**Pre-Conference ECREA 2022**

News Media, disinformation, and hate speech's promotion

Dra. Tamara Antona Jimeno, Dr. Elías Said-Hung y Julio Montero Díaz


# Hate Speech's taxonomy in Spanish professional news media

unir LA UNIVERSIDAD EN INTERNET

HATE MEDIA

Universidad Internacional de La Rioja

# Presentation

*Pre-Conference ECREA 2022. News Media, disinformation, and hate speech's promotion*

- Research Project: *Taxonomía, presencia e intensidad de las expresiones de odio en entornos digitales vinculados a los medios informativos profesionales españoles* – Hatemedia

PID2020-114584GB-I00, Funding by Agencia Estatal de Investigación - Ministerio de Ciencia e Innovación

# Objetives

- Presentation of preliminary results
- Map of hate speech promoted by users and digital spaces linked to the media in Spain
- Helping the detection and monitoring of hate speech.
- Tangible result: an algorithm that is under construction (we will see DEMO) and that serves to monitor, in real time, hate speech from institutional environments associated with the media.

# Methodology

- It takes data extracted from Twitter users and institutional portals linked to the main media outlets.

- Semiotic discursive analysis of the contents published both by the media and their followers during a full year.

- The contents are classified and analyzed with hate speech according to the proposed levels of intensity

# Methodology: Media

- The professional digital media analyzed are:

EL PAÍS

EL MUNDO.es

20 minutos

Antena 3

elDiario.es
Periodismo a pesar de todo

- The study analyzes the messages published by the professional digital media considered on their institutional portals and on Twitter and Facebook.

# Methodology: Why Twitter?

- Twitter is the social network where the influence of micro-stories has grown the most.

- Its users tend to prefer breaking news and the monitoring of current affairs.

- It is the main social network for political actors in terms of contact with citizens, especially during electoral processes.

# Methodology: Why Twitter?

- Its strength is growing with respect to the dissemination of content generated by the media themselves and also for the exercise of journalistic work through immediate contact (either with the facts, or with other users).

- It is the network that inspires the greatest confidence in Spanish media managers to access reliable data in digital environments.

- It is one of the social networks with the largest presence of users linked to traditional and digital media in Europe.

# Methodology: What do we analyze?

In the main digital media:

- Headlines and headlines of published messages associated with news published by these actors.

- Comments published by them in reaction to messages published by other users in their institutional portals and on Twitter.

# Methodology: Data core

- Data were collected directly from users associated with professional digital media for 12 continuous months.

- This collection was performed on a daily basis using the Twitter academic API, following a CRISP-DM methodology.

- As much of the data was obtained from Twitter, it was unstructured (e.g., text and images). Therefore, a preprocessing was first performed to normalize the content linked to each tweet

# Methodology: Collected data

- Data were collected directly from users associated with professional digital media for 12 continuous months.

- This collection was performed on a daily basis using the Twitter academic API, following a CRISP-DM methodology.

- As much of the data was obtained from Twitter, it was unstructured (e.g., text and images). Therefore, a preprocessing was first performed to normalize the content linked to each tweet

# Methodology: Data processing

- Once the messages have been standardized, they have been classified according to the presence or absence of hate speech in them.

- Hate levels are also established: scale from 0 to 5. If it does not fit in any of the intensities, it is not catalogued.

- Types of hate: select as many as you find

- Is there humor? select yes or no

- Modifiers: if we find intensifiers or attenuators

# Methodology: Hate Levels

- **Level 0:** A term is used that could objectively define a trait of a group and is not used in a derogatory sense, but socially has negative connotations.

- **Level 1:** No verbal violence, but a fact is presented in a factual manner in order to stigmatize a particular social group.

- **Level 2:** Attribution of bad intentions, abusive expressions, also attributing to one or several persons clearly negative actions with the intention of spreading a negative image of the social group to which they belong.

