




<https://hatemedia.es>

Proyecto PID2020-114584GB-I00,  
ejecutado con el apoyo de:




# Discurso de odio y medios digitales en España. Retos de su detección en la web

Elias Said-Hung, Max Römer  
Pieretti y Julio Montero

The background features several large, overlapping, abstract shapes in shades of purple and blue. One large shape is at the top left, a smaller one is to its right, and another is partially visible at the bottom right corner.

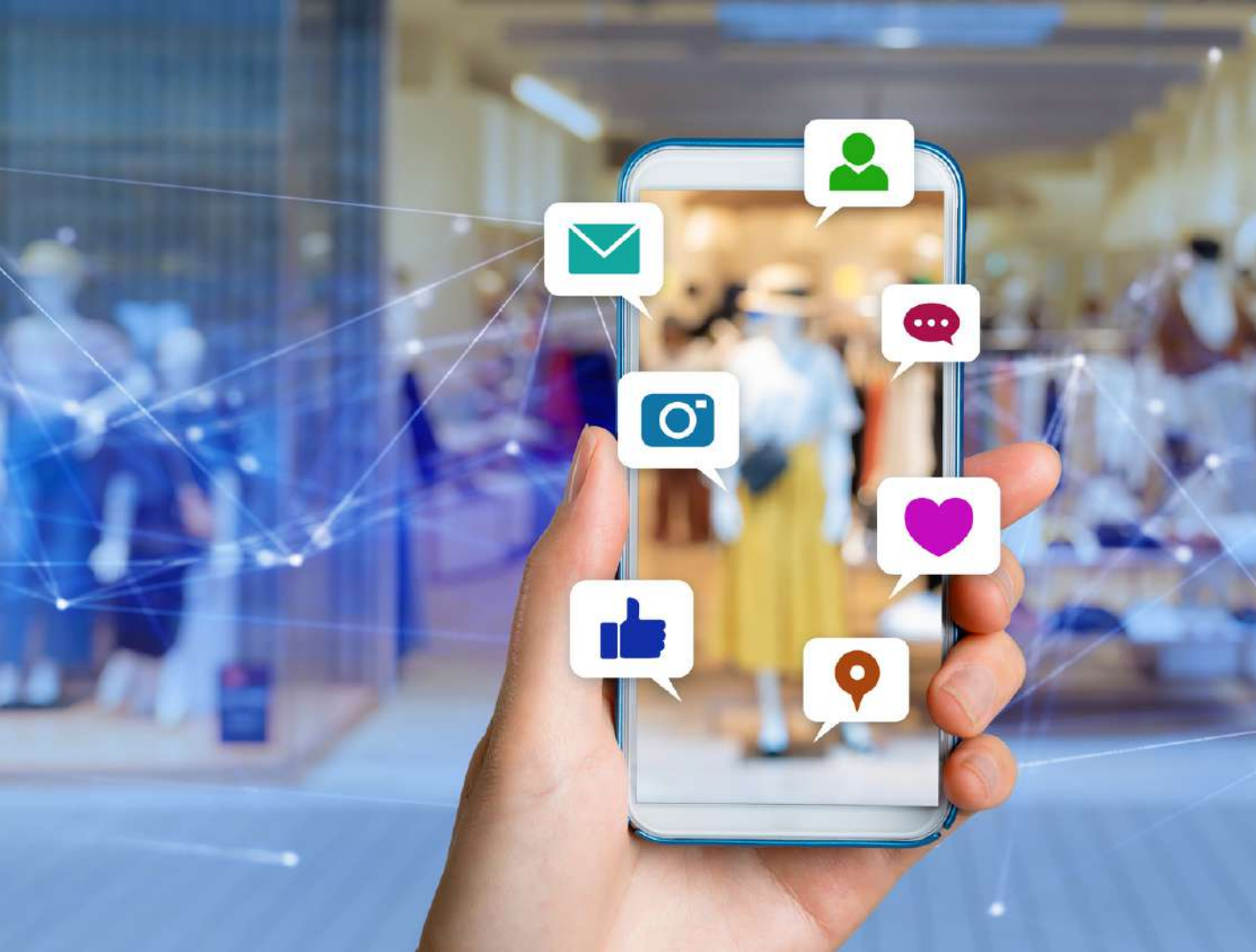
**El estudio de las expresiones de odio, se ha convertido en un tema de creciente interés académico y a nivel de las instituciones públicas...**



