

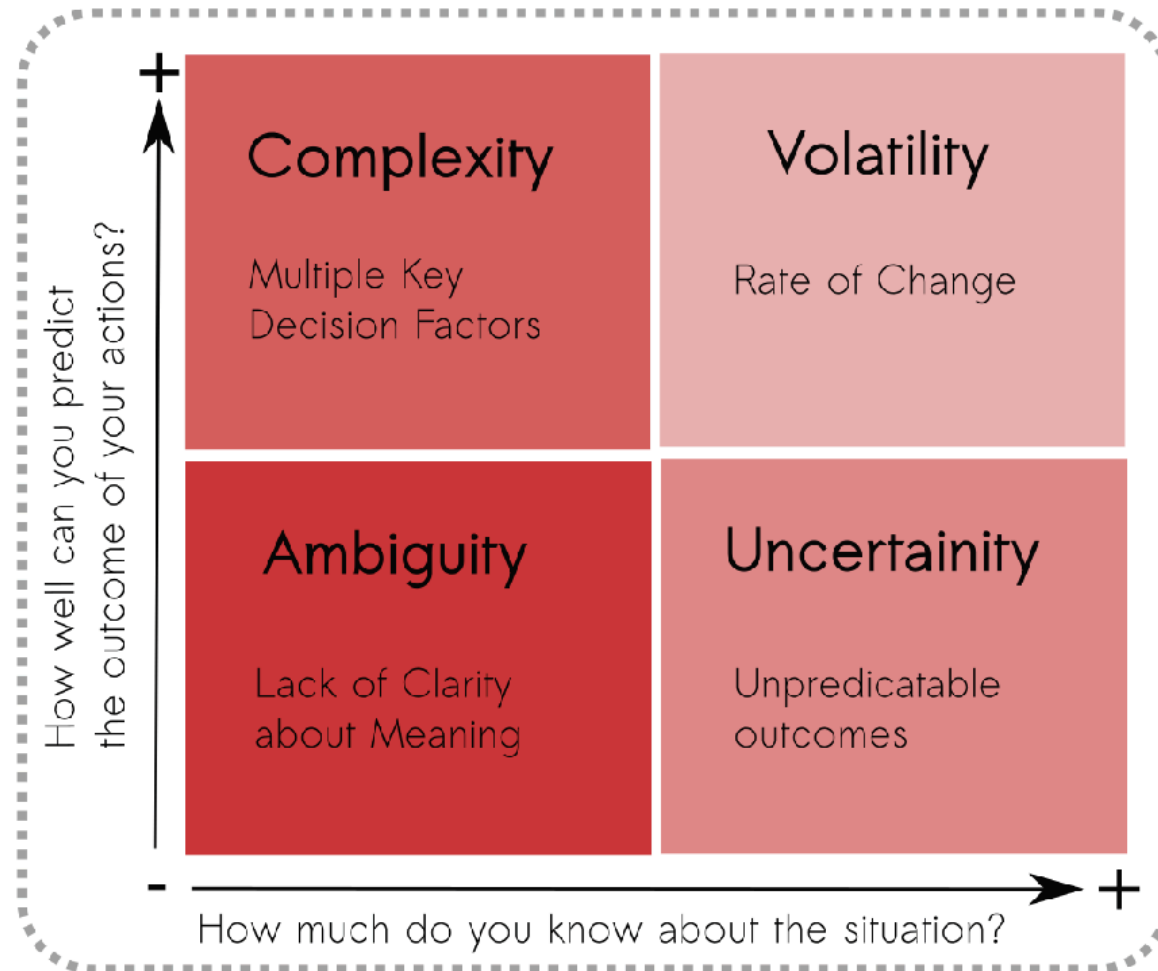
Technology behind hate speech

J. Oriol López-Plana

Hate speech: definition

- Facebook: “We define hate speech as a direct attack on people based on what we call protected characteristics—race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability. We define attack as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation.”