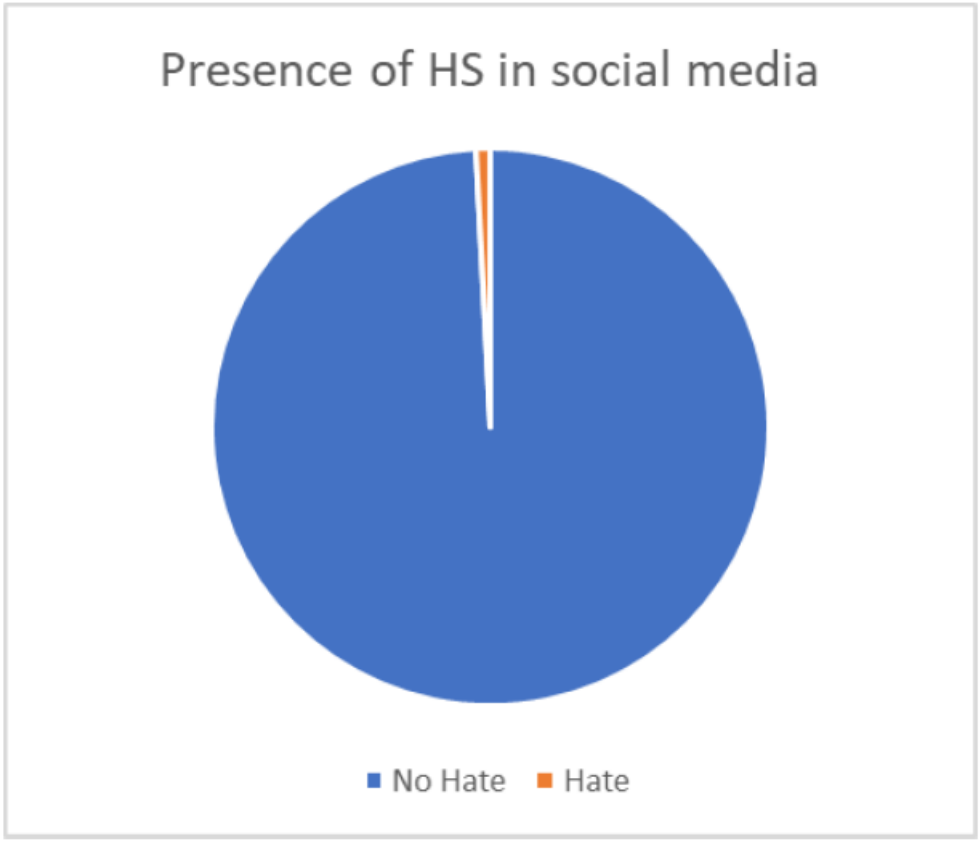
# Methodology: Hate Levels

- **Level 3:** the area of mere verbal violence. This includes all messages intended to insult, offend, belittle or humiliate a person(s) for belonging to a group that the sender hates (usually containing insults).

- **Level 4:** Veiled or implied threats, Expressions of positive emotion in the face of death, aggression or physical harm to someone or something. Intimidating expre**ssions with consequences (real or otherwise) that are not physical.**

