# Desarrollo y aplicación de un sistema de Inteligencia Artificial para la detección de noticias falsas.

**Proyecto desarrollado por KPI RISK ETHICS & COMPLIANCE**

Coordinadora: Antonia Ferrer-Sapena.

**Colaboradores:** Enrique A. Sánchez Pérez, José Manuel Calabuig, Fernanda Peset, Lluis Miquel García, Isabel Sánchez del Toro, Christian Vidal

## Resumen

Se presenta un nuevo soporte técnico para el análisis de noticias falsas en Internet, en concreto, en Twitter. En la primera parte, se expone una visión panorámica de cuáles son las técnicas de detección de “fake news”, “bots” y otros elementos propios de los “sistemas de desinformación”. Con posterioridad presentamos nuestro nuevo soporte, basado en la estructuración de la información en forma de grafo. Aunque el sistema no tiene implementado aún ningún procedimiento automático de clasificación basado en el riesgo de que una cierta información sea falsa, permite el desarrollo de diferentes métodos heurísticos de análisis por parte de los usuarios potenciales. **Se muestra un caso aplicado: el estudio del conjunto de tweets relacionado con el Día Mundial del Medio Ambiente.**



## I. Introducción – Presentación

Desde las últimas elecciones en los EEUU, que ganó Donald Trump utilizando supuestamente procedimientos de difusión de noticias cuestionables, el análisis de la información en Twitter se ha convertido en una prioridad para muchos gestores de medios de difusión y control de los datos informativos. La gran cantidad de noticias falsas, *fake news*, o en general, elementos de desinformación, que se puede detectar diariamente en las redes sociales, hace muy cuestionable su uso como fuente de información, y condiciona de forma determinante la opinión pública. Al aparecer esta dinámica en temas socialmente sensibles o de importancia fundamental, como son las elecciones presidenciales de los países, se convierte en un problema de primera magnitud; véase por ejemplo el informe de Hindman y Barash, 2018, sobre la influencia de Twitters y los “bots” en las campañas electorales; véase también el reciente informe Xnet, 2019.

La detección automática de noticias falsas en la red sigue siendo difícil. No existe aún un procedimiento de trabajo bien establecido y testado para su detección. Son muchas las grandes empresas que están invirtiendo en ello: por ejemplo, Facebook, Google, Twitter y WhatsApp; pero en muchos casos se continúa utilizando procedimientos heurísticos en el que los analistas (personas físicas) confirman la veracidad de las noticias. Es lo que se conoce habitualmente como “Fact-Checking”, actividad que desarrollan grupos especializados interesados en aportar claridad a un universo

informativo cada vez más confuso. En nuestro país están proliferando empresas y organizaciones que se dedican a ello, entre ellas se encuentra Newtral o Maldita.es.

Tampoco existe una única definición acerca de lo que es una información errónea. El espectro de datos que pueden considerarse fraudulentos desde el punto de vista de la desinformación es muy amplio. Encontramos elementos en los que la información que se ofrece no es totalmente exacta, que van desde lo que pueden ser directamente datos falsos hasta información descontextualizada, entre otras muchas categorías.

A la vista de lo anterior, proponemos en este proyecto un sistema nuevo, que estamos testeando, y que tiene como objetivo **la detección de noticias que pueden ser falsas, confusas, poco claras o erróneas para el lector**.

Hay distintos sistemas que pueden servir para la detección de la posible inexactitud o falsedad en las noticias. Si se estudia desde un punto de vista matemático se puede hacer un primer acercamiento al análisis a través de varias técnicas, que especificaremos con mas detalle en la siguiente sección. Fundamentalmente, hay dos enfoques analíticos. El primero consiste en un acercamiento semántico rastreando palabras sobre un tema, que, relacionadas, pueden identificar el campo semántico de un concepto como primer paso para el análisis. Sin embargo, es conocido que la detección de la información falsa a través de este sistema tiene un cierto nivel de inexactitud (véase Conroy et al, 2015). Otro enfoque puede consistir en analizar la dinámica de difusión de las noticias en el tiempo; para ello se pueden utilizar sistemas de observación de medios en los que la transmisión de la información se hace de manera rápida, como sucede en Twitter. Así, por ejemplo, una de las posibles hipótesis es que las noticias falsas suelen propagarse de manera más rápida que las verdaderas.

A priori se identifican en la actualidad dos posibles vías de detección de la información inexacta y sus sistemas de difusión en la red.

Una de ellas se basaría en la identificación de las propiedades semánticas que comparten los elementos de información erróneos: mismo tipo de expresiones contundentes, fuentes de información sospechosas, junto con la dinámica de su difusión, que podría ser anormalmente rápida en muchos casos.

La otra vía iría ligada a un procedimiento experimental, diseñando un sistema de aprendizaje centrado por ejemplo en el entrenamiento de una red neuronal sobre la base

de un patrón representado en forma de grafo, a partir de un conjunto relevante de noticias falsas.

Nuestra propuesta presenta una tercera vía de análisis, ubicada conceptualmente en el contexto de las fuentes de información *Open Source Intelligence* -OSINT-, que provienen de datos desclasificados y a los que se puede acceder a través de Internet de modo público. La búsqueda se realiza tanto en Internet como en la Internet profunda (aquella que no se encuentra indexada por los motores de búsqueda tradicionales). Así, son fuentes de información los Blogs, las páginas web de empresas, foros, bases de datos gratuitas, e información proveniente de las administraciones públicas de acceso abierto. De este modo, proponemos un nuevo modelo de organización de la información recuperada a través de estas fuentes abiertas, que nos permita construir interactivamente un "Gráfico de Conocimiento", el cual constituiría la base para la creación de un sistema analítico que pueda contar con componentes automáticos y heurísticos.



## II. Contexto y estado de la cuestión

Centrándonos ya en el contexto concreto de la red social Twitter, hasta el momento las investigaciones realizadas de la información en este medio se han basado en tres fuentes principales, a saber: 1/ la obtención de información de los usuarios, 2/ la obtención de grandes conjuntos de tweets y 3/ el establecimiento de las relaciones entre los mismos. Es posible encontrar referencias en la literatura científica a métodos analíticos específicos sobre Twitter, tanto desde el punto de vista de la tecnología utilizada y los planteamientos formales, como desde el punto de vista del posicionamiento contextual sociológico de dichos análisis. Los presentaremos a continuación en esta sección.

Para entender el porqué de algunos de estos métodos, es necesario comentar previamente algunos datos específicos sobre Twitter. Es sabido que este medio habría podido utilizar la duplicación de cuentas, "likes" falsos y otra información falsa en la red, lo que permitiría a los administradores influir en los usuarios de la misma en la toma de decisiones mediante la introducción de información crítica. Atodiresei et al, 2018, han creado un método para controlar tales prácticas. Para este fin, han establecido un sistema de asignación de puntos a tweets particulares, cuyo cómputo permite su clasificación. Básicamente, lo que hace su algoritmo es comparar el tweet con otros

almacenados en sus bases de datos, cuya veracidad está contrastada, siguiendo procedimientos de garantía que incluyen, por ejemplo, la repetición del dato, o el prestigio de la fuente. Para asignar la similitud a cada tweet se usa un método mixto, que considera, por ejemplo, tanto el análisis semántico automático como el de los aspectos formales. Se asigna así una puntuación, tomando en consideración diversos elementos, como la referencia a instituciones reconocidas, y, a mayor similitud con tweets garantizados, mayor puntuación. Además, se ponderan también las características del usuario que emite el tweet, dando lugar a un algoritmo muy preciso que permite la automatización. La credibilidad de los tweets está también relacionada con el análisis de los aspectos psicológicos del usuario, que también ha sido objeto de estudio (Gamallo et al, 2014, Kharde y Sonawane, 2016). Más información sobre iniciativas particulares puede encontrarse en Iozzio, 2016.