# Fase inicial de análisis de las expresiones de odio

Muchos de estos trabajos publicados  
atienden una mirada específica del  
fenómeno





<https://hatemedia.es>

Institución ejecutora:

**unir** LA UNIVERSIDAD  
EN INTERNET

Proyecto PID2020-114584GB-I00,  
ejecutado con el apoyo de:



Instituciones colaboradoras:

 Universidad  
Rey Juan Carlos

 UNIVERSIDAD  
DE BURGOS

 Universidad  
de Navarra

Universida de Vigo

 UNIVERSIDAD DE  
MURCIA

 Universidad  
Camilo José Cela

# Objetivos del proyecto

## OG1

Analizar cómo se difunden las expresiones de odio en los entornos digitales asociados a los medios informativos profesionales en España

## OG2

Favorecer la detección y monitoreo de este tipo de expresiones en los entornos digitales asociados a los medios informativos profesionales en España

1.100.611 mensajes analizados y etiquetados manualmente de usuarios de Twitter y Web de los principales medios informativos digitales en España (1,6% odio detectado, mayoritariamente con una intensidad centrada más en la promoción de la violencia verbal y otros tipos de intensidades menores)



# Algunos retos asociados al estudio del odio, identificados en Hatemedia

## Librerías de palabras

Ausencia de librerías públicas (actualizables) de palabras de odio robustas (en cuanto a número de elementos) de léxicos y expresiones apegadas a la semántica española

## Modelos binarios

Predominio de modelos binarios de clasificación, de lo que es o no odio, que si bien ayudan a poder establecer una primera aproximación del fenómeno (desde la misoginia, migración, entre otros), no facilita la detección de lo que se entiende como expresiones de odio y el discurso ofensivo o agresivo (que no llega a ser violatorio de derechos o punible como delito)