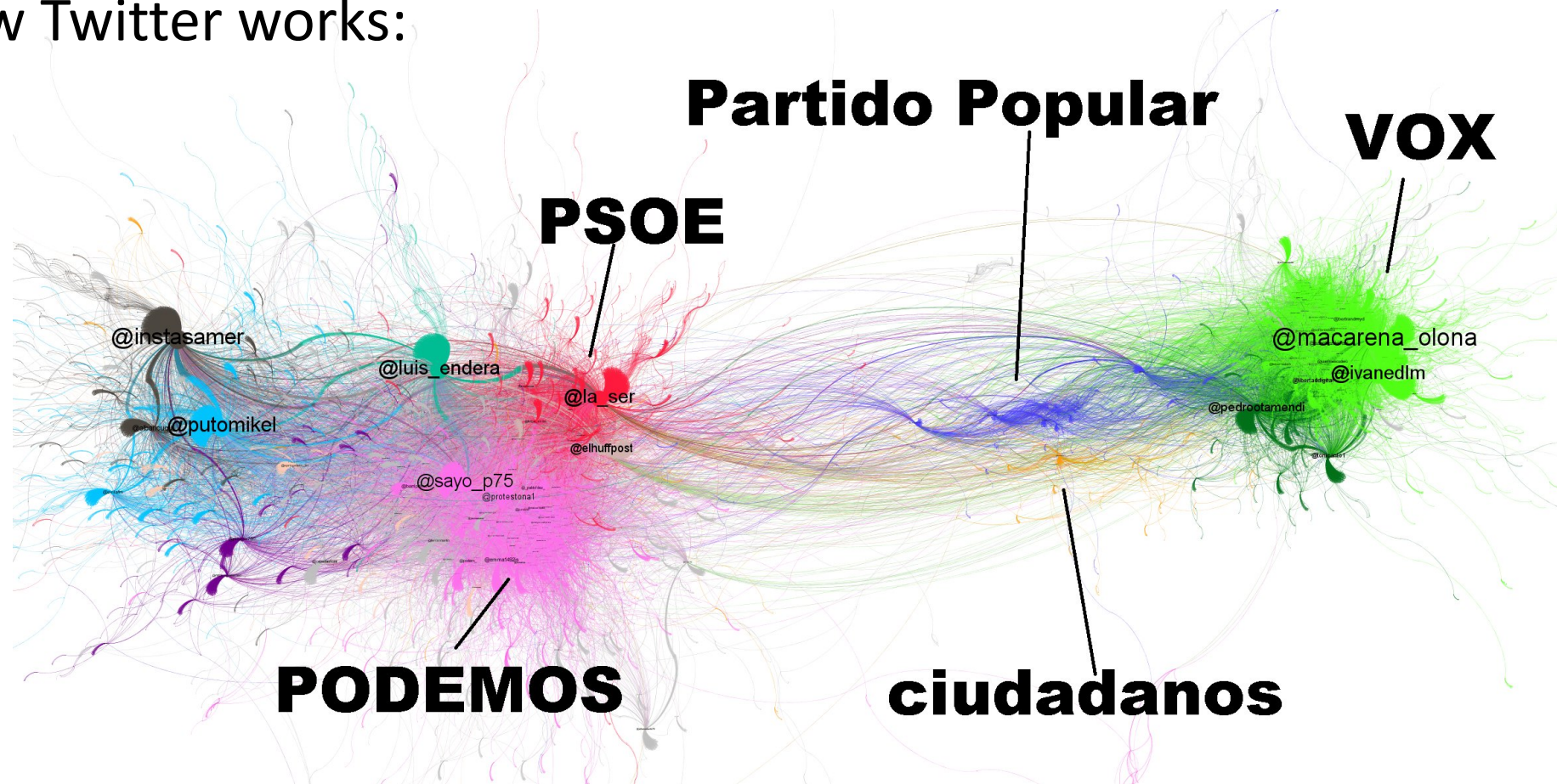
VUCA concept.



Berber (1992)

Hate speech on Twitter.

- How Twitter works:



130k tweets #mociondecensura. Analysis by @barrripdmx

Hate speech on Twitter.

Types of nodes:

- Fake accounts and cyborgs: The user is a human who uses automation to amplify their messaging.
- Bots: Fully automatic accounts
- Sockpuppets: Real accounts that have not been used for a while. They amplify messages.
- Teams: Groups of people that generate previously agreed content.

Hate speech and racist discourses have increased since 2017.



UNITED NATIONS STRATEGY AND PLAN OF ACTION ON HATE SPEECH

Foreword

Around the world, we are seeing a disturbing groundswell of xenophobia, racism and intolerance – including rising anti-Semitism, anti-Muslim hatred and persecution of Christians. Social media and other forms of communication are being exploited as platforms for bigotry. Neo-Nazi and white supremacy movements are on the march. Public discourse is being weaponized for political gain with incendiary rhetoric that stigmatizes and dehumanizes minorities, migrants, refugees, women and any so-called “other”.