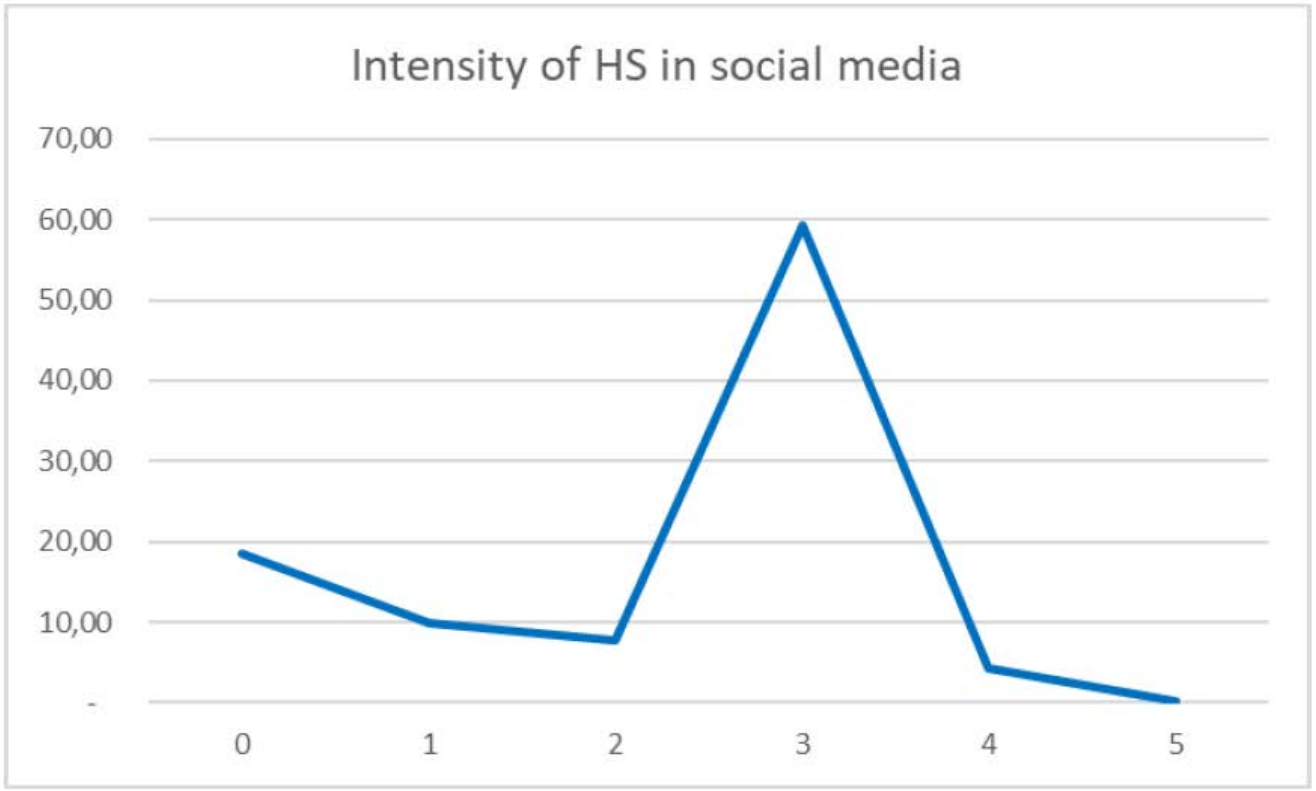
- **Level 5:** Expressions that call for physical violence against or linked to them. Also expressions that express a desire for another to die or come to harm.