Una clasificación general de las técnicas utilizadas en bibliografía aplicadas a la detección de noticias falsas se puede encontrar en el trabajo de Conroy et al, 2015. Según estos autores, tal y como hemos indicado anteriormente, existen dos enfoques metodológicos principales.

1) El primero de ellos consiste en un análisis lingüístico, en el que el contenido de los mensajes fraudulentos se compara con conciertos patrones que son reconocidos como propios de las noticias falsas. Dentro de este enfoque, se pueden reconocer los siguientes tipos: a) El método más simple consiste en el uso de conjuntos de palabras que se identifican con mensajes falsos, que al aparecer en un tweet pueden indicar su falsedad. b) Análisis profundo de la sintaxis, mediante la construcción de árboles estructurales, que en ocasiones identifican los mensajes falsos, o que al menos puede asignarles una probabilidad de fraude (Feng et al, 2012). c) Análisis semántico de los tweets, que estudia la analogía del tweet con contenidos similares verificados. d) Análisis del discurso, que también presenta una estructura característica en el caso de noticias falsas. e) Finalmente, el uso de clasificadores automáticos que podrían diferenciar entre noticias verdaderas y falsas, entrenando redes neuronales o usando técnicas propias de *Support Vector Machines*.

2) El segundo enfoque consiste en el análisis de la red. Metadatos, informaciones externas o de uso, pueden permitir diferenciar informaciones verdaderas y falsas. Se utiliza para ello procedimientos matemáticos automáticos, como técnicas de *Machine Learning*. Se pueden destacar dos aspectos

relevantes en este contexto. a) El uso de *linked data*, que permite el *Fact-Checking* (comprobación de los hechos) contrastando el elemento de información con datos conocidos sobre la realidad existente a los que se puede acceder. El método depende de la posibilidad de consultar las redes de conocimiento existentes, o bases de datos estructuradas públicas disponibles, como la ontología de DBpedia o el *Google Relation Extraction Corpus* (GREC). b) El comportamiento en la red social particular que facilita la transmisión de la información también puede indicar la falsedad de dicha información. Por ejemplo, la inclusión de hipervínculos o de metadatos asociados que pueden ser analizados para establecer la veracidad (Cook et al, 2014).



## III. Iniciativas concretas directamente relacionadas con nuestro proyecto

Tras la exposición del contexto general, podemos centrar la atención en propuestas más concretas. Así, un posible diseño de análisis se basaría en analizar los tweets publicados por un usuario, y estudiar quiénes son sus seguidores, perfiles de los mismos y perfiles de sus contactos. El análisis de esta información tiene una serie de limitaciones, tal y como señala Congosto[1], (Congosto, 2019) ya que utiliza la API de Twitter, que por otra parte permite la composición del grafo y todas las relaciones existentes entre los distintos usuarios, sus seguidores, los tweets y los retweets (véase Congosto, 2017, para experiencias sobre la relación con los bots).

La siguiente tabla muestra estas limitaciones, que son bastante restrictivas y que pueden entorpecer el tratamiento de los datos, sobre todo en lo que respecta al uso de algoritmos para el análisis automático en tiempo real.

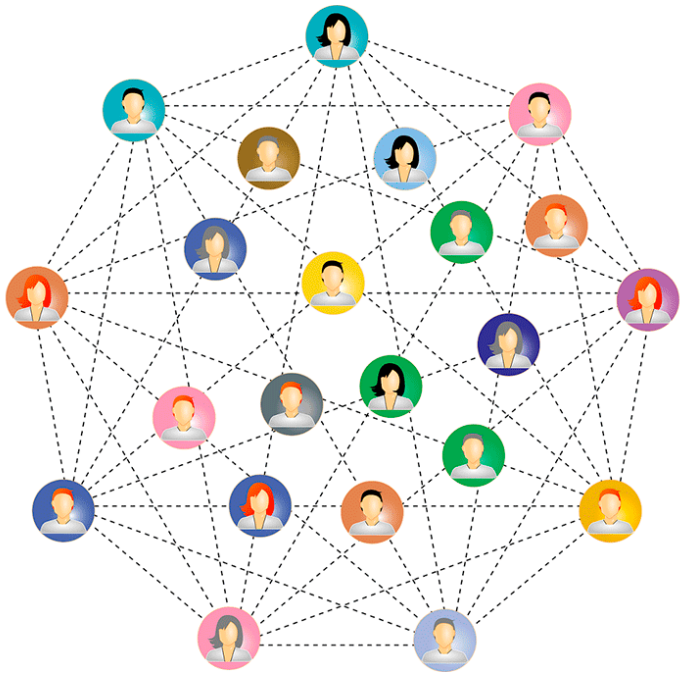
[1] http://www.ctranspa.webs.upv.es/wp-content/uploads/2019/07/t1_datostw.pdf

| Operación | Método | Limitaciones |
|---|---|---|
| --profiles | GET users/show | 3.600 perfiles/hora |
| --followers | GET followers/list | 12.000 perfiles/hora |
| --following | GET friends/list | 12.000 perfiles/hora |
| --relations | GET followers/list GET friends/list | 12.000 perfiles/hora |
| --tweets | GET statuses/user_timeline | 720.000 tweets/hora |
| Buscar tweets | GET search/tweets | 72.000 tweets /hora |
| Bajar tweets en tiempo real | POST statuses_filter | Máximo de 180.000 tweets/hora |
| Relaciones declaradas | GET followers/ids GET friends/ids | 60 peticiones hora Máximo 300.000 ids /hora<br>(5.000 Ids por petición)<br>Con–fast 60 conexiones de usuarios /hora |
| Relaciones dinámicas | No necesita la API | |

*Fuente: https://developer.twitter.com/en/docs/basics/rate-limits*

Como explicaremos más adelante, uno de los avances fundamentales de nuestro método es que estas limitaciones de la aplicación se han logrado salvar.

Desde el punto de vista puramente formal, el análisis posterior propuesto por buena parte de los métodos de detección de noticias falsas, una vez extraídos los datos, se basa en la estructuración de la información en forma de grafos, al igual que se propone en nuestra técnica, donde los individuos suelen ser los nodos y las relaciones, las aristas.



*Fuente: https://revistadigital.inesem.es/informatica-y-tics/teoria-grafos/*

La teoría de grafos aplicada al análisis de la información se basa en la organización de un cierto sistema mediante el establecimiento de vértices, que representan los individuos del sistema, y aristas, que representan las relaciones entre ellos. En general constituye un procedimiento muy útil para el análisis de procesos que involucran a varios actores y cuya evolución se refleja en la dinámica de las relaciones entre ellos.