This is not an isolated phenomenon or the loud voices of a few people on the fringe of society. Hate is moving into the mainstream – in liberal democracies and authoritarian systems alike. And with each broken norm, the pillars of our common humanity are weakened.

A handwritten signature in black ink, which appears to be 'António Guterres', is written over a light blue background.

United Nations Secretary-General
António Guterres

May 2019

13 UN key commitments

Hate speech and racist discourses have increased since 2017.



English > News and Events > DisplayNews

Share 478K Tweet 407K Share 14.4K

Joint open letter on concerns about the global increase in hate speech

Signed by 26 mandates, see list below

We are alarmed by the recent increase in hateful messages and incitement to discrimination and hatred against migrants, minority groups and various ethnic groups, as well as the defenders of their rights, in numerous countries. Hate speech, both online and offline, has exacerbated societal and racial tensions, inciting attacks with deadly consequences around the world. It has become mainstream in political systems worldwide and threatens democratic values, social stability and peace. Hate-fuelled ideas and advocacy coarsen public discourse and weaken the social fabric of countries.

<https://www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=25036&LangID=E> - Genève, 23.09.2019

Hate speech and racist discourses have increased since 2017.

- In some sensitive places top politicians and local municipalities have taken these discourses as part of their political agenda.

Orbán compares EU to Soviet Union



Kester Eddy in Budapest MARCH 15 2012



Dice la líder del Frente Nacional en Francia, la 'ultra' Marine Le Pen: "Para mí la Unión Europea es como la Unión Soviética"

9:41 · 14 mar. 14 · [Twitter Web Client](#)



Quando un'immagine vale più di mille parole. Sì a [#brexit](#), l'inizio della fine dell'Unione Sovietica Europea!



Jeremy Hunt is using my language, the EU is the new Soviet Union.

[Traducir Tweet](#)

10:53 · 01 oct. 18 · [Twitter for iPhone](#)



TV en DIRECTO | Abascal (Vox) centra ahora su intervención en un discurso euroesceptico y contra el Estado de las autonomías: "A España la están desvalijando a dúo, por la maquinaria de Bruselas y las comunidades separatistas" [#mocióndecensura](#) bit.ly/35gdyz2

Hate speech and racist discourses have increased since 2017: COVID19

United Nations Guidance Note on Addressing and Countering COVID-19 related Hate Speech

11 MAY 2020

RECOMMENDATIONS:

- To united nations departments, agencies, funds and programmes.
- To member states.
- To social media and tech companies.
- To media.
- To civil society and other stakeholders.

The COVID-19 pandemic has amplified existing concerns related to the spread and use of hate speech globally. It has also created new vulnerable groups targeted by hate speech.

Technology for identifying hate speech.

- Hate speech detection and monitoring: database.

HATEBASE

Search

ADVANCED SEARCH
HOW IT WORKS
RECENT SIGHTINGS
VOCABULARY UPDATES

CITIZEN LINGUIST LAB
OPEN TECHNOLOGY PROJECTS

The world's largest structured repository of regionalized, multilingual hate speech

Companies » Nonprofits » Academia » Government » Media »

3,725 TERMS 580,866 SIGHTINGS 97 LANGUAGES 178 COUNTRIES

Technology for identifying hate speech.

- Hate speech detection and monitoring: verification.



FAke News discovery and propagation from big
Data ANalysis and artificial intelliGence Operations

The aim of FANDANGO is to aggregate and verify different typologies of news data, media sources, social media, open data, so as to detect fake news and provide a more efficient and verified communication for all European citizens.

Fandango-project.eu has received funding from the European Union's H2020 EU research and innovation programme.

Technology for identifying hate speech.

- Hate speech detection and monitoring: network analysis.

Klausen et al. (2012): 1) Nodes are coded in a formalised manner and subjected to statistical analysis. 2) Social Network Analysis (SNA) can be used to map communication structures and provide intuitive understanding of different types of communication networks.

Mathew et al. (2019): Understand the dynamics of user interaction that facilitate the spread of hateful content.



Technology for identifying hate speech.

