




# El impacto del discurso científico en Twitter: El caso de Neil deGrasse Tyson

The impact of science communication on Twitter:  
The case of Neil deGrasse Tyson

 Elena Denia, Investigadora Predoctoral en Comunicación de la Ciencia, Instituto de Gestión de la Innovación y del Conocimiento (INGENIO), CSIC/Universidad Politécnica de Valencia (España) (elenadeniac@gmail.com) (<https://orcid.org/0000-0002-9187-9051>)

## RESUMEN

La percepción social de la ciencia se ha estudiado ampliamente desde mediados del siglo XX. El presente proyecto pretende abordar la interacción ciencia-público en el marco de la vida digital para complementar los estudios clásicos sobre impacto social de la ciencia, en particular en la red social Twitter. Se presenta así una propuesta metodológica con el diseño de un algoritmo que opera sobre conjuntos representativos de tweets para analizar su contenido utilizando técnicas computacionales de minería de datos y procesamiento del lenguaje natural, fácilmente reproducible por otros investigadores y de bajo coste. Para probar la herramienta, se analiza el discurso del popular divulgador Neil DeGrasse Tyson. El impacto de la información se calcula en términos de: 1) likes y retuit; 2) medidas sugeridas para la popularidad y el grado de contenido polémico; y 3) la red semántica. Tras identificar y clasificar los elementos relevantes del discurso por las categorías «ciencia», «cultura», «político-social», «creencias», «medios» y «emocional», los resultados revelan que una transmisión con carga emocional en el mensaje del divulgador despierta una respuesta sustancialmente más profunda en el público, así como la alusión a cuestiones socio-políticas. Además, numerosos conceptos periféricos a la discusión científica suscitan mayor interés que los propios centrales en el discurso. Ambos resultados sugieren que la ciencia interesa en mayor medida cuando va ligada a otros aspectos.

## ABSTRACT

Public perceptions of science have been studied extensively since the mid-twentieth century. The aim of this project is to explore the interaction between science and the public in the digital world as a complement to traditional studies on the societal impact of science, particularly on the social network Twitter. It thus proposes a low-cost, easily reproducible methodology involving the design of an algorithm operating on representative sets of tweets to analyse their content by using computational techniques of data mining and natural language processing. To test this methodology, I analyse the communications of the popular science communicator Neil DeGrasse Tyson. The impact of the information is calculated in terms of 1) likes and retweets; 2) suggested formulas for measuring the popularity and controversial nature of the content; and 3) the semantic network. Relevant elements of the communications are then identified and classified according to the categories of “science”, “culture”, “political-social”, “beliefs”, “media” and “emotional”. The results reveal that content with an emotional charge in the communicator’s message triggers a substantially more profound response from the public, as do references to socio-political issues. Moreover, numerous concepts peripheral to the scientific discussion arouse more interest than the concepts central to the communication. Both these results suggest that science is more interesting when it is linked to other issues.

## PALABRAS CLAVE | KEYWORDS

Twitter, comunicación, ciencia, divulgación, impacto, público, participación, análisis computacional.  
Twitter, communication, science, dissemination, impact, public, participation, computational analysis.

## 1. Introducción

La percepción pública de la ciencia es un aspecto social ampliamente estudiado desde mediados del siglo XX a través de encuestas a la población, como los Eurobarómetros en Europa, los cuestionarios de la National Science Foundation (NSF) en Estados Unidos o los informes de la Fundación Española para la Ciencia y la Tecnología (FECYT) en España, entre otros. Desde sus inicios, estos sondeos tratan de realizar medidas sobre el interés, conocimiento y actitudes de los ciudadanos hacia la ciencia (Davis, 1958), aunque no están exentos de crítica (Bauer et al., 2007; Pardo, 2001). En particular, la comunicación de la ciencia se ha presentado en numerosas ocasiones como una estrategia esencial para favorecer la permeabilidad entre la ciencia y el público (European Commission, 2008).

Siendo la actual corriente en comunicación científica la de involucrar al público en el diálogo (Nisbet & Scheufele, 2009), la labor comunicadora cobra especial interés en su desempeño en las redes sociales, que cuentan con cantidades masivas de cuentas de usuario y una elevada participación. Desde la academia se advierte de que es necesaria una mejor comprensión acerca de cómo afectan los nuevos entornos virtuales a esta práctica (Brossard & Scheufele, 2013), que a su vez brindan la posibilidad de investigar de cerca los debates públicos en torno a temas científicos prestando atención a nuevas voces y contextos diversos (Kapoor et al., 2018; Shan et al., 2014), y en los que el acceso a los contenidos, en muchas ocasiones, se produce sin mediadores. Es por ello que los análisis en redes sociales pueden constituir un complemento útil a las encuestas clásicas de percepción social de la ciencia (Li et al., 2019).

En cuanto a la participación del público, se ha resaltado la necesidad de evaluar su implicación en discusiones abiertas sobre ciencia en estos entornos, teniendo en cuenta aspectos como la facilidad de acceso a los contenidos, el tipo de información diseminada e incluso el tipo de público, entre otros (López-Pérez & Olvera-Lobo, 2019). Se ha sugerido, no obstante, que las audiencias que siguen cuentas de ciencia no suelen interactuar con ellas, siendo utilizadas en gran medida para que el usuario se mantenga actualizado (Álvarez-Bornstein & Montesi, 2019).

Cabe destacar que el contenido que se genera a diario en Twitter está íntimamente ligado a la actualidad científica (Veltri, 2013; Wilkinson & Thelwall, 2012; Zhao et al., 2011), rasgo que le confiere especial atractivo para la investigación en comunicación pública de la ciencia (Büchi, 2016), y que constituye uno de los motivos más robustos para la elección de esta red social en el presente trabajo, junto a su accesibilidad a datos masivos en abierto. Por otro lado, se ha advertido que la mejor estrategia para la difusión en Twitter es enfocarse activamente en acumular seguidores, y no tanto confiar en las búsquedas de palabras clave para visibilizar contenidos científicos (Mohammadi et al., 2018). Así, en este trabajo se ha optado por tomar el perfil del popular divulgador Neil de Grasse Tyson, quien acumula un elevado número de seguidores (más de 13 millones en 2019), para probar la herramienta propuesta.

