

[ZHO21] Xinyi Zhou, Reza Zafarani. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys*, Volume 53, Issue 5, septiembre 2021.

[KSD21] Karen Santos-D'Amroim, Májory K. Fernandes de Oliveira Miranda. Misinformation, disinformation, and malinformation: Clarifying the definitions and examples in disinfodemic times. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, Volume 26, marzo 2021.

[CSI23] Misinformation and Disinformation: Thinking Critically about Information Sources. Consultado a https://library.csi.cuny.edu/misinformation

[LEE21] Artificial intelligence may not actually be the solution for stopping the spread of fake news. Consultado a https://theconversation.com/artificial-intelligence-may-not-actually-be-the-solution-for-stopping-the-spread-of-fake-news-172001

[NIM20] Ben Nimmo. The breakout scale: Measuring the impact of influence operations. Consultado a https://www.brookings.edu/wp-content/uploads/2020/09/Nimmo_influence_operations_PDF.pdf

[MET21] Threat Report: Combating Influence Operations. Consultado a https://about.fb.com/news/2021/05/influence-operations-threat-report/

[ZUH17] Mohammed Zuhair Al-Taie, Seifedine Kadry. *Python for Graph and Network Analysis*. Springer International Publishing AG, Cham, primera edición, 2017.

[SNA22] Social network analysis. Consultado a https://en.wikipedia.org/wiki/Social_network_analysis

[LUN05] Lun Li, David Alderson, Reiko Tanaka, John C. Doyle, Walter Willinger. Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. *Internet Mathematics*, enero 2005.

[FOR10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486, 75-174, enero 2010.

[TRA19] V.A. Traag, L. Waltman, N.J. van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Nature*, Scientific Reports, marzo 2019.

[KHA17] Bisma S. Khan, Muaz A. Niazi. Network Community Detection: A Review and Visual Survey. *COMSATS Institute of Information Technology*, agosto 2017.

[KIM13] Yong-Hyuk Kim, Sehoon Seo, Yong-Ho Ha, Seongwon Lim, y Yourim Yoon. Two Applications of Clustering Techniques to Twitter: Community Detection and Issue Extraction. *Discrete Dynamics in Nature and Society Volume*, Article ID 903765, octubre 2013.

[DIL21] Community Detection Algorithms. Consultado a https://towardsdatascience.com/community-detection-algorithms-9bd8951e7dae

[AMA21] Aman Ullah, Bin Wang, JinFang Sheng, Jun Long, Nasrullah Khan y ZeJun Sun. Identification of nodes influence based on global structure model in complex networks. *Nature*, Scientific Reports, marzo 2021.

[SAR19] Dipanjan Sarkar. *Text Analytics With Python*. Springer Science, New York, segunda edición, 2019.

[RIC21] Short-Text Topic Modelling: LDA vs GSDMM. Consultado a `https://towardsdatascience.com/short-text-topic-modelling-lda-vs-gsdmm-20f1db742e14`

[BLE03] David M. Blei, Andrew Y. Ng y Michael I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 993-1022, enero 2003.

[YIN14] Jianhua Yin, Jianyong Wang. A Dirichlet Multinomial Mixture Model-based Approach for Short Text Clustering. *KDD'14*, Association for Computing Machinery, agosto 2014.

[PAA94] Pentti Paatero, Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Fourth International Conference on Statistical methods for the Environmental Sciences 'Environmetrics'*, Volume 5 Issue 2, páginas 111-126, junio 1994.

[AMR19] Short Text Topic Modeling. Consultado a `https://towardsdatascience.com/short-text-topic-modeling-70e50a57c883`

[STO21] 8 Limitations of Topic Modelling Algorithms on Short Text. Consultado a `https://lazarinastoy.com/topic-modelling-limitations-short-text/#6-vulnerability-of-overfitting`

[MEN22] Graph Neural Network and Some of GNN Applications: Everything You Need to Know. Consultado a `https://neptune.ai/blog/graph-neural-network-and-some-of-gnn-applications`

[SMI21] Steven T. Smith, Edward K. Kao, Erika D. Mackin, Danelle C. Shah, Olga Simek y Donald B. Rubin. Automatic detection of influen-

tial actors in disinformation networks. *PNAS*, Vol. 118, No. 4, enero 2021.

[GOL21] Social Network Analysis: From Graph Theory to Applications with Python. Consultado a `https://towardsdatascience.com/social-network-analysis-from-theory-to-applications-with-python-d12e9a34c2c7`

[ZHA14] Huiyuan Zhang, Subhankar Mishra, My T. Thai. Recent Advances in Information Diffusion and Influence Maximization of Complex Social Networks. *Opportunistic Mobile Social Networks*, capítulo 2, julio 2014.

[JIN13] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, Naren Ramakrishnan. Epidemiological Modeling of News and Rumors on Twitter. *The 7th SNA-KDD Workshop '13*, Association for Computing Machinery, agosto 2013.