- Hate speech detection and monitoring: network analysis.



[Download](#) [Blog](#) [Wiki](#) [Forum](#) [Support](#) [Bug tracker](#)

[Home](#) [Features](#) [Learn](#) [Develop](#) [Plugins](#) [Services](#) [Consortium](#)

The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

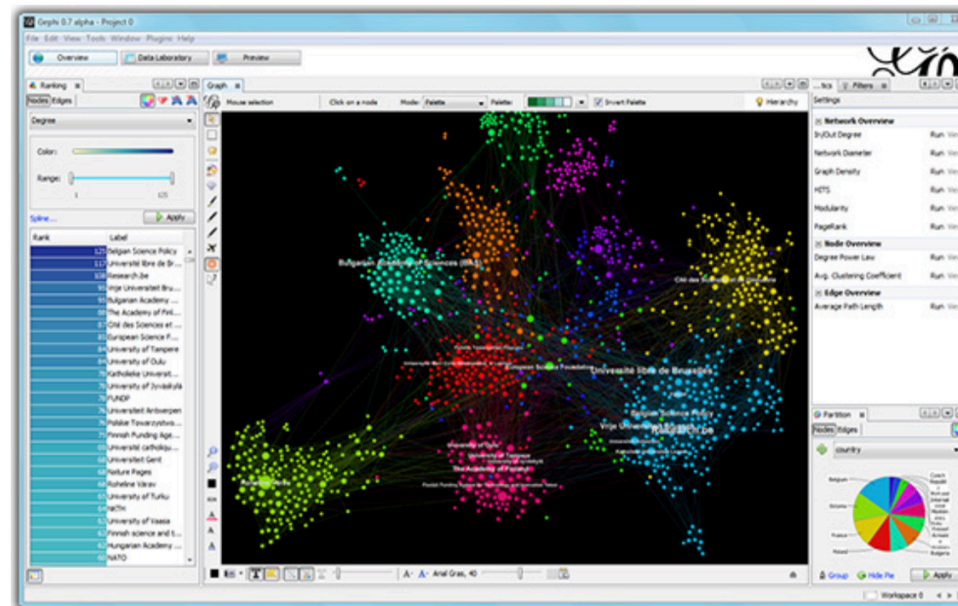
Runs on Windows, Mac OS X and Linux.

[Learn More on Gephi Platform »](#)



[Release Notes](#) | [System Requirements](#)

- ▶ [Features](#)
- ▶ [Screenshots](#)
- ▶ [Quick start](#)
- ▶ [Videos](#)



- ✓ **Exploratory Data Analysis:** intuition-oriented analysis by networks manipulations in real time.
- ✓ **Link Analysis:** revealing the underlying structures of associations between objects.
- ✓ **Social Network Analysis:** easy creation of social data connectors to map community organizations and small-world networks.

Technology for identifying hate speech.

- Communication and visualisation.



The screenshot shows the Twitter profile of 'Odiómetro' (@odiometrobot). The profile header features a dark red banner with the 'ODIÓMETRO' logo, which consists of two hands shaking. Below the banner is a circular profile picture with the same logo. The bio states: 'Rastreamos el odio en Twitter. Te lo mostramos en un resumen diario — y en tiempo real en odiometro.es. Nuestro objetivo: desactivar el odio en Twitter. Únete.' It also includes a website link, the date 'Joined February 2015', and follower statistics: '118 Following' and '4,136 Followers'. A 'Following' button is visible. The navigation sidebar on the left includes Home, Explore, Notifications, Messages, Bookmarks, Lists, Profile, and More. A 'Tweet' button is at the bottom of the sidebar.



Technology for identifying hate speech.

- Communication and visualisation.

simon wiesenthal center
digitalterrorism
2020 +**hate**

REPORT CARD REPORT ACTIVITY STRATEGIES 2.0 TERRORISM GEOGRAPHIC

Strategies 2.0
Web 2.0

- Blogs-North America
- Blogs-International
- Apps
- Networking
- Video on Demand
- Miscellaneous
- Vk.com
- Alt-Tech
- Twitter
- Gab
- Instagram

Targeting the Enemy

- Religion
- Re-Writing History
- Conspiracy
- Images
- Miscellaneous
- Social Networking
- Covid-19 / Coronavirus

Recruitment

- U.S./Foreign
- Gender/Education
- Discussion Forums
- Networking
- Blogs/Publications
- Messaging/Apps

Terrorism
Strategies

- Twitter
- Recruiting
- Propaganda/Threats
- Images
- Broadcast/Video
- Networking
- Islamic State
- Instagram

Instructional

- Explosives
- Guerilla Tactics
- Manuals/Other
- Cyberwars

Support

- Social Networking
- Fatwas
- Recruiting
- Women/Children
- Fundraising
- Blogs
- Media/Publications
- Online Magazines
- Hashtags

Who's Who

- United States
- International

Geographic
U.S./Canada

- Web 2.0
- Social Networking
- Nazi / Supremacy
- Klan / Skinhead
- Other U.S. Groups
- Miscellaneous
- Canada Groups
- Canada Other
- Twitter

Europe

- Western Europe
- Northern Europe
- Central Europe
- Eastern Europe
- Social Networking
- Vk.com

Latin / S. America

- Youth/Music
- Aryan
- Other
- Social Networking

Asia/Africa/Australia

- Asia
- Australia/New Zealand
- Africa/Middle East
- Social Networking

print search site map

simon wiesenthal center snider social action institute
To report a hate or terror website, blog, newsgroup or video please email: ireport@wiesenthal.com

Twitter alternatives?

- Mastodon (<https://mastodon.social>)

Mastodon isn't a single website like Twitter or Facebook, it's a network of thousands of communities **operated by different organizations** and individuals that provide a seamless social media experience. Without an incentive to sell you things, Mastodon allows you to consume content you enjoy uninterrupted. Your feed is chronological, ad-free and **non-algorithmic**—you decide who you want to see!



Twitter alternatives?

- Retroshare (<https://retroshare.cc>)

Retroshare establish encrypted connections between you and your friends to create a **network of computers**, and provides various distributed services on top of it: forums, channels, chat, mail...

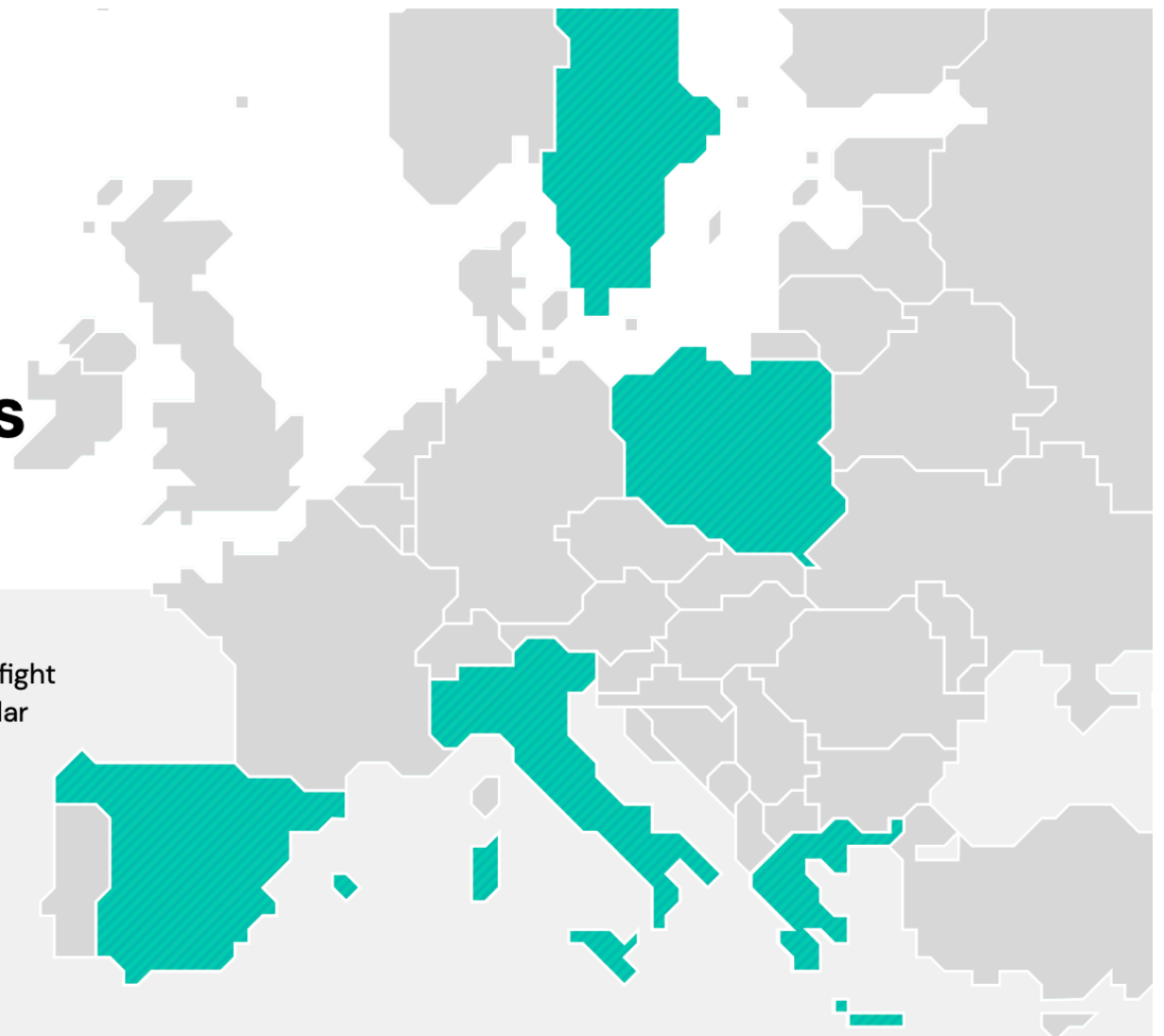
Retroshare is fully decentralized, and designed to provide maximum security and anonymity to its users beyond direct friends. Retroshare is entirely free and open-source software.



SO-CLOSE: H2020 project

From so far to so close

Enhancing Social Cohesion through Sharing the Cultural Heritage of Forced Migrations



SO-CLOSE is a three-year project created to contribute to social cohesion and fight refugee marginalization or exclusion by facilitating the encounters between similar life stories, through the mediation of innovative digital and artistic tools.

Based on theories of cultural heritage-making, SO-CLOSE will improve social cohesion and promote mutual understanding between refugees and their local communities.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 870939

References

- Barber, H-F. (1992). Developing Strategic Leadership: The US Army War College Experience. *Journal of Management Development* 11, no. 6: 4-12.
- MacAvaney, S., Yao, H-R., Yang, E., Russell, K., Goharian, N., & Frieder, O. (2019). Hate speech detection: Challenges and solutions. PLoS ONE 14(8): e0221152. <https://doi.org/10.1371/journal.pone.0221152>
- Yang, K-C., Varol, O., Hui, P-M., & Menczer, F. (2020). Scalable and Generalizable Social Bot Detection through Data Selection. Vol 34 No 01: AAAI-20 Technical Tracks 1. DOI: <https://doi.org/10.1609/aaai.v34i01.5460>
- Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.
- Klausen, J., Barbieri, E., Reichlin-Melnick, A., & Zelin, A. (2012). The YouTube Jihadists: A Social Network Analysis of Al-Muhajiroun's Propaganda Campaign. *Perspectives on Terrorism*, 6(1), 36-53. Retrieved October 22, 2020, from <http://www.jstor.org/stable/26298554>
- Mathew, B., Dutt, R., Goyal, P., & Mukherjee, A. (2019). Spread of Hate Speech in Online Social Media. In Proceedings of the 10th ACM Conference on Web Science (WebSci '19). Association for Computing Machinery, New York, NY, USA, 173–182. DOI: <https://doi.org/10.1145/3292522.3326034>



TransMedia Catalonia is a research group funded by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, reference code 2017SGR113.

The project REBUILD has received funding from the European Union's Horizon 2020 Research and Innovation Programme, grant agreement No 822215.

Texts, marks, logos, names, graphics, images, photographs, illustrations, artwork, audio clips, video clips, and software copyrighted by their respective owners are used on these slides for personal, educational and non-commercial purposes only. Use of any copyrighted material is not authorized without the written consent of the copyright holder. Every effort has been made to respect the copyrights of other parties. If you believe that your copyright has been misused, please direct your correspondence to: maria.jimenez.andres@uab.cat stating your position and we shall endeavour to correct any misuse as early as possible.

This document and its contents reflect the views only of the authors. TransMedia Catalonia and the European Union H2020 Research and Innovation Programme cannot be held responsible for any use which may be made of the information contained therein.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.