Es el caso del uso de Twitter por una comunidad de usuarios determinada, se pueden establecer las siguientes relaciones.

Existen unas relaciones declaradas (permanentes en el tiempo) y unas relaciones dinámicas (independientes del tiempo).

| **Relaciones declaradas** | **Relaciones dinámicas** |
|---|---|
| A sigue a B | A retuitea a B |
|  |  |
| A es seguido por B | A es retuiteado por B |
|  |  |
| A y B se siguen | A y B se retuitean |
|  |  |



El elemento innovador que nosotros hemos propuesto, como explicaremos más adelante, es el modo en que podemos acceder a la información para, de forma automática, construir un grafo para el análisis de esa información.

**Generación de elementos de desinformación: los *bots***

La utilización de *bots* para la propagación de noticias falsas en la red ha sido también un elemento estudiado y testeado con profundidad (véase por ejemplo Bessi y Ferrara,

2016). El hecho determinante que ha disparado el interés por el análisis de los *bots* ha sido la evidencia de que están siendo utilizados para manipular la opinión pública a todos los niveles, sobre todo en lo relativo a las preferencias en procesos electorales (véase Bessi y Ferrara, 2016, y Hindman and Barash, 2018).

Estos *bots* consiguen una falsa popularidad de manera muy rápida. Entre sus objetivos está el conseguir una manipulación política. Ha habido otros casos de manipulación a través de estos *bots*, como es el caso de un equipo mexicano de primera división que, estando en un momento bajo de moral, se consiguió cambiar su actitud a través de la creación de un gran conjunto de seguidores falsos. Estos *bots* fueron creados por la empresa mexicana Victory Lab, que se dedica al negocio de la creación de seguidores falsos y *spam* político en las redes sociales[2].

Una de las páginas que mejor ha tratado cómo funcionan los *bots* políticos es: https://botsdetwitter.wordpress.com/2018/06/10/seguidores-falsos-a-fondo/. En ella, se muestra, por ejemplo, que no todos los seguidores falsos se han conseguido a través de la compra.

Un elemento que permite su detección es la visualización de la información de sus seguidores. Así, siguiendo las indicaciones que figuran en la página, encontramos que:

1. Los mejores *bots* políticos son los que tienen una imagen en el perfil, un país de origen identificado, unen su perfil a otras redes sociales, a veces tienen el perfil protegido y emiten algún tweet, pocos.

2. En los *bots* se observa que en un momento determinado aparece una carga de seguidores puntual, permaneciendo sus seguidores en el tiempo. Estas inyecciones de seguidores en momentos puntuales pueden detectarse fácilmente a través de los gráficos de diagramas de dispersión.

3. La API de Tweeter únicamente ofrece información de los seguidores de un usuario, pero no la fecha en que se hace el seguimiento. Esta circunstancia limita la representación gráfica de la información y su análisis.

Sin embargo, aunque los que tienen las características indicadas son a veces difíciles de clasificar, existen otros *bots* que son más o menos fáciles de detectar: los

[2] https://elpais.com/internacional/2018/04/03/actualidad/1522769651_850596.html

automáticos que tuitean en un horario continúo y fijo sin disimular. Hay otros que son más difíciles de seguir ya que van mutando e innovándose. Estos ya no publican en horarios fijos, se adaptan mejor a los horarios de los humanos, reutilizan perfiles de antiguos usuarios, por lo que la detección a veces es casual, por la denuncia de algún usuario, o por un análisis del histórico del usuario en Tweeter.

Una de las metodologías propuestas para la detección de *bots* que más claramente puede seguirse es la presentada por Ferrara y sus colaboradores en una serie de trabajos (Ferrara et al, 2016, Ferrara, 2017, Stella et al, 2018). El siguiente cuadro muestra de forma esquemática como se puede aplicar esta técnica.

| Metodología Ferrara para la detección de *bots* | |
|---|---|
| Características de los sistemas de detección de *bots* | |
| Tipos | Descripción |
| Redes | Las características de la red captan varias dimensiones del patrón de difusión de la información. Las características estadísticas pueden ser extraídas de redes basadas en retweets, menciones y co-ocurrencias hashtag. Ejemplos de ello son el grado de distribución, el coeficiente de agrupación y las medidas de centralidad. |
| Usuarios | Las características del usuario se basan en los metadatos de Twitter y que se encuentran relacionados con una cuenta, incluyendo el idioma, la ubicación geográfica y la hora de creación de la cuenta. |
| Amigos | Las características de los amigos incluyen estadísticas descriptivas relativas a los contactos sociales de una cuenta, como la mediana, los momentos y la entropía o las distribuciones de su número de seguidores, seguidores y mensajes. |
| Tiempo | Las características de tiempo capturan patrones temporales de generación de contenido (tweets) y consumo (retweets); los ejemplos incluyen la señal similar a un proceso Poisson o el tiempo promedio entre dos mensajes consecutivos. |
| Contenido | Las características del contenido se basan en señales lingüísticas computadas a través del procesamiento del lenguaje natural, especialmente el etiquetado parcial del habla; los ejemplos incluyen la frecuencia de verbos, sustantivos y adverbios en los tweets. |
| Sentimiento | Las características de los sentimientos se construyen utilizando algoritmos de análisis de sentimientos de propósito general y específicos de Twitter, incluyendo felicidad, excitación-dominio-violencia y la puntuación de las emociones. |

*Fuente: Congosto, Maríluz. Buscando bots en twitter. http://www.ctranspa.webs.upv.es/wp-content/uploads/2019/07/t2_botstw.pdf*

Aunque no es la única propuesta metodológica para la detección de *bots*, resulta ilustrativa para mostrar cuáles son los elementos que se deben considerar para un sistema de detección y análisis.



## IV. Nuestra metodología

Como se ha explicado en la introducción, nuestro objetivo es definir un método que permita la detección automática de noticias falsas. Concretamente, se ha desarrollado un algoritmo que permite construir un gráfico interactivo a través del uso de la información obtenida de fuentes abiertas de Internet, sobre un tema específico. Aunque en la actualidad se están desarrollando algunas iniciativas directamente relacionadas con técnicas de Inteligencia Artificial, promoviendo el uso de métodos automáticos de detección de noticias falsas (véase por ejemplo Monti et al, 2019, y las referencias en ese trabajo), nuestro procedimiento aporta la base para representar la información de forma que se puedan aplicar sobre ella diferentes estrategias específicas de análisis. Para ponerlo a prueba hemos elegido un evento concreto: el Día Mundial del Medio Ambiente establecido por la Organización de Naciones Unidas para el 5 de junio de 2019. **El análisis se ha centrado en twitter y en páginas web,** con el objetivo de crear un grafo que contenga toda la información relevante sobre el tema.