El objetivo del artículo es presentar las posibilidades que ofrece una herramienta metodológica diseñada para analizar conjuntos de tweets y evaluar el impacto de discursos en Twitter sobre temas científicos, con la intención de revelar dimensiones inexploradas sobre el interés del público en la ciencia y tratando de responder a una pregunta que ha sido inspirada por los resultados de los sondeos clásicos: ¿qué despierta el interés en la ciencia, los propios avances y descubrimientos científicos o bien los aspectos ligados a la vida diaria de los seres humanos, como factores culturales, políticos e incluso emocionales? Se procura así descubrir tendencias sobre qué formas de comunicar la ciencia son más efectivas. La hipótesis de partida es que la red social Twitter, entendida como un espacio de participación ciudadana, puede resultar útil para averiguar estas cuestiones.

Para abordar los desafíos expuestos, la presente investigación incluye las siguientes fases: 1) Elaboración de un algoritmo en el lenguaje de programación Swift capaz de analizar grandes conjuntos de datos y que genere medidas del grado de interés e impacto de la información liberada en Twitter a partir de conversaciones sobre ciencia; 2) Aplicación del algoritmo sobre un conjunto de datos públicos extraídos de la cuenta del astrofísico Neil de Grasse Tyson; 3) Representación gráfica e interpretación de resultados; 4) Valoración de su alcance.

Debe advertirse previamente que existe poco acuerdo sobre qué métodos son confiables en la investigación con Twitter y qué información puede revelarnos (Veltri & Atasanova, 2015), al no haberse desarrollado metodologías rigurosas que permitan análisis sistemáticos fiables (Kahle et al., 2016). No

obstante, existen diversos estudios académicos de carácter exploratorio sobre la comunicación de la ciencia en Twitter en ámbitos específicos, por ejemplo, los que han suscitado mayor atención académica son los relacionados con la percepción del riesgo de los ciudadanos, como los relativos al debate climático (Pearce et al., 2014) o los estudios en temas de salud para entender la postura emocional del público (Becker et al., 2016).

Para el estudio del contenido de los tweets, se utilizan técnicas computacionales para realizar análisis sistemáticos de grandes volúmenes de datos, aplicando minería de textos, procesamiento del lenguaje natural, análisis de redes, etc., y enfoques cualitativos en los que participa un codificador humano, consciente del contexto conceptual del discurso que pretende investigarse y por tanto capaz de revelar cierta información sutil a partir de los tweets (Uren & Dadzie, 2015). A consecuencia, el presente estudio combina una metodología mixta para aprovechar la potencialidad de ambos tipos de análisis.

Conviene apuntar que, debido al formato de Twitter, los usuarios se expresan de forma breve y por tanto seleccionan palabras pertinentes para reflejar sus ideas, lo que en principio permite explorar elementos clave en términos de palabras frecuentes. Así, el contenido de los tweets admite una representación semántica (Narr et al., 2011) que permite tanto estudiar elementos centrales al discurso como periféricos de forma sencilla, asignándoles grados de relevancia.

Siendo además común el llamado «análisis de sentimientos» aplicado al contenido de los tweets, en este trabajo también se trata de medir la carga emocional del discurso, dado que los mensajes emocionales en Twitter tienen una mayor tendencia a retuitearse (Stieglitz & Dang-Xuan, 2013; Veltri & Atanasova, 2015) y reflejan las percepciones emocionales de los usuarios expresadas en lenguaje natural (Dehkharghani et al., 2014), identificadas generalmente a partir de listas predeterminadas de palabras.

Por último, queda resaltar que algunos estudios sugieren que a la hora de tuitear los usuarios utilizan la plataforma más para publicar de manera personal sobre sus actividades diarias que sobre publicaciones informativas (un 80% frente al 20%) (Dann, 2010; Naaman et al., 2010). Nótese aquí que Twitter es a la vez red de información y red social (Myers et al., 2014), y su uso como herramienta de investigación puede ofrecer enfoques muy diversos, desde el estudio del potencial para implicar audiencias mediante el retuit (Kwak et al., 2010), pasando por el desempeño de los periodistas científicos en la plataforma (Arrabal & De-Aguilera, 2016), hasta el rol de los docentes en la motivación y participación de los alumnos (Santoveña & Bernal, 2019). Como nota final, queda puntualizar la idea de que funcionan mejor los perfiles personales que los institucionales en términos de interacciones con el público (Pérez-Rodríguez et al., 2018).

## 2. Material y métodos

### 2.1. Propuesta metodológica

En la presente investigación se planea una propuesta metodológica orientada a la detección de tendencias en los públicos de la ciencia mediante el estudio de los discursos disponibles en abierto en Twitter. El conjunto de tweets a estudiar puede recolectarse fácilmente a través de la interfaz de programación (API) que ofrece la propia plataforma (Twitter, 2019), por ejemplo, utilizando un algoritmo sencillo programado en el lenguaje R. Si es el caso, se recomienda limpiar el texto de los mensajes mediante el uso del paquete *tidytext* R (Silge & Robinson, 2016), que procesa el texto y lo prepara para su análisis, eliminando palabras sin sentido como conectores o stopwords. A continuación, para analizar el contenido de conjuntos representativos de tweets se ha diseñado un algoritmo en lenguaje Swift, cuyas funciones se describen en el presente apartado. Además de las técnicas cuantitativas para realizar análisis sistemáticos, se combina una aproximación cualitativa para la clasificación de los términos por categorías.