# First resoults



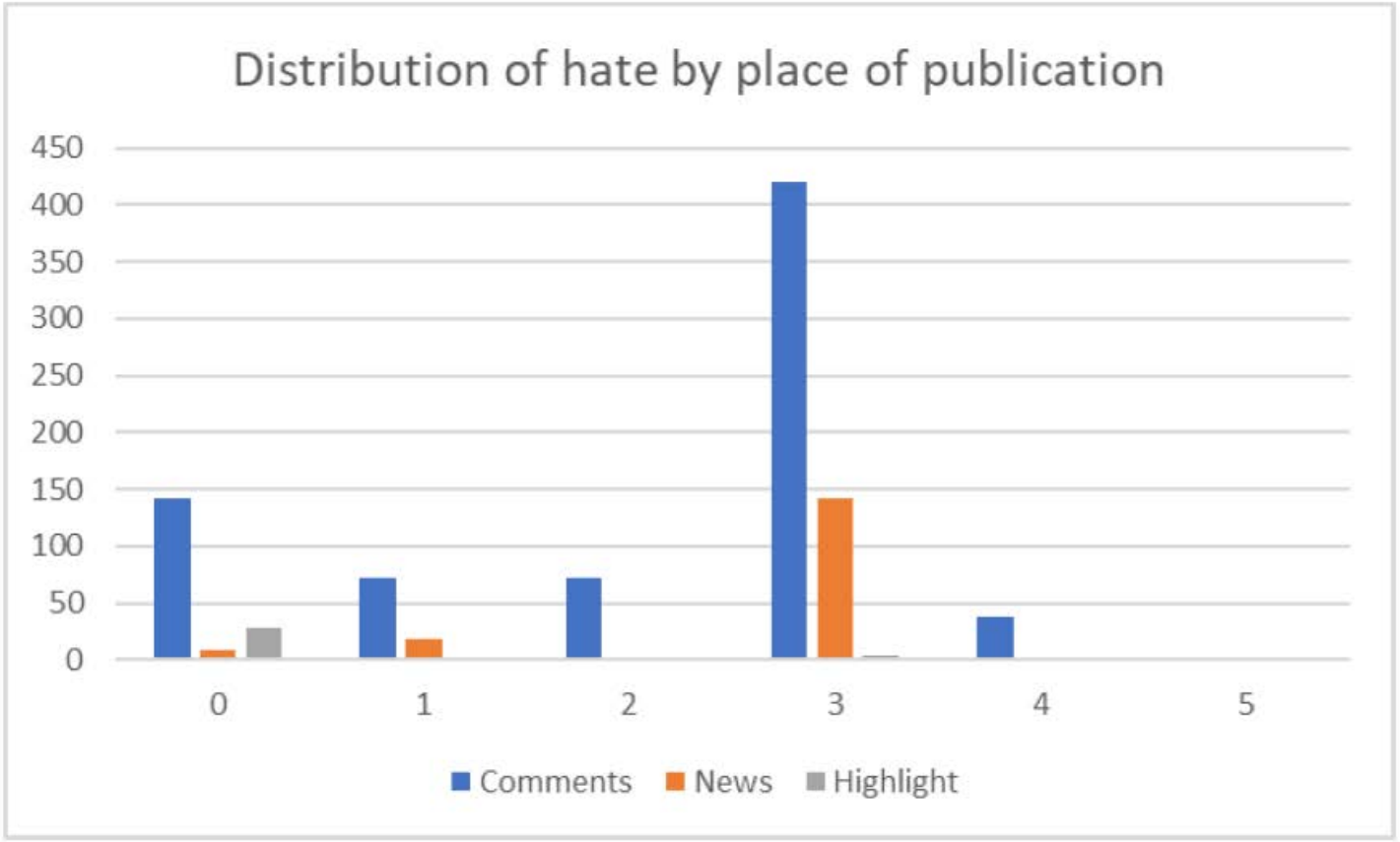

Source: Corpus of 115,190 messages of social media. Own production.