En el gráfico, los vértices son los usuarios, y las aristas son las relaciones existentes entre los usuarios, por ejemplo, la publicación de tweet sobre un tema concreto. Una vez se establece el conjunto primario, en este caso el de los tweets en castellano sobre este evento, se va enriqueciendo selectivamente siguiendo un patrón determinado. Este conjunto primario sirve para definir el conjunto secundario de usuarios elegidos siguiendo una serie de reglas. Un de ellas es la selección del conjunto de seguidores de los usuarios del conjunto primario. Este segundo conjunto enriquece el gráfico y conserva los vértices y aristas anteriores, añadiendo algunos nuevos, los correspondientes al nuevo conjunto añadido.

La estructura creada nos permite realizar búsquedas particulares. Por ejemplo, se pueden usar los *hashtags* para encontrar subgráficos en los que se pueda obtener la información estructurada que se desea. El gráfico completo puede enriquecerse también añadiendo ciertos elementos semánticos. De esta manera, utilizando rutas aleatorias en el grafo, se pueden establecer relaciones de proximidad entre elementos conceptuales que aparecen en el grafo, lo que permite la caracterización del campo semántico de las

expresiones lingüísticas. Este método proporciona una estructura topológica al grafo, en la que se basa la herramienta que se ha desarrollado.

La estructura formal de grafo es dinámica, ya que las relaciones representadas (los vértices y las aristas) dependen del tiempo. La construcción de este grafo (a este tipo de estructuras se les denomina grafos de conocimiento) es un proceso dinámico, que sigue los siguientes pasos:

1) El conjunto A de elementos de información basado en tweets se ha seleccionado utilizando los siguientes criterios:

   a) El algoritmo transforma A en un grafo GA. Este se implementa en una base de datos con estructura de grafo, utilizando software ya existente. El conjunto de vértices V0 es el primario, definido por los usuarios de twitter que envían y reciben el conjunto original de información A. Además, se pueden añadir algunos vértices nuevos al conjunto original V0 considerando, por ejemplo, tweets relevantes como vértices, hashtags u otros elementos que se convertirán en elementos relacionales relevantes en el gráfico. Los vértices E y los arcos R que los relacionan tiene características propias, que se obtienen a partir de la información relacional proporcionada por la fuente de información, por ejemplo, la dirección de difusión o por las propiedades intrínsecas de los nodos.
   b) El grafo primario GA se enriquece con la introducción de nuevos elementos siguiendo el criterio del analista. En este modelo que se presenta se han introducido los usuarios inicialmente implicados en V0, y los usuarios que siguen a los elementos de V0 en un segundo paso. Alternativamente, si se toma V0 en el tiempo t = t0, el conjunto V1 se puede definir de la misma manera utilizando la información en t = t1, y se puede definir un nuevo grafo mediante el conjunto de información A1 = At0 ∪ At1, donde At0 = A y At1 = A1. Esto proporciona una regla para producir un grafo dinámico.

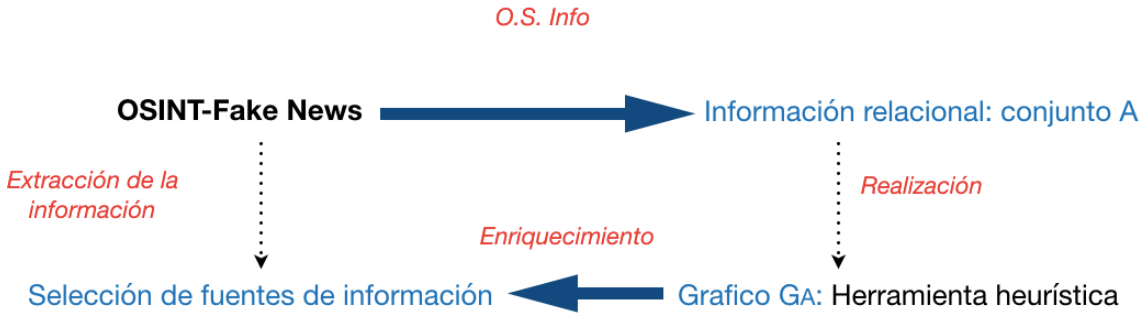
2) La estructura de grafo, que ha sido implementada en Neo4j, es el campo de juego para el analista, y es la base para el procedimiento heurístico que éste quiera desarrollar. Se puede utilizar para estudiar algunas ramas particulares del grafo GA, seleccionando por ejemplo el subgrafo definido por la existencia de algunas palabras clave en el texto de los ítems de información, o siguiendo el subgrafo relacional que comienza en un determinado vértice sospechoso. La

estructura resultante no es todavía un algoritmo para la detección automática de noticias falsas, sino una herramienta de apoyo a los analistas.

3) En el siguiente paso, una vez que la información se ha estructurado, se han implementado unos algoritmos para proporcionar listas de fuentes de información sospechosa, noticias equivocadas, rumores o declaraciones sin fundamento. Esto se ha hecho desde dos perspectivas diferentes:

   1. Definiendo algoritmos basados en propiedades dinámicas de la difusión de noticias falsas ("explosión local" de un elemento de información, detección de fuentes originales sospechosas...), o bien
   2. Entrenando una red neuronal o cualquier otro procedimiento de aproximación sobre el comportamiento de diseminación de información errónea con un gran conjunto de datos de gráficos dinámicos asociados a noticas falsas reconocidas.

El diagrama que se presenta a continuación muestra el esquema de nuestra técnica.




*Fuente: elaboración propia*



## V. Las *fake news* en el Día Mundial del Medio Ambiente

Para probar nuestro sistema de estructuración de la información y detección de noticias falsas, hemos utilizado los tweets que se publicaron el 5 de junio de 2019, Día Mundial del Medio Ambiente. Seguimos los siguientes pasos:

1. Se hizo una búsqueda con los términos "Día Mundial del Medio Ambiente". El grafo de conocimiento fue construido con el conjunto original de tweets que contenían esta expresión y con las relaciones entre ellos.

2. Se enriqueció este conjunto con los tweets de los seguidores de los usuarios involucrados en este conjunto original.
3. Se incluyó también a los usuarios que emitieron esos nuevos tweets.
4. Se hicieron pruebas selectivas en las que se cruzaba la información de estos tweets con algunos hashtags.
5. El resultado final es un grafo extendido de usuarios (vértices) y flechas (aristas) entre ellos, que representan diferentes propiedades ligadas al Día Mundial del Medio Ambiente.
6. En este contexto, se pueden hacer distintos análisis:

    A. Buscar en este conjunto las relaciones existentes entre los usuarios en el conjunto que siguen el hashtag “reciclaje”.

**La figura 1** es la representación gráfica de la difusión de los tweets asociados a un tema determinado. Los puntos verdes indican la línea de tiempo. Cada tweet generado está representado por un punto rosa en cada momento concreto. Los puntos amarillos con el símbolo de una persona representan al usuario que envió el tweet. Las flechas entre ellas indican quién sigue a quién. Cuando el mismo usuario ha publicado más de un tweet, tiene varias flechas que salen del punto amarillo que lo representa.