En primer lugar, para averiguar de qué trata el discurso contenido en la muestra y cómo puede cuantificarse con el fin de estimar su impacto, conviene aclarar que la unidad de investigación básica y discreta utilizada en el método propuesto es la palabra o término (Blei et al., 2003). Las medidas para estimar dicho impacto están basadas en su frecuencia de uso y en el número de retuits y de likes que tenga asociados, apropiadamente normalizados en base al número de veces que aparece el término en la muestra. Además, para revelar en qué medida cierto tipo de información es más interesante que otra para el público receptor, se proponen dos coeficientes adicionales, la «popularidad» y la «polemicidad». La idea que subyace al indicador de popularidad es que cuanto más se retuitea una palabra (es decir, acumula

mayor número de retuits) esta se considera más popular, teniendo en cuenta también su frecuencia de aparición en la muestra. Por ejemplo, una palabra poco habitual que posea muchos retuits se considerará especialmente popular. Para su definición, se ha empleado la variable *retweetRate* (ratio de retuits) de manera que proporcione una medida del interés que despierta el término, es decir, que se comparta en mayor o menor grado. En este sentido la popularidad puede ocurrir tanto de forma positiva como negativa, y así un contenido que cierto usuario desaprobe puede retuitearse y ganar visibilidad. Por su parte, la polemicidad, entendida como el grado de controversia de la información, se define como el cociente entre el *retweetRate* y el *favoriteRate* (ratio de likes), por lo que el valor será mayor si el contenido del tweet se retuitea generosamente, pero acumula menos likes. Las variables mencionadas y los indicadores de popularidad y polemicidad para palabras y categorías se presentan en la Tabla 1, siendo por definición comparables entre diferentes conjuntos de datos.

| Tabla 1. Variables integradas en el algoritmo |                                                                                                                           |                                                                                                                                |
|-----------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Variable                                      | Definición                                                                                                                | Ecuación                                                                                                                       |
| Frecuencia (palabra)                          | Número de apariciones de una palabra sobre el número total de palabras clasificadas presentes en el discurso del usuario. | $\text{freq}(w_i) = \frac{\text{wordCount}(w_i)}{\sum_j^k \text{wordCount}(w_j)}$                                              |
| favoriteRate                                  | Número de likes que acumula una palabra determinada sobre el número de tweets que contienen dicha palabra.                | $\text{favoriteRate}(w_i) = \frac{\text{favoriteCount}(w_i)}{\#\text{tweetsContaining}(w_i)}$                                  |
| retweetRate                                   | Número de retuits que acumula una palabra determinada sobre el número de tweets que contienen dicha palabra.              | $\text{retweetRate}(w_i) = \frac{\text{retweetCount}(w_i)}{\#\text{tweetsContaining}(w_i)}$                                    |
| Popularidad (palabra)                         | Número de retuits (ponderados) sobre la frecuencia de una palabra determinada.                                            | $\text{popularity}(w_i) = \frac{\text{retweetRate}(w_i)}{\text{frequency}(w_i)}$                                               |
| Polemicidad (palabra)                         | Número de retuits (ponderados) sobre el número de favoritos (ponderados) de una palabra determinada.                      | $\text{polemicity}(w_i) = \frac{\text{retweetRate}(w_i)}{\text{favRate}(w_i)}$                                                 |
| Popularidad (categoría)                       | Número de retuits (ponderados) sobre la frecuencia de una determinada categoría.                                          | $\text{pop}_{\text{cat}} = \sum_{\text{cat}} \frac{\text{wordCount}(w_i) \cdot \text{pop}(w_i)}{\text{wordCount}(\text{cat})}$ |
| Polemicidad (categoría)                       | Número de retuits (ponderados) sobre el número de favoritos (ponderados) de una determinada categoría.                    | $\text{pol}_{\text{cat}} = \sum_{\text{cat}} \frac{\text{wordCount}(w_i) \cdot \text{pol}(w_i)}{\text{wordCount}(\text{cat})}$ |

Nota. Donde  $w_i$  es una palabra o término de la muestra y  $\text{cat}$  una categoría.

Por otro lado, para averiguar el grado de interés que suscitan los contenidos científicos por sí mismos o bien correlacionados con otros tipos de información en el tweet, se plantea una clasificación de palabras por categorías, método que ya ha sido empleado en otros estudios que investigaron en Twitter el interés en ciencia en comparación con otras temáticas (Zhao et al., 2011). En el presente artículo las categorías propuestas, descritas en la Tabla 2, son: «ciencia», «cultura», «político-social», «creencias», «medios» y «emocional», que se han definido mediante un criterio propio inspirado por trabajos previos.

En particular, la categoría «cultura» viene motivada por las propuestas que invitan a considerar los factores culturales en los estudios de percepción social de la ciencia (Bauer et al., 2012; Pardo, 2001); la «político-social» por estudios en Twitter que resaltan preocupaciones de esta índole en relación con controversias científicas, usuales en el ámbito del cambio climático (Pearce et al., 2014); la de «creencias» por las interacciones frecuentes entre ciencia y religión así como por el interés creciente en estudios sobre pseudociencias (Moreno-Castro et al., 2019); la de «medios» por considerarse pertinente en estudios de comunicación; y la «emocional», por los estudios de análisis de sentimientos en tweets, sumados a la observación de que «una conexión emocional [...] puede ser una poderosa "entrada" a una experiencia científica para los no expertos, capturando la atención inicial y aumentando los sentimientos de vinculación con el comunicador u otros participantes» (Kaiser, 2014: 28).

La clasificación de palabras en categorías se realiza de forma manual en base a su significado, ya que los métodos de investigación asistidos por ordenador resultan poco adecuados para el pretendido análisis de contenido al asumir que los términos tienen el mismo significado en cualquier contexto (Matthes & Kohring, 2008), mientras que el empleo de un codificador humano conduce a una mejor interpretación del contexto de la discusión.