# First resoults



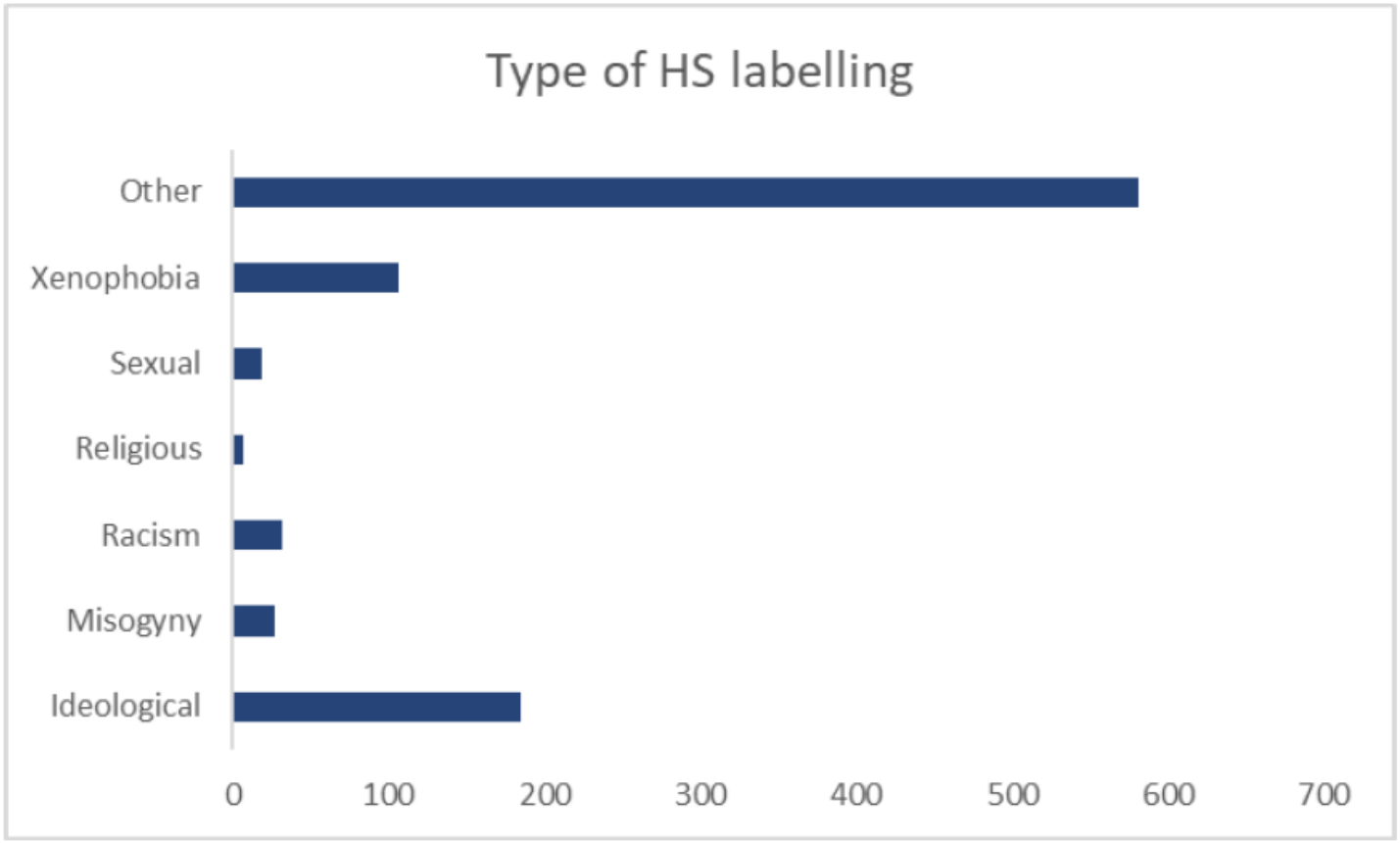

Source: Corpus of 115,190 messages of social media. Own production.

# First resoults



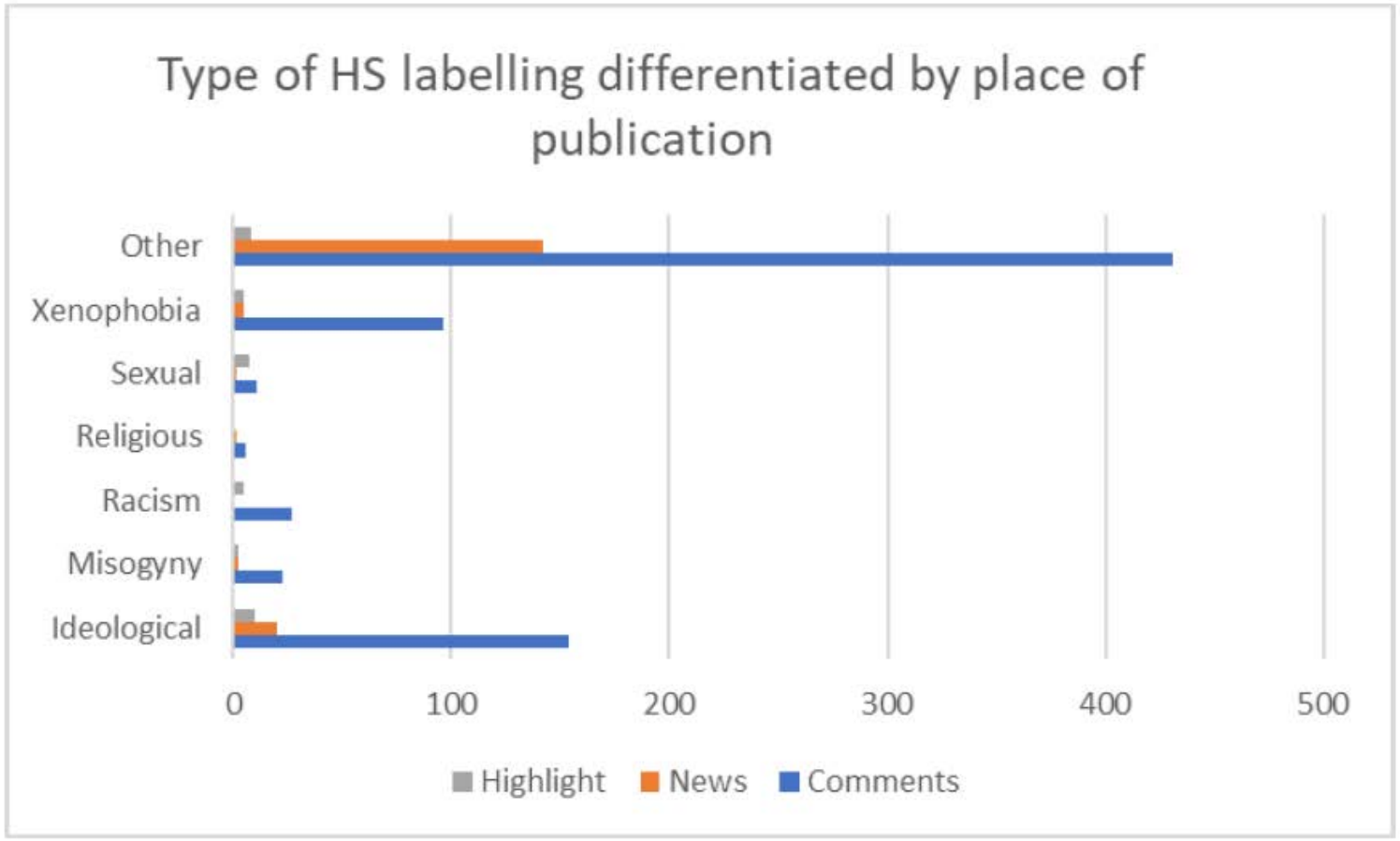

Source: Corpus of 115,190 messages of social media. Own production.

# First resoults




Source: Corpus of 115,190 messages of social media. Own production.

# First resoults



# Reflexions

- The most preferred place for hate speech is in user comments.
- This is good news as it means that, in general terms, hate speech from the media is very residual, which shows a correct democratic functioning.
- Hate speech detected in news and headlines is very low.

# Reflexions

- It is not very present (little = less than 1 percent), sometimes it has a lot of media impact.
- It is not a mainstream phenomenon, however, a broader analysis is needed
- The algorithm will be able to monitor hate in real time.

# Demo Hate Monitor - Hatemedia

https://www.youtube.com/watch?v=h4qyKeCtsgc&list=TLPQMzAwOTIwMjJJ3rbmhRyV1w&index=2


www.unir.net