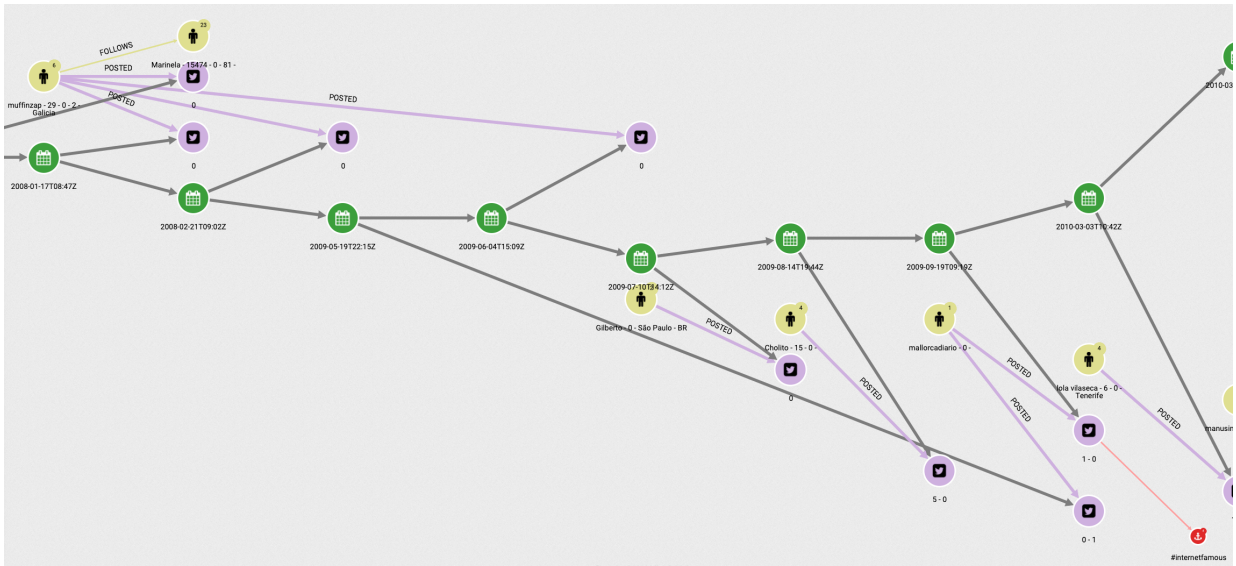



*Figura 1- Ejemplo de grafo para ser sometido a análisis de tweets: tweets y usuarios en una línea de tiempo.*
*Fuente: elaboración propia*

**La figura 2** muestra las relaciones entre dos usuarios. El sistema permite también mostrar el gráfico de relaciones entre usuarios (seguidores).

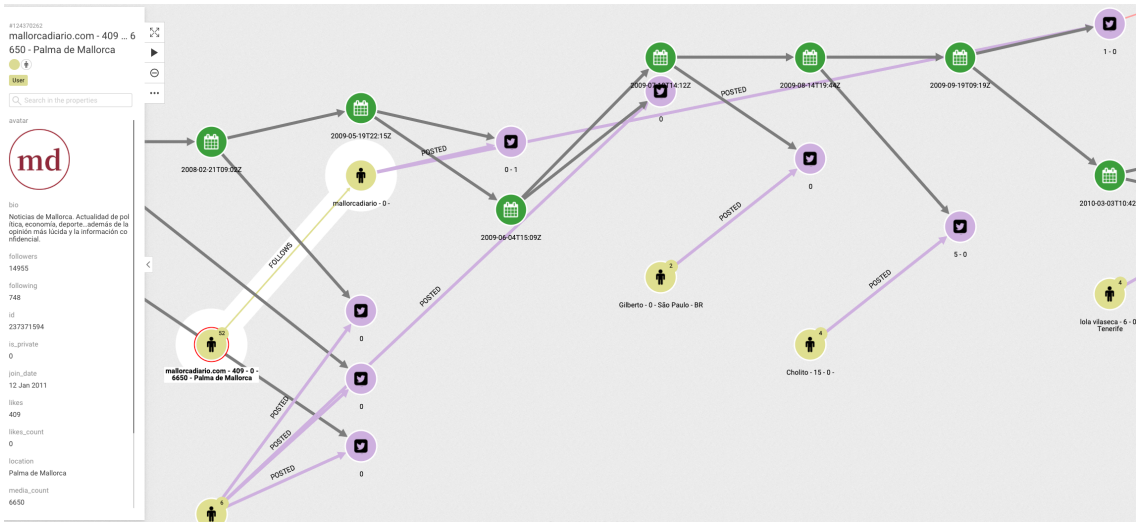



*Figura 2 - Análisis de los Tweets.*
*Fuente: elaboración propia.*

La plataforma también tiene capacidad para mostrar la red de tweets relacionados con un *hashtag* determinado. El mismo gráfico relacional puede ser enriquecido para mostrar todos los elementos que en cada momento fueran considerados relevantes (**Figura 3**).

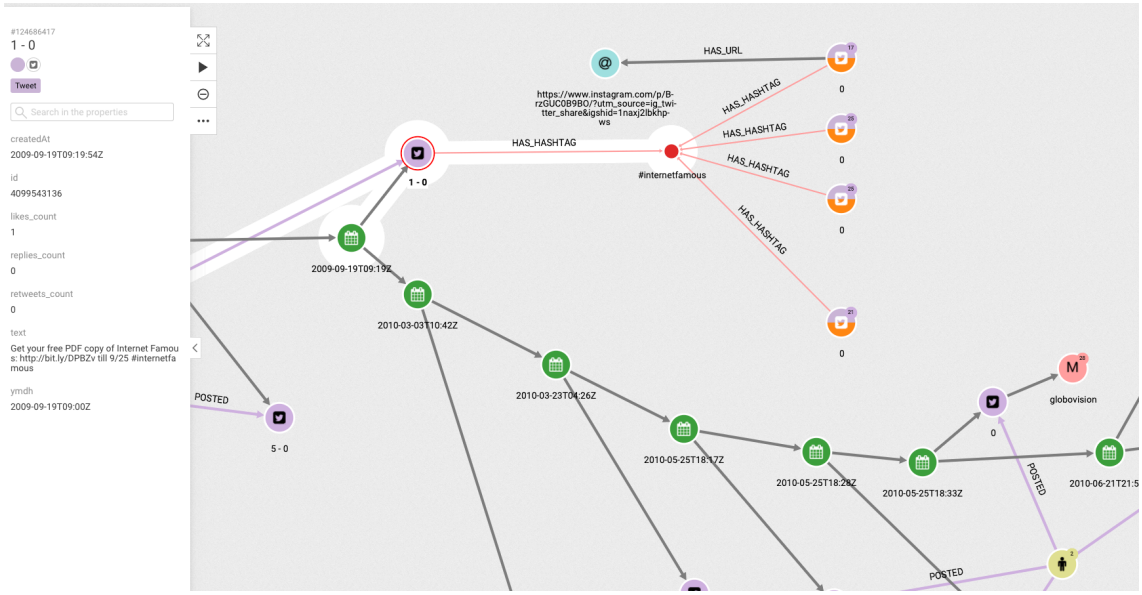



*Figura 3 - Tweets e Hashtag en el tiempo*
*Fuente: elaboración propia*

En la **figura 4** a continuación se muestran todos los tweets junto con los usuarios y otros elementos asociados al tema "Reciclaje".

Se puede también hacer una representación general de los tweets asociados con un tema específico.