**Tabla 2. Descripción de las categorías para analizar los discursos en Twitter**

| Categoría       | Descripción                                                                                                                                                                                                          |
|-----------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ciencia         | Palabras que se refieren a cuestiones científicas, incluyendo: conceptos científicos («planet», «physics»), metodológicos («deduction») y personajes (ya sean históricos como «Galileo», o modernos como «bgreene»). |
| Emoción         | Palabras que contienen carga emocional, que expresan sentimientos, incluyendo: emociones («happy», «fear», «awesome»), expresiones («yup», «yeah») y apreciaciones («lovely»).                                       |
| Cultura         | Palabras que se refieren a la cultura: industria («film», «song»), personajes populares («ladygaga») y actividades culturales («football»).                                                                          |
| Político-social | Palabras que se refieren al contenido político y los aspectos sociológicos («electorate», «president», «americans») así como indicadores sociales («poverty»).                                                       |
| Medios          | Palabras referidas a los medios de comunicación, incluyendo: redes sociales («facebook») y mass media («audience», «news»).                                                                                          |
| Creencias       | Palabras que se refieren a creencias: religión («faith»), pseudociencia, esoterismo y ufología («homeopathic», «astrology»).                                                                                         |

Gracias al diseño del algoritmo, esta categorización es relativamente sencilla al disponer las palabras en orden de relevancia en una lista en base a su frecuencia de uso y la acumulación de likes y retuits. Por otro lado, el algoritmo también genera dos archivos con nodos y aristas dispuestos para su representación visual en forma de red semántica (por ejemplo, mediante el popular software Gephi), en la que se podrán apreciar relaciones entre las métricas propuestas y el contenido de los tweets recabados, como visualizar el peso de cada categoría en comparación con las demás u observar si los conceptos científicos presentes en la muestra de datos son centrales en el discurso o bien periféricos. La consecución de tareas para el análisis viene detallada en la Tabla 3.

**Tabla 3. Tareas propuestas para el análisis de una muestra de tweets**

| Herramienta        | Tarea                             | Descripción                                                                                                                                                                                                                                                                     |
|--------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Algoritmo en Swift | Estadísticas de los términos      | Se definen distintos parámetros estadísticos basados en frecuencias de uso, retuits y likes: la frecuencia «frequency» sobre el número total de palabras; los ratios de retuits y de likes (retweetRate y favoriteRate respectivamente); y los índices popularity y polemicity. |
|                    | Lista de palabras                 | Se genera un archivo CSV que contiene una lista de palabras ordenadas por su relevancia en el conjunto de tweets mediante una medida aproximada calculada a partir de las variables ponderadas: frequency, retweetRate y favoriteRate.                                          |
|                    | Asociaciones de palabras          | Se identifican las coincidencias de palabras en los tweets de forma automatizada.                                                                                                                                                                                               |
|                    | Visualización de la red semántica | Se preparan los datos para la representación visual de las coincidencias entre palabras en los tweets para evaluar el grado de centralidad de los conceptos que resultan más atractivos.                                                                                        |
| Codificador humano | Categorización de palabras        | Se realiza una clasificación manual de términos por categorías («ciencia», «cultura», «política y social», «creencias», «medios» y «emocional») en el archivo CSV con la lista de palabras.                                                                                     |

Por último, conviene resaltar que la lista de palabras categorizadas es acumulativa, por lo que siempre crecerá en tamaño y favorecerá así análisis posteriores aplicados a nuevos conjuntos de tweets, para los que gran parte del trabajo ya estará hecho, a falta de clasificar las palabras más relevantes del nuevo discurso ejecutando el algoritmo. Esto proporcionará análisis cada vez más afinados de los discursos de ciencia, fácilmente reproducibles por otros investigadores y que supongan un bajo coste en cuanto a recursos intelectuales y económicos, en contraposición a las encuestas de población que requieren grandes despliegues y sustanciosa financiación, y cuyos resultados son estancos. Naturalmente, este diseño algorítmico no se restringe al ámbito de la ciencia, ya que sus variables son compatibles con otros espacios conceptuales. Queda señalar que esta herramienta está a disposición del investigador que la requiera, previa solicitud vía email a la autora.

## 2.2. Confiabilidad de la categorización

Para validar la categorización de palabras, seis miembros del grupo de investigación ScienceFlows (ScienceFlows, 2019) realizaron de forma independiente el ejercicio de clasificar, en base a las categorías creadas, el conjunto de las primeras 50 palabras relevantes detectadas por el algoritmo de la muestra de datos (ver apartado 3.1). Para cada palabra se ha calculado el grado de acierto en base a la clasificación de

la autora, considerándose como válida cuando supera un umbral mínimo del 0.75 sobre 1. Se ha estimado así un nivel de confiabilidad de la clasificación manual del 82%, si bien solo se trata de una estimación orientativa útil para explorar ciertas tendencias en el discurso a estudiar.

### 2.3. Limitaciones de Twitter para la extracción de datos

Si bien la API de Twitter es gratuita y posibilita el acceso a millones de tweets incluyendo metadatos, cada consulta devuelve alrededor de 3000 tweets, los cuales constituyen muestras aleatorias de conjuntos más grandes, por lo que existen limitaciones en cuanto a la información que el investigador puede recabar. Para las búsquedas de palabras clave, debe considerarse que la muestra devuelta por la API pertenece a los últimos 9 días, por lo que la realización de estudios transversales en este medio requiere poner en funcionamiento un mecanismo para la extracción sistemática de datos en tiempo real, obteniendo muestras para periodos más amplios. En cambio, cuando se recaban los tweets de cuentas de usuario específicas, la muestra aleatoria incluye tweets desde la apertura del perfil.

Otra debilidad manifiesta en el empleo de Twitter es la somera distinción entre países que ofrece la herramienta, resultando en datos demográficos insuficientes y en una descripción homogénea del público. Además, algunos autores apuntan que los usuarios de las redes sociales solo son un subconjunto del público general, por lo que no resultarían representativos (Murphy et al., 2014). No obstante, puede aducirse que a medida que el acceso a Internet aumenta en todo el mundo, cada vez son más las personas representadas en las redes sociales. En todo caso, dado que el presente estudio tiene como objetivo explorar tendencias, esta problemática no resulta un inconveniente grave.

## 3. Análisis y resultados. El impacto del discurso público de Neil deGrasse Tyson

### 3.1. Datos

Para probar la herramienta, se extrae un conjunto de tweets a través de la API de la cuenta de un usuario específico, Neil deGrasse Tyson («@neiltyson»). El archivo resultante contiene el texto completo de los mensajes junto a una serie de propiedades almacenadas en columnas con sus valores asociados, siendo los datos de especial relevancia para el análisis del impacto de la información el número de likes y el número de retuits. La muestra resultante, que solo incluye tweets escritos por el usuario (no retuits), contiene 3.005 tweets emitidos entre el 2012-10-05 y el 2019-06-19. Esta descarga supone el 49,5% de los tweets publicados por Tyson en las fechas de la consulta, sobre un total de 6.974 tweets. Tras limpiar el texto, se obtiene un archivo con 24.484 palabras relevantes y sus estadísticas asociadas.

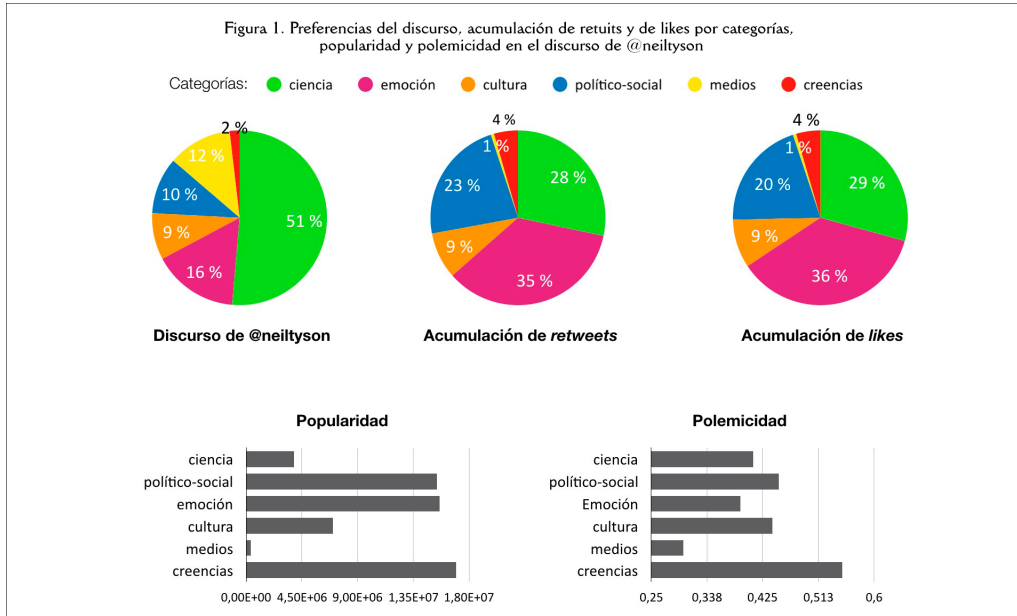
### 3.2. Análisis

Tras identificar las palabras más relevantes con el algoritmo, los primeros 1.250 términos se han clasificado manualmente y permiten hacerse una idea sobre las preferencias temáticas del usuario en cuestión. Entre las más utilizadas, aparecen conceptos científicos como «mars» «space» o «physics», un resultado esperable dado que se trata de un divulgador que opera en el dominio de la astrofísica. Hay asimismo conceptos muy frecuentes alejados de la ciencia en sí; por ejemplo, la palabra «film» hace referencia a la industria cultural del cine y la palabra «happy» a un estado emocional.

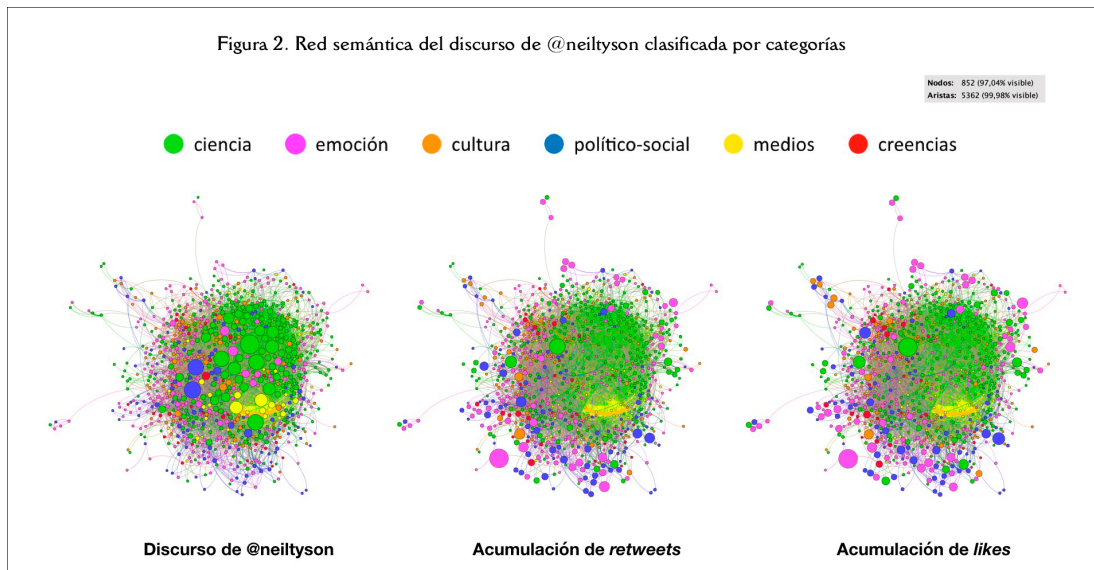
En la Figura 1 se representan los datos en gráficos circulares de sectores prestando atención a la fracción de cada categoría, que se muestra de forma porcentual. Estas gráficas muestran el peso de las distintas categorías en el discurso de Tyson en base a: 1) Las preferencias temáticas del divulgador; 2) La acumulación de likes por categoría; 3) La acumulación de retuits por categoría.

En el primer caso, queda manifiesto que la frecuencia de uso de palabras con contenido científico es alta en el discurso de Tyson, conformando más de la mitad de la muestra categorizada, un 51%. Además, su discurso cuenta con un 16% de expresiones emocionales; un 12% de palabras que se refieren a los medios de comunicación (por ejemplo, citando su propio programa de radio «startalk»); un 8% de palabras con contenido político y social; un 9% de las palabras relacionadas con el ámbito cultural; y un 2% relacionado con creencias. A continuación, para estimar el impacto del discurso, se muestra cómo estas preferencias temáticas del usuario suscitan diferente cantidad de likes y retuits para cada categoría. Al representar los términos del discurso de Tyson, destaca que el impacto es sensiblemente mayor en aquellos que contienen

carga emocional (por ejemplo «joy», «shit», «hostile» o «cry»), seguidos de los que aluden a cuestiones sociales o políticas; mientras que para los términos científicos se reduce. Por otro lado, los que contienen información cultural y aquellos que se refieren a creencias mantienen un impacto proporcional. Los porcentajes para la acumulación de likes son: «ciencia» 28%, «emoción» 35%, «cultura» 9%, «político-social» 23%, «medios» 1%, «creencias» 4%; mientras que para la acumulación de retuits son: «ciencia» 29%, «emoción» 36%, «cultura» 9%, «político-social» 20%, «medios» 1%, «creencias» 4%.



Después, para medir lo popular o controvertido que es el contenido del discurso de Tyson, se calculan los coeficientes de popularidad y polemicidad para las diferentes categorías, también representados en la Figura 1 mediante un gráfico de barras (siendo medidas comparativas cuyos valores no tienen significado por sí mismos).



Nota. Para su representación se ha empleado el programa Gephi, aplicando el algoritmo Force Atlas sobre una muestra de 852 nodos y 5.362 aristas. En la red, los nodos se repelen y las aristas se atraen.

Los valores más populares son los relacionados con creencias, con cuestiones políticas o sociales y con carga emocional, siendo el más impopular el relacionado con los medios. Asimismo, el grado de contenido polémico en el discurso aparece claramente por encima del resto de categorías cuando se trata de elementos relacionados con creencias.

El último paso es averiguar qué sucede cuando la información científica se combina con otros tipos de información, cuestión que se explora mediante la representación de la red semántica del discurso de Tyson también diferenciada por categorías, mostrando las relaciones y conexiones entre palabras. Esta visualización puede encontrarse en la Figura 2, en la que el tamaño de los nodos representa: 1) Las preferencias del discurso por parte del divulgador (un nodo es más grande a medida que cierto término aparece más veces en la muestra); 2) Los términos que reciben más retuits; 3) Los términos que reciben más likes. Nótese que el usuario construye su discurso mediante conceptos científicos centrales (red de la izquierda), pero se observa en las otras dos redes conceptuales representadas que el impacto de la información sobre los usuarios receptores es superior para términos periféricos al discurso, especialmente para las categorías «emoción» y «político-social».

#### 4. Discusión y conclusiones

Los resultados obtenidos en el apartado previo tienen un doble escenario, el propiamente metodológico y el correspondiente al caso elegido. Según el análisis del estudio de caso con la herramienta planteada, el primer resultado llamativo es que el contenido científico pierde relevancia con respecto a otros tipos de información en el conjunto de tweets recabados del divulgador Neil deGrasse Tyson. Aunque el discurso predominante de Tyson se compone de palabras asociadas con contenido científico, son los tweets con carga emocional los que reciben mayor atención por parte del usuario receptor, tanto en términos de likes como de retuits, seguidos de los que aluden a contenido social y político. Por su parte, la categoría de medios de comunicación es claramente la que ofrece el resultado más débil en cuanto al interés que suscita, a pesar de que su presencia en el discurso es elevada, presentando además la tasa de popularidad más baja. Este fenómeno queda abierto a interpretación. A la luz de estos resultados, este trabajo constituye una aproximación con datos que sugiere que la ciencia por sí sola no interesa al público tanto como otros temas. Si bien solo se estudia un caso específico, el de Tyson, debe considerarse que se trata de un divulgador con gran influencia y elevado número de seguidores, habiéndose utilizado una muestra con la mitad de sus tweets en este trabajo. Nótese también que los mensajes emitidos por el divulgador llegan, en primer lugar, a sus seguidores, un público supuestamente interesado en la ciencia (si bien también se difunden a través del retuit).

En cuanto a los valores generados para los dos índices propuestos de popularidad y polemicidad, se comprueba que las diferencias patentes entre ambos para las distintas categorías pueden dar cuenta de rasgos interesantes. Así, la categoría «ciencia», a pesar de no ser especialmente popular para los usuarios receptores, presenta cierto nivel de contenido polémico, quizá por el activismo social de Tyson en cuestiones como la crisis climática. Mientras que la idea de la polemicidad estriba en investigar qué temas generan más debate o controversia, la popularidad es una medida de la atención que reciben, ya sea positiva o negativa. Las características que se manifiestan con estos cálculos de otro modo podrían pasar desapercibidas si hay una baja incidencia de palabras utilizadas en la muestra sobre determinada categoría, como sucede con «creencias», que presenta un carácter radicalmente popular y controvertido, pero escasamente aparece en el conjunto estudiado. Presumiblemente debido a su naturaleza sensible, son cuestiones que no abundan en el discurso del usuario estudiado pero que gozan de gran impacto en el ecosistema Twitter. Como nota adicional, debe considerarse que, si bien los resultados basados en likes y retuits no difieren mucho entre sí, una interpretación que parece plausible es que la acción del retuit favorece la visibilidad del mensaje y, por tanto, la cantidad de likes potenciales crece.

Respecto a las coincidencias de palabras en la red semántica, representada para evaluar el grado de centralidad de los términos más o menos atractivos, queda patente para la muestra examinada que los conceptos periféricos a la discusión científica, es decir, aquellos que tratan cuestiones adyacentes, presentan mayor interés que los centrales, siendo además mayoritariamente no científicos. Una interpretación es que existen temas puntuales que suscitan una atención repentina, pero sin ser los



habituales en el discurso de Tyson, aspecto que requiere un análisis más profundo para examinar las causas. Por otro lado, al observar las redes semánticas considerando el número de likes y de retuits que acumulan los nodos, los resultados también son similares entre sí, por lo que el motivo de nuevo puede estar relacionado con que la ramificación que ocurre en Twitter con la propagación de un tweet favorece la acumulación de likes al otorgar mayor visibilidad al mismo y extenderlo al escrutinio de otros usuarios.

Por otra parte, este estudio de caso sugiere estrategias específicas para reforzar la comunicación de la ciencia: ligar la información científica con cuestiones sociopolíticas y/o transmitirla de forma emocional. Como es natural, al circunscribir el estudio a los tweets del caso particular de Tyson los resultados obtenidos no son extrapolables a todo el conjunto del universo de estudio de la divulgación de la ciencia, por lo que no se pretende hacer tal generalización sino indicar una tendencia que debe investigarse en mayor profundidad. Siendo razonablemente coherentes los resultados al aplicar la herramienta propuesta al caso particular de un divulgador famoso, cabe preguntarse: ¿el estudio del discurso de otros divulgadores ofrecerá resultados similares? ¿En qué variará si se realiza con perfiles institucionales? ¿Y si se escogen temas de actualidad científica en lugar de usuarios? Conviene, en conclusión, aplicar la herramienta a otros perfiles específicos y también a discusiones descentralizadas que reciban una significativa atención mediática. Otro estímulo para posteriores investigaciones es el potencial de la herramienta para proporcionar valoraciones periódicas comparables y beneficiar, por ejemplo, a las administraciones en la elaboración de planes concretos de divulgación científica o acciones similares.

En suma, debido a las limitaciones de las encuestas de percepción social de la ciencia (Bauer et al., 2012; Pardo, 2001), este trabajo contempla la inclusión de aspectos inexplorados, corroborando que la investigación en redes sociales puede ser un complemento de gran potencialidad para las mismas (Li et al., 2013). En particular, resulta especialmente llamativa la ratificación con datos empíricos del efecto de la comunicación emocional en el discurso científico (Kaiser, 2014).

### Apoyos

Investigación incluida en el proyecto I+D «Excelencia científica, transferencia de conocimiento, factores organizativos, antecedentes individuales, impacto social» (CSO2013-48053-R), financiada con la ayuda BES-2014-069584 del Ministerio de Ciencia, Innovación y Universidades de España (2015-2019).

### Referencias

- Alvarez-Bornstein, B., & Montesi, M. (2019). Who is interacting with researchers on Twitter? A survey in the field of Information Science. *JLIS, 10*(2), 87-106. <https://doi.org/10.4403/jlis.it-12530>
- Arrabal, G., & De-Aguilera, M. (2016). Comunicar en 140 caracteres. Cómo usan Twitter los comunicadores en España. [Communicating in 140 characters. How journalists in Spain use Twitter]. *Comunicar, 46*, 9-17. <https://doi.org/10.3916/C46-2016-01>
- Bauer, M.W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science, 16*(1), 79-95. <https://doi.org/10.1177/0963662506071287>
- Becker, B.F., Larson, H.J., Bonhoeffer, J., van Mulligen, E.M., Kors, J.A., & Sturkenboom, M.C. (2016). Evaluation of a multinational, multilingual vaccine debate on Twitter. *Vaccine, 34*(50), 6166-6171. <https://doi.org/10.1016/j.vaccine.2016.11.007>
- Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022. <https://bit.ly/2wQLaGj>
- Brossard, D., & Scheufele, D.A. (2013). Science, new media, and the public. *Science, 339*(6115), 40-41. <https://doi.org/10.1126/science.1232329>
- Büchi, M. (2017). Microblogging as an extension of science reporting. *Public Understanding of Science, 26*(8), 953-968. <https://doi.org/10.1177/0963662516657794>
- Dann, S. (2010). Twitter content classification. *First Monday, 12*(12), 15-15. <https://doi.org/10.5210/fm.v15i12.2745>
- Davis, R.C. (1958). *The public impact of science in the mass media*. Institute for Social Research, University of Michigan. <https://stanford.io/2w9teGk>
- Dehkharghani, R., Mercan, A., & Saygin, Y. (2014). Sentimental causal rule discovery from Twitter. *Expert Systems with Applications, 41*(10), 4950-4958. <https://doi.org/10.1016/j.eswa.2014.02.024>
- European Commission (Ed.) (2008). *Public engagement in science*. Publications Office of the European Union. <https://bit.ly/2uB98Vg>
- Kahle, K., Sharon, A.J., & Baram-Tsabari, A. (2016). Footprints of fascination: Digital traces of public engagement with particle physics on CERN's social media platforms. *PLoS One, 11*(5), e0156409. <https://doi.org/10.1371/journal.pone.0156409>
- Kaiser, D., Durant, J., Levenson, T., Wiehe, B., & Linett, P. (2014). *The evolving culture of science engagement: An exploratory initiative*. MIT & Culture Kettle. <https://bit.ly/2VVjy5hG>
- Kapoor, K.K., Tamilmani, K., Rana, N.P., Patil, P., Dwivedi, Y.K., & Nerur, S. (2018). Advances in social media research: Past,

- present and future. *Information Systems Frontiers*, 20(3), 531-558. <https://doi.org/10.1007/s10796-017-9810-y>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In Rappa, M., & Jones, P. (Eds.), *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600). ACM. <https://doi.org/10.1145/1772690.1772751>
- Li, R., Crowe, J., Leifer, D., Zou, L., & Schoof, J. (2019). Beyond big data: Social media challenges and opportunities for understanding social perception of energy. *Energy Research & Social Science*, 56, 101217. <https://doi.org/10.1016/j.erss.2019.101217>
- López-Pérez, L., & Olvera-Lobo, M.D. (2019). Participación digital del público en la ciencia de excelencia española: Análisis de los proyectos financiados por el European Research Council. *El Profesional de la Información*, 28, 1-10. <https://doi.org/10.3145/epi.2019.ene.06>
- Matthes, J., & Kohring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, 58(2), 258-279. <https://doi.org/10.1111/j.1460-2466.2008.00384.x>
- Mohammadi, E., Thelwall, M., Kwasny, M., & Holmes, K.L. (2018). Academic information on Twitter: A user survey. *PLoS One*, 13(5), e0197265. <https://doi.org/10.1371/journal.pone.0197265>
- Moreno-Castro, C., Corell-Doménech, M., & Camano-Puig, R. (2019). Which has more influence on perception of pseudo-therapies: The media's information, friends or acquaintances opinion. *Communication & Society*, 32, 35-49. <https://doi.org/10.15581/003.32.3.35-48>
- Murphy, J., Hill, C., & Dean, E. (2013). *Social media, sociality, and survey research*. John Wiley & Sons. <https://doi.org/10.1002/9781118751534.ch1>
- Murphy, J., Link, M.W., Childs, J.H., Tesfaye, C.L., Dean, E., ... Harwood, P. (2014). Social media in public opinion research: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly*, 78(4), 788-794. <https://doi.org/10.1093/poq/nfu053>
- Myers, S.A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? the structure of the twitter follow graph. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 493-498). ACM. <https://doi.org/10.1145/2567948.2576939>
- Naaman, M., Boase, J., & Lai, C.H. (2010). Is it really about me? message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work* (pp. 189-192). ACM. <https://doi.org/10.1145/1718918.1718953>
- Narr, S., Luca, E.W.D., & Albayrak, S. (2011). Extracting semantic annotations from twitter. In *Proceedings of the fourth workshop on Exploiting semantic annotations in information retrieval* (pp. 15-16). ACM. <https://doi.org/10.1145/2064713.2064723>
- Nisbet, M.C., & Scheufele, D.A. (2009). What's next for science communication? Promising directions and lingering distractions. *American Journal of Botany*, 96(10), 1767-1778. <https://doi.org/10.3732/ajb.0900041>
- Pardo, R. (2001). La cultura científico-tecnológica de las sociedades de la modernidad tardía. *Treballs de la Societat Catalana de Biologia*, 51, 35-63. <https://bit.ly/2T0n8B5>
- Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate Change on Twitter: Topics, communities and conversations about the 2013 IPCC working group 1 report. *PLoS One*, 9(4). <https://doi.org/10.1371/journal.pone.0094785>
- Pérez-Rodríguez, A.V., González-Pedraz, C., & Berrocal, J.L.A. (2018). Twitter como herramienta de comunicación científica en España. Principales agentes y redes de comunicación. *Communication Papers*, 7(13), 95-95. [https://doi.org/10.33115/udg\\_bib/cp.v7i13.21986](https://doi.org/10.33115/udg_bib/cp.v7i13.21986)
- Santoveña, S., & Bernal, C. (2019). Explorando la influencia del docente: Participación social en Twitter y percepción Académica. [Exploring the influence of the teacher: Social participation on Twitter and academic perception]. *Comunicar*, 58, 75-84. <https://doi.org/10.3916/C58-2019-07>
- ScienceFlows (Ed.) (2019). ScienceFlows. <https://bit.ly/2wGZ8dB>
- Shan, L., Áine Regan, Brún, A.D., Barnett, J., van der Sanden, M.C.A., ... Áine McConnon (2014). Food crisis coverage by social and traditional media: A case study of the 2008 Irish dioxin crisis. *Public Understanding of Science*, 23(8), 911-928. <https://doi.org/10.1177/0963662512472315>
- Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *The Journal of Open Source Software*, 1(3), 37-37. <https://doi.org/10.21105/joss.00037>
- Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media: Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248. <https://doi.org/10.2753/mis0742-1222290408>
- Twitter (Ed.) (2019). Application programming interface. <https://developer.twitter.com>
- Uren, V., & Dadzie, A.S. (2015). Public science communication on Twitter: A visual analytic approach. *Aslib Journal of Information Management*, 67(3), 337-355. <https://doi.org/10.1108/ajim-10-2014-0137>
- Veltri, G. (2013). Microblogging and nanotweets: Nanotechnology on Twitter. *Public Understanding of Science*, 22(7), 832-849. <https://doi.org/10.1177/0963662512463510>
- Veltri, G.A., & Atanasova, D. (2017). Climate change on Twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science*, 26(6), 721-737. <https://doi.org/10.1177/0963662515613702>
- Wilkinson, D., & Thelwall, M. (2012). Trending Twitter topics in English: An international comparison. *Journal of the American Society for Information Science and Technology*, 63(8), 1631-1646. <https://doi.org/10.1002/asi.22713>
- Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., ... Li, X. (2011). Comparing Twitter and traditional media using topic models. In *Lecture Notes in Computer Science: Vol 6611. Advances in Information Retrieval* (pp. 338-349). Springer. [https://doi.org/10.1007/978-3-642-20161-5\\_34](https://doi.org/10.1007/978-3-642-20161-5_34)