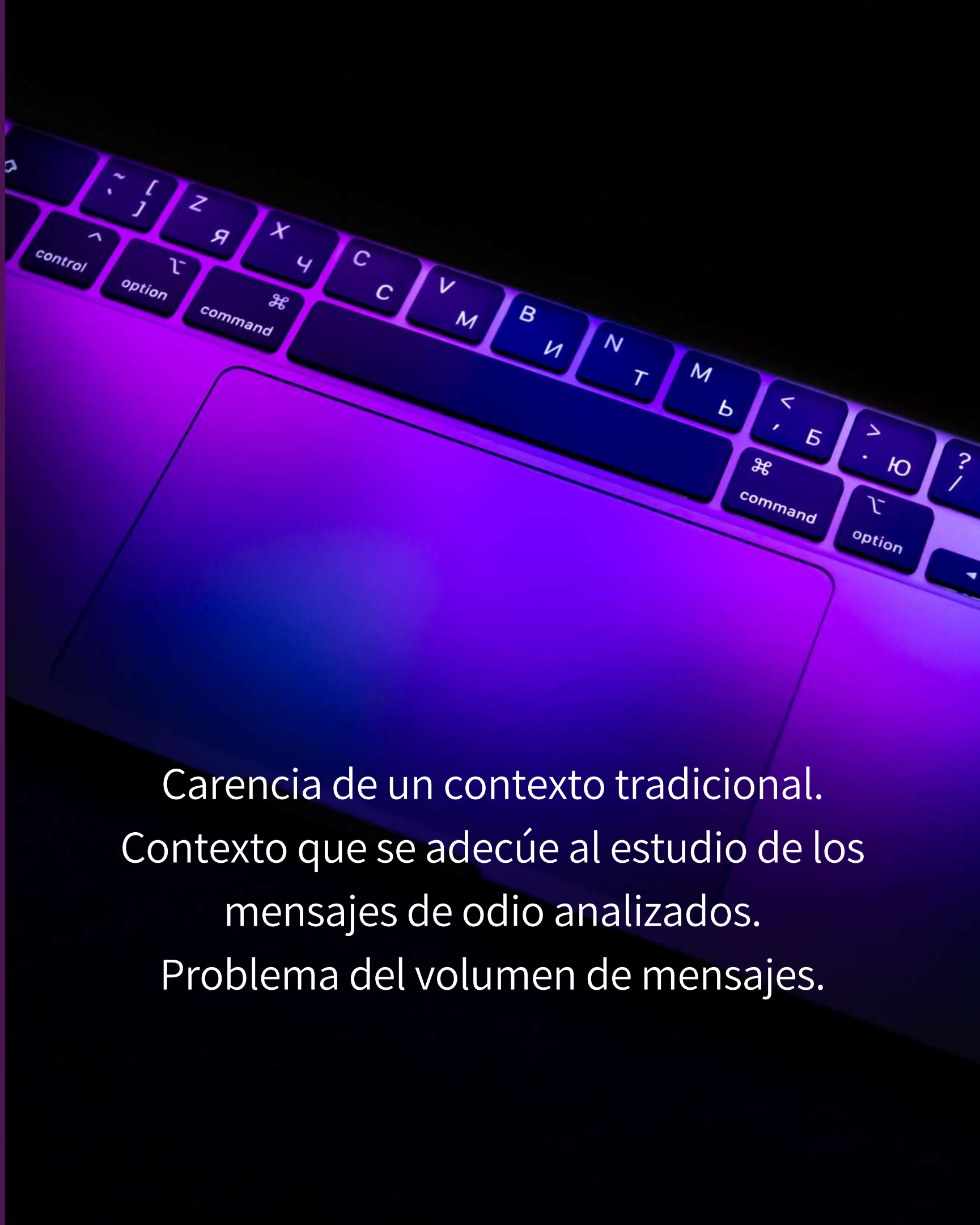
## Criterios comunes de monitoreo

Falta de criterios comunes que ayuden al monitoreo de este fenómeno de forma global, es decir, que analice la presencia de este tipo de expresiones en medios informativos digitales.

## Técnicas de clasificación

Identificación de técnicas más adecuadas de clasificación de expresiones de odio (con el mayor nivel de accuracy), desde un enfoque no binominal, en nuestro caso en español, haciendo uso de métodos tradicionales (SVM, Naive Bayes, Random Forest o Logistic Regression), como nuevos (Deep Learning Methods)

# Semiótica del odio



Carencia de un contexto tradicional.  
Contexto que se adecúe al estudio de los mensajes de odio analizados.  
Problema del volumen de mensajes.



# Desde Hatemedia hemos.....

recurrido a las estructuras generativas de Greimas  
como solución para nuestro estudio





# Las estructuras generativas

Son universales e independientes de un código específico.

Describen una competencia taxonómica y sintagmática.

En el nivel más profundo, la semántica fundamental contiene categorías semánticas que adoptan formas elementales de significación; y, la sintaxis fundamental contiene las relaciones y transformaciones que constituyen esas estructuras.

Por otro lado, en el nivel superficial la sintaxis narrativa analiza los sintagmas narrativos que son concebidos como categorías antropomórficas de relaciones entre los actantes.