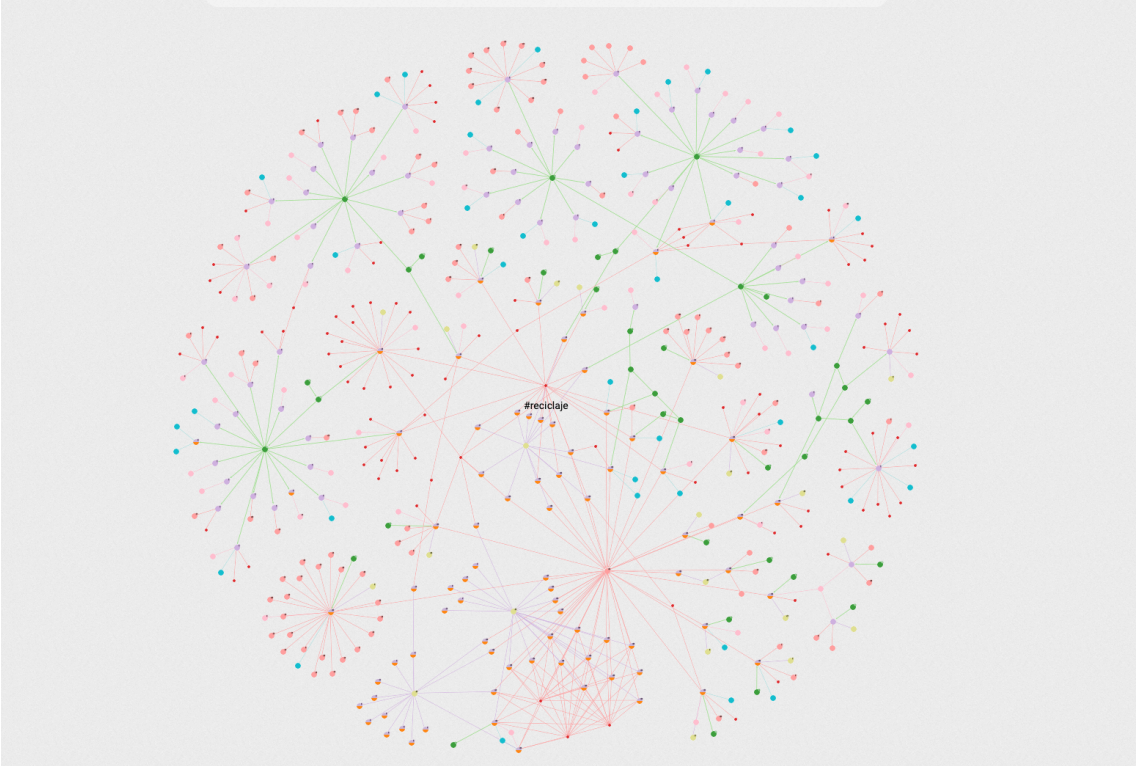



*Figura 4. El grafo completo sobre el tema del "Reciclaje".*
*Fuente: elaboración propia.*

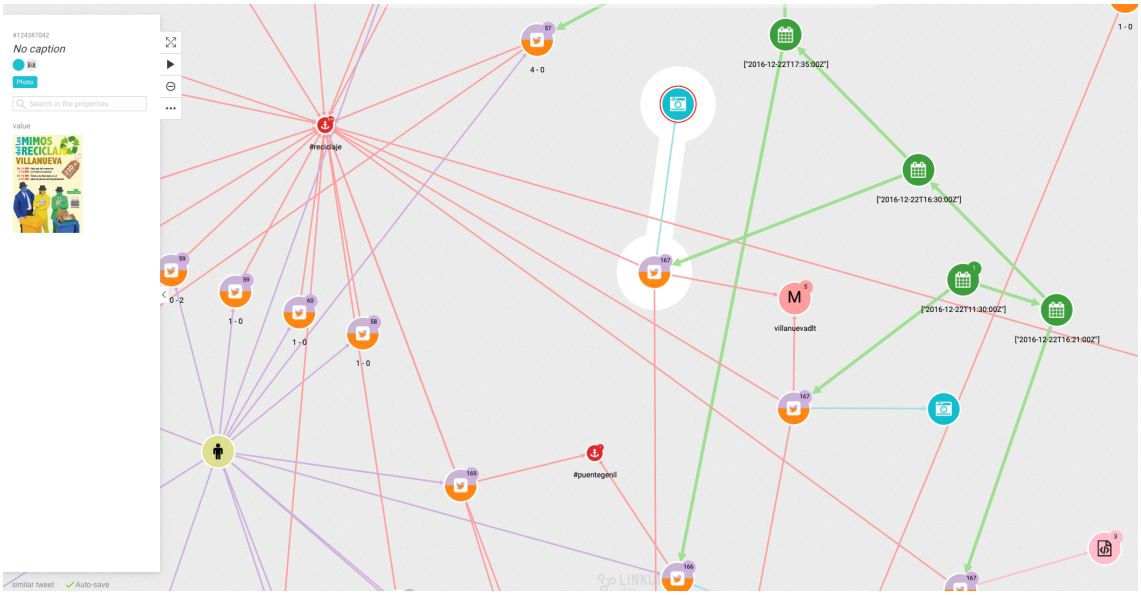



*Figura 5. Cadena de difusión de tweets.*

*Fuente: elaboración propia.*

**La figura** *5 (anterior) nos permite ver:*

- **la introducción de imágenes en la difusión** de la cadena de tweets.
- junto con **el usuario que ha originado la primera acción**.

**La figura 6** a continuación presenta una imagen global del grafo.

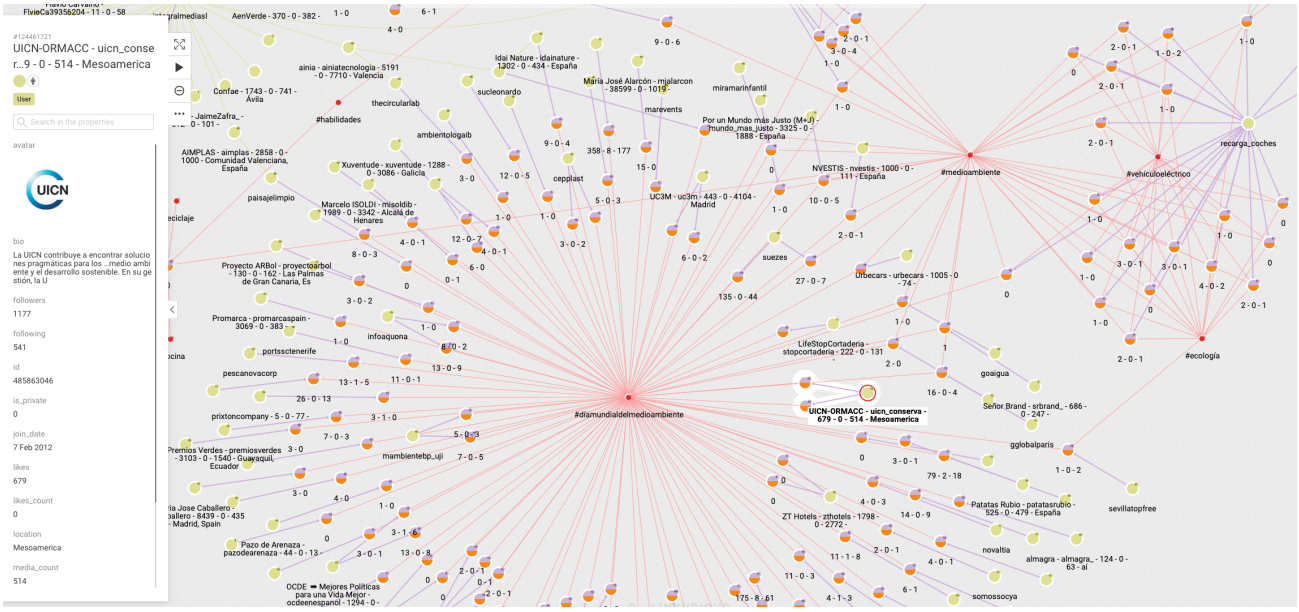



*Figura 6. Imagen del grafo global*
*Fuente: elaboración propia*

Además, se puede navegar por el grafo, planteando sus propias preguntas.

## VI. Conclusiones

6

Esta plataforma es un instrumento, aplicado a un caso de estudio en Twitter, sobre el análisis de la difusión de información a través de Internet, y de las probabilidades de que la información no sea correcta, utilizando fuentes existentes en código abierto.

El grafo de conocimiento que se muestra es una estructura gráfica de los datos asociados a un proceso de difusión de noticias y comentarios sobre un tema concreto de actualidad. Tal y como se ha explicado, se puede ver el hilo de un determinado hashtag o de una cadena de tweets, y de qué modo esta información se propaga en la red, en el tiempo.

Esta plataforma es la base de un sistema heurístico de análisis de información dinámica, que pretende adaptarse al proceso de cómo funciona la difusión de las noticias y otras

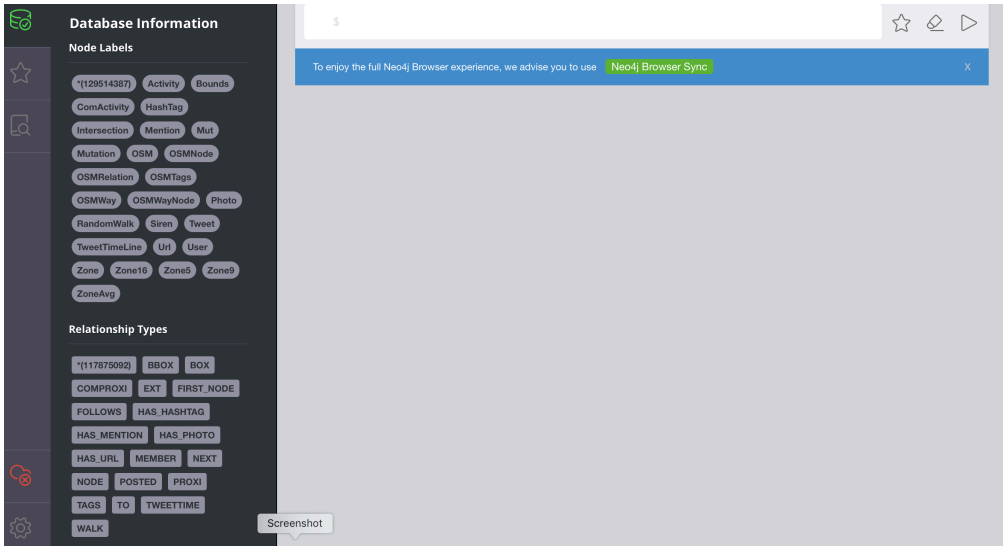
informaciones en la red**. Este sería el punto de partida para continuar implementando algoritmos automáticos de detección de noticias falsas.**

Ahora juega con ello!!

Te ayudamos con tres casos.

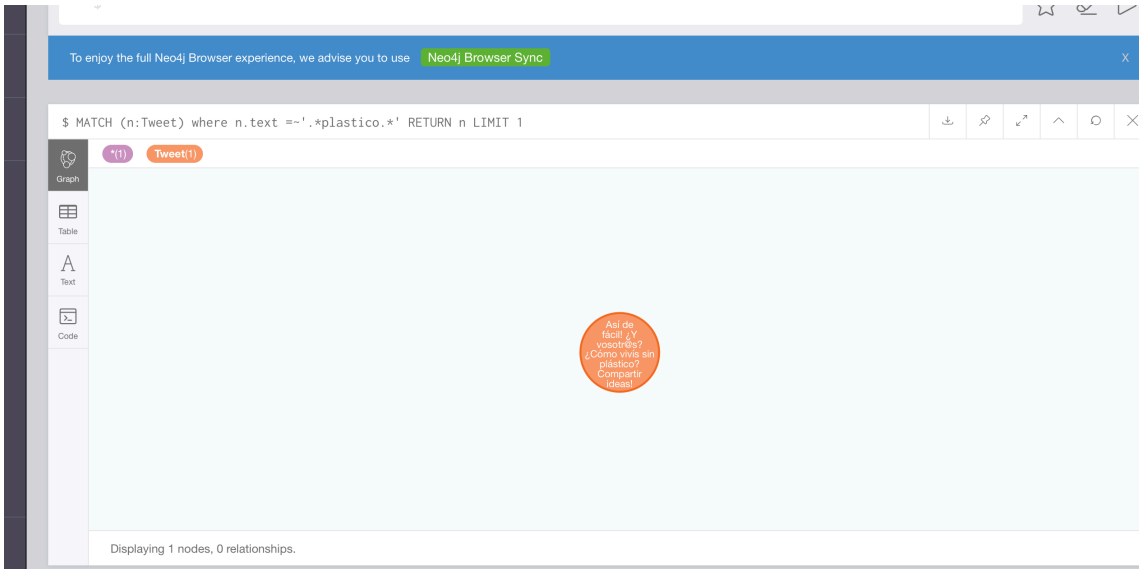
## A. La primera pregunta.

1. Conectate a http://kpi-compliance.com:7474
   Usurario: upv
   Contraseña: Eu9D3nV

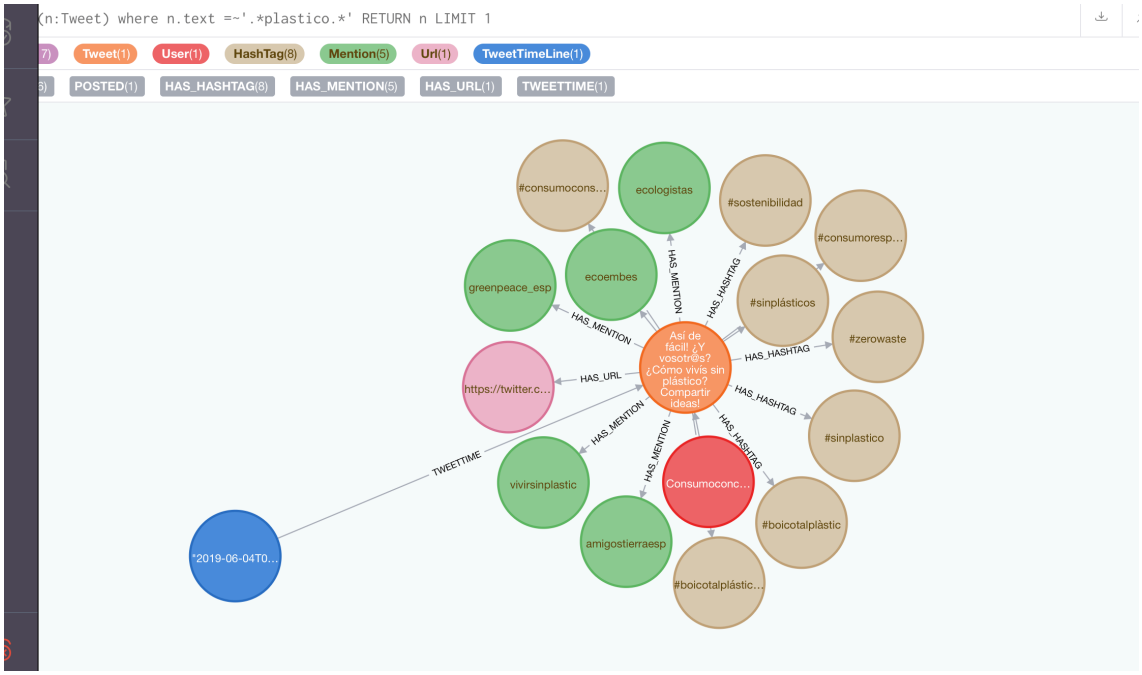



2. **Selecciona** : MATCH (n:Tweet) where n.text =~'.*plastico.*' RETURN n LIMIT 1

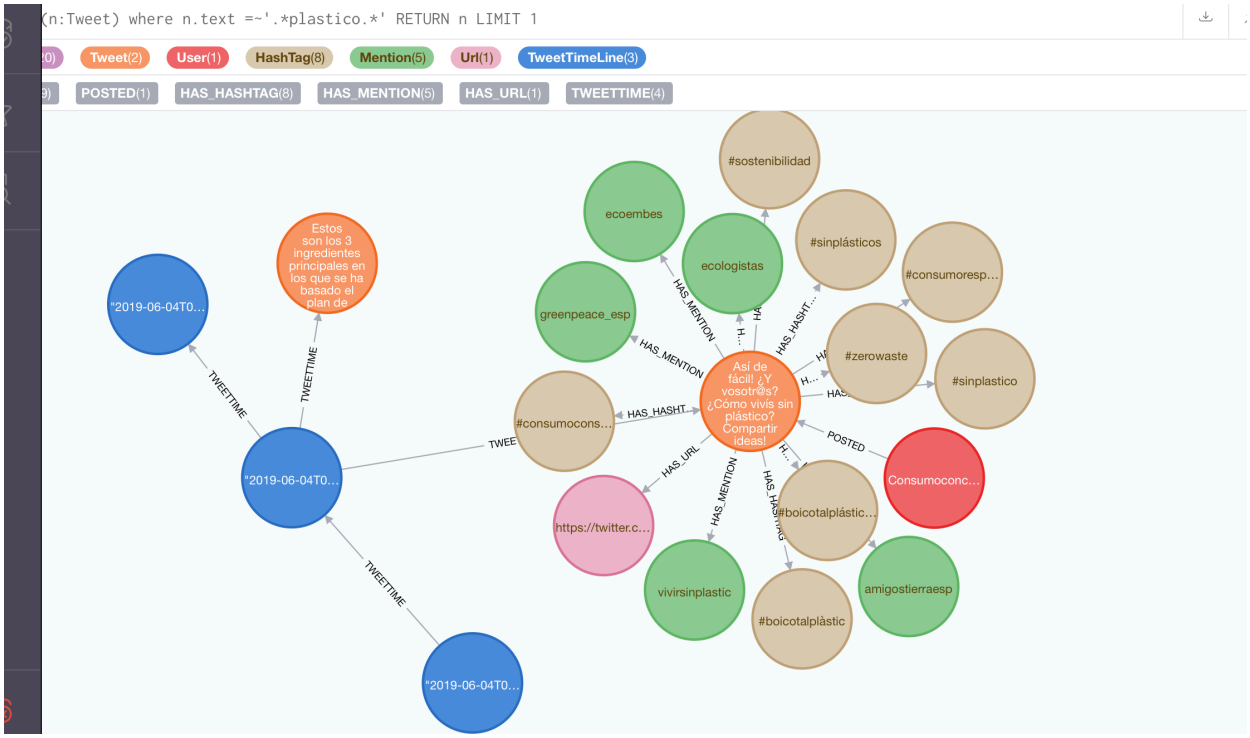
Esta es una búsqueda con una expresión regular, sobre ”plástico”, que devuelve el siguiente resultado:



3. A partir de este nodo central puedes ir pinchando para ver las relaciones.




4. Pinchando en el nodo azul se sigue la línea temporal.



5. En este resultado vemos que aparece un tweet más, que viene relacionado con uno de la misma fecha que el anterior y dos nodos temporales más.

6. Solo con los tweets de contenidos similares, enviados por distintos individuos en distintos momentos, se pueden detectar posibles estrategias de manipulación.

## Bibliografía

Atodiresei, C. S., Tănăselea, A., & Iftene, A. (2018). Identifying Fake News and Fake Users on Twitter. Procedia Computer Science, 126, 451-461.

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. First Monday, 21(11-7).

Congosto, M. (2017). Barriblog: Una prueba de concepto de socialbot realizada en el 2011. Consulta 30 de julio de 2019. http://www.barriblog.com/2017/12/una-prueba-concepto-socialbot-realizada-2011/

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.

Cook, D. M., Waugh, B., Abdipanah, M., Hashemi, O., & Rahman, S. A. (2014). Twitter deception and influence: Issues of identity, slacktivism, and puppetry. Journal of Information Warfare, 13(1), 58-71.

Feng, S., Banerjee, R., & Choi, Y. (2012, July). Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 (pp. 171-175).

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election.

Ferrara,E.,Varol,O.,Davis,C.,Menczer,F.,& Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96- 104.

Gamallo G, Garcia M. Citius: A Naive-Bayes Strategy for Sentiment Analysis on English Tweets. 8th International Workshop on Semantic Evaluation (SemEval 2014), 2014, p. 171-175.

Hindman, Matthew, and Vlad Barash. Disinformation, and Influence Campaigns on Twitter. Knight Fundation. (2018).

Iozzio C. Reuters built a bot that can identify real news on Twitter. Who says AI can't spot fake news? Popular Science, Technology. 2016: https://www.popsci.com/artificial-intelligence-identify-real-news-on-twitter-facebook. Acceso 28 de julio de 2019.

Kharde VA, Sonawane SS. Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications (0975 – 8887), 2016: 139 (11), p. 5-15.

Monti, F., Frasca, F., Eynard, D., Mannion, D., & Bronstein, M. M. (2019). Fake News Detection on Social Media using Geometric Deep Learning. arXiv preprint arXiv:1902.06673.

Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots sustain and inflate striking opposition in online social systems. arXiv preprint arXiv:1802.07292.