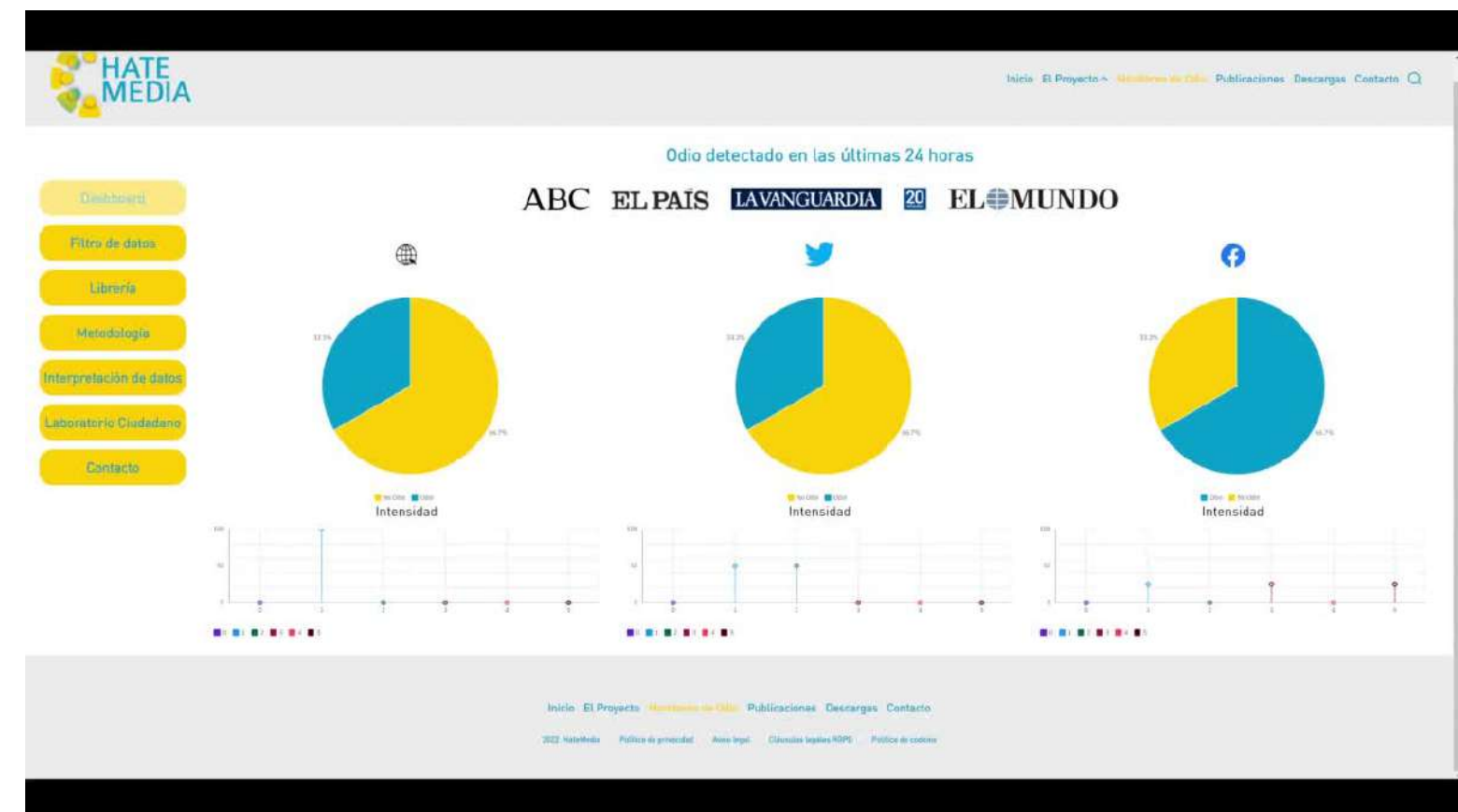
# De esa forma y, siguiendo con los elementos greimasianos

Determinamos cinco niveles en el que los actantes adoptan formas narrativas de odio

Niveles de intensidad de odio	
Nivel 5	Expresiones que llaman a la violencia física contra grupos sociales o personas
Nivel 4	Expresiones intimidatorias
Nivel 3	Violencia verbal
Nivel 2	Atribuciones de actos negativos contra grupos sociales
Nivel 1	Atribuciones que invitan al odio hacia grupos sociales
Nivel 0	Atribuciones de referencia hacia grupos sociales sin intención de estigmatizar



# Aportes de Hatemedia al estudio del odio en medios informativos digitales en España - A modo de conclusión





Gracias...



<https://hatemedia.es>

Proyecto PID2020-114584GB-I00,  
ejecutado con el apoyo de:

