Empowering Cyber Intelligence with Natural Language Processing

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

BSIDES Cybersecurity conference
Bogotá, Colombia, April 23th ,2023

Andrés Zapata Rozo

Cybersecurity engineer, Telefónica
Professional in Applied Mathematics and Computer Science, University of Rosario

andres.zapatarozo@telefonica.com
andresf.zapata@urosario.edu.co

Alejandra Campo Archbold

Cybersecurity analyst, Telefónica
Professional in Applied Mathematics and Computer Science (C), University of Rosario

alejandra.campoarchbold@telefonica.com
alejandra.campo@urosario.edu.co

Daniel Díaz, PhD

Principal Professor,
School of Engineering, Science and Technology, University of Rosario

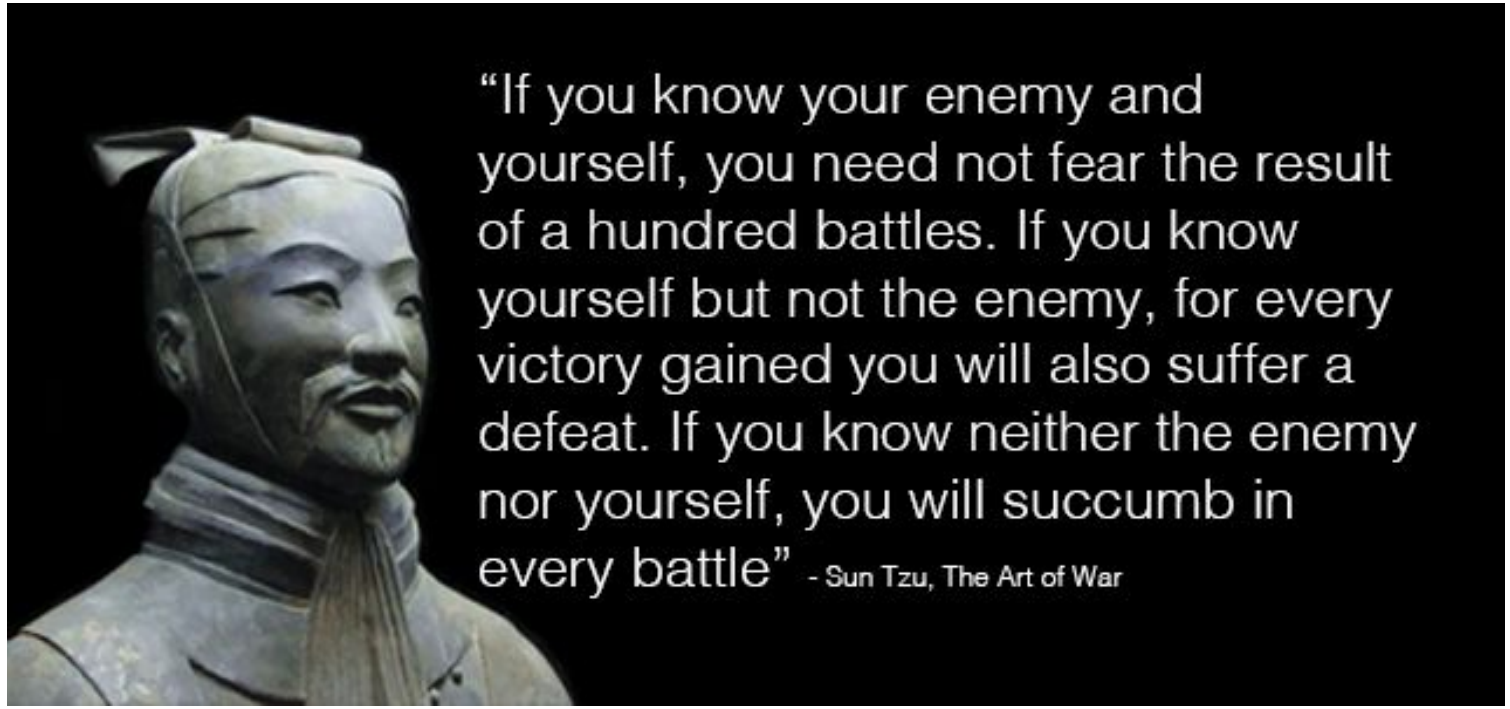
danielo.diaz@urosario.edu.co

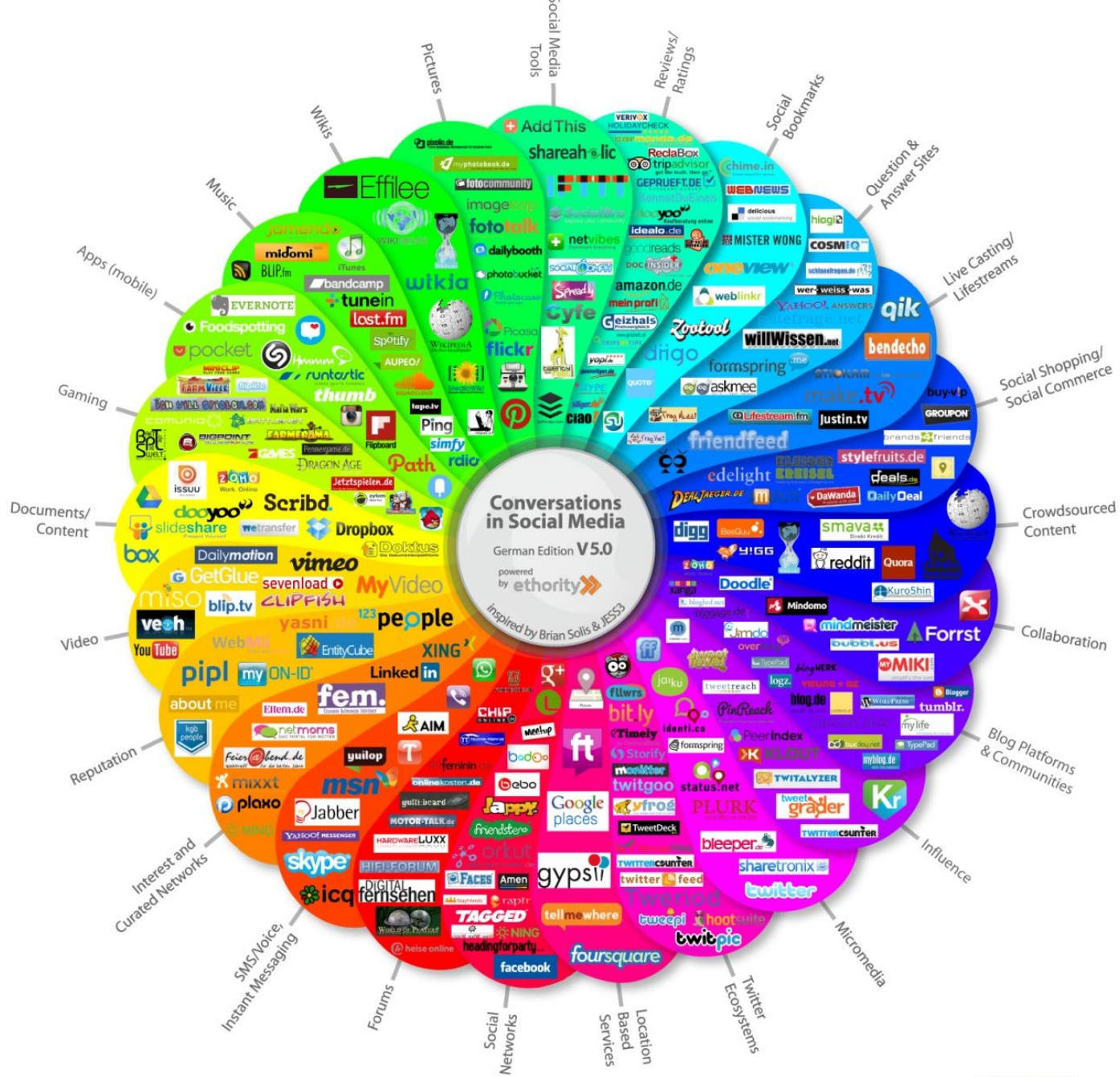
Agenda

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

1. OSINT
2. Hostile Social Manipulation and Social Warfare
3. Artificial intelligence and NLP
4. Preventing cyber crimes in Colombia, Ecuador and USA

An inspiring phrase…





OSINT (Open Source Intelligence)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Through OSINT it is possible to:

- Identify **potential adversaries** according to their level of hostility toward a topic of interest, an individual or an organization.
- Identify the **motivation** and the **means** that an adversary can have to affect a valuable asset of a company.
- Prevent the realization of (cyber) terrorist attacks, coordinated activities with high impact on a **assets** of a Nation.

Know the enemy
(Means, Motivation)

Know yourself
(Opportunities/vulnerabilities)

HSM and Information Warfare

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Hostile Social Manipulation (HSM)

The purposeful, systematic generation and dissemination of information to produce **harmful** social, political, and economic outcomes in a target area by affecting beliefs, attitudes, and behavior

Techniques

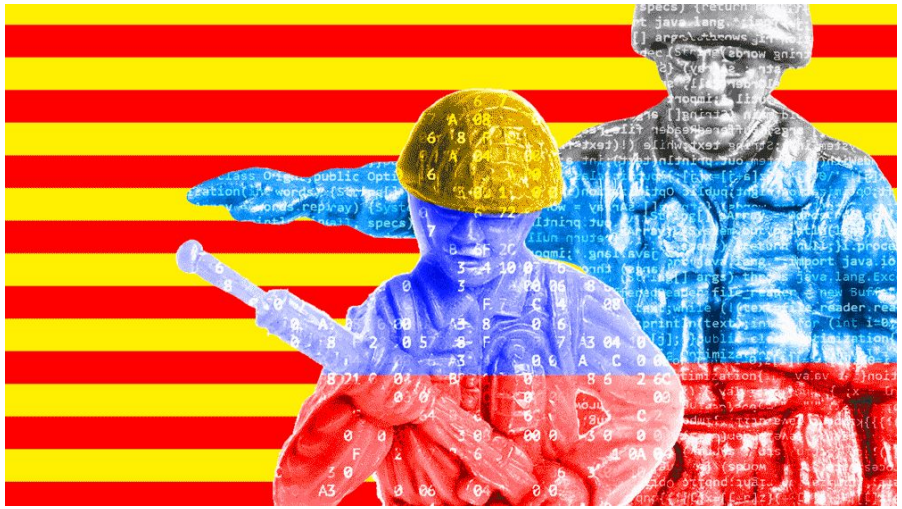
- Rumors
- Conspiracy theories
- Fake news
- Cyber bullying
- Harassment

Venezuela and Russia Teamed Up to Push Pro-Catalan Fake News

NOTICIAS FALSAS

Itxu Díaz

Updated Nov. 28, 2017 10:46AM ET
Published Nov. 28, 2017 5:00AM ET



MADRID—Europe is at war. Digital war. And it's very much the same fight that's taken place in the United States: facing an attack meant to sow distrust, heighten divisions, and undermine established democratic processes.

https://www.thedailybeast.com/why-is-venezuela-waging-cyber-war-in-europe

Twitter admits far more Russian bots posted on election than it had disclosed

Company says it removed more than 50,000 accounts and reported them to investigators, marking latest upward revision of figures

Twitter told Congress in October that it had found 36,746 Russian accounts that had posted about the election. Photograph: Anadolu Agency/Getty Images

Jon Swaine

@jonswaine
Sat 20 Jan 2018 00.46 GMT

Twitter has admitted that more than 50,000 Russia-linked accounts used its service to post automated material about the 2016 US election - a far greater number than previously disclosed.

https://www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed

The problem

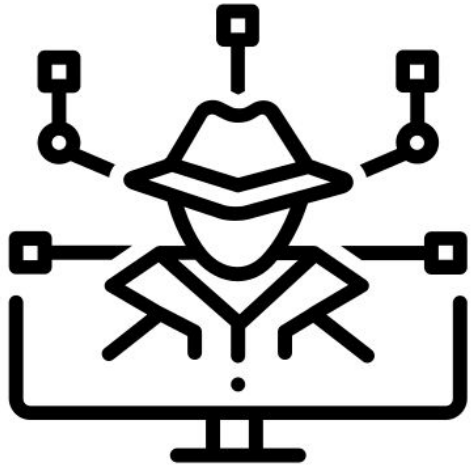
Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

How to uncovering Cybercrimes, like HSM (Hostile Social Manipulation) in Social Media through Natural Language Processing and Open Source Intelligence (OSINT)?

Our Goal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Design and implement the architecture of a cyber intelligence solution that allows to collect and process data from open sources, allowing to detect and prevent cyber crimes, specifically Hostile Social Manipulation (HSM)



Natural Language Processing

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

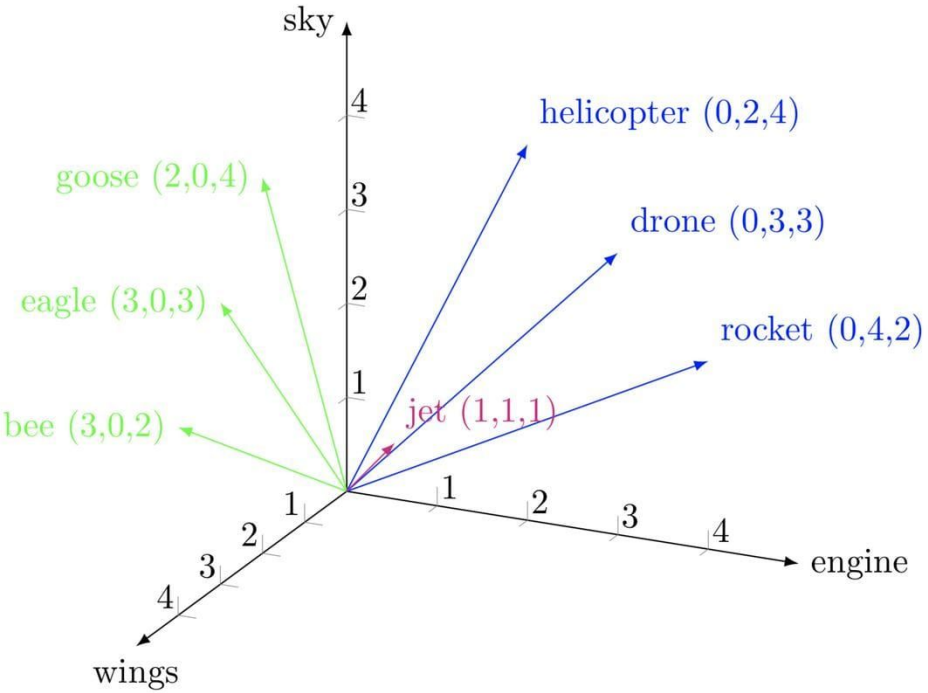
A set of machine learning technologies that gives computers the ability to understand, manipulate and structure human language.

Similarity models

The goal of similarity models is to measure the similarity between two or more objects based on their features or attributes.

Once the similarity between the objects is measured, they can be grouped, ranked, or classified based on their similarity scores (ej. **Cosine Similarity**).

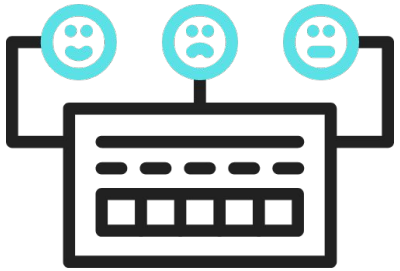
$$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta$$



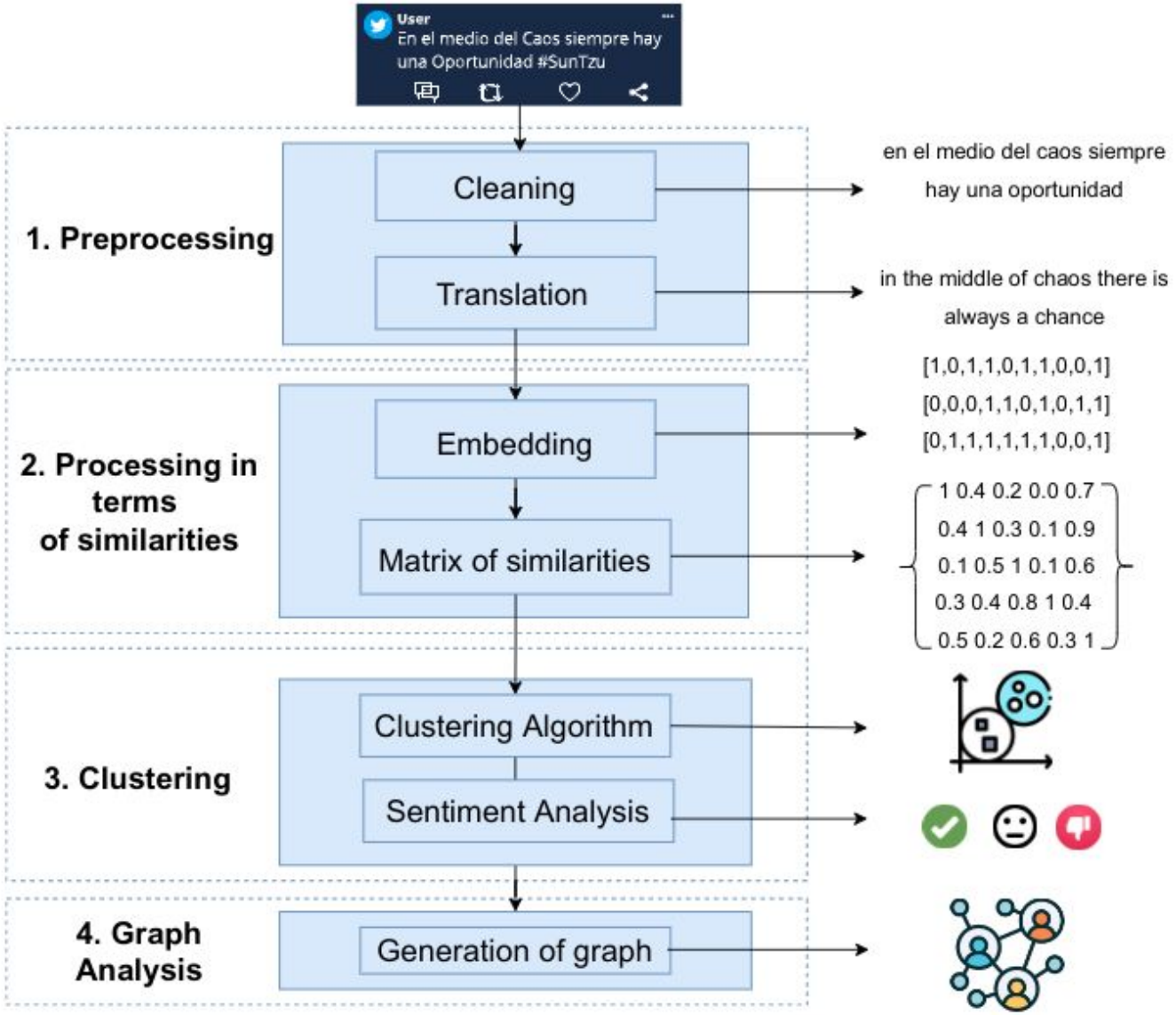
Sentiment and Polarity Models

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

An NLP technique that allows obtaining information about the sentiments contained in the text (Sentiment Analysis) and how positive or negative are the emotions printed in the analyzed text (Polarity Analysis).



Our proposal of architecture v.1.0



Solution that uses Neural Network (NN) to monitor suspicious activities in social networks allowing to identify and prevent related cyber crimes.

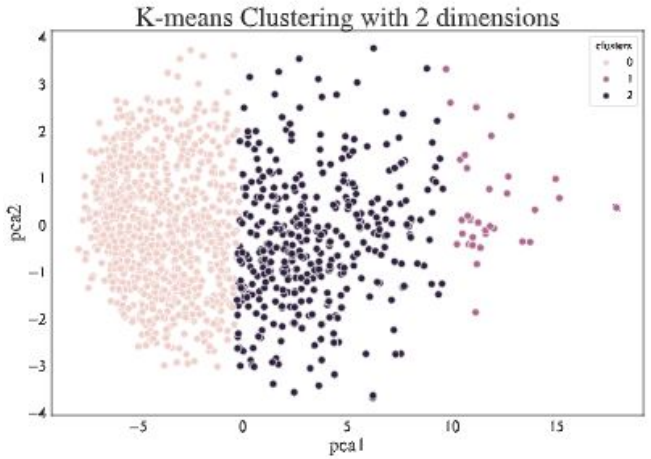
A LEA can find similar posts grouped in clusters, determine their level of polarity, and identify a subset of user accounts that promote violent activities to be reviewed extensively to prevent Hostile Social Manipulation (HSM).

Different experiments were conducted to prove the feasibility of the proposal.

Results (Violent protest on June 9th 2021 in Colombia)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Three clusters of tweets with their corresponding word maps identified from the analysis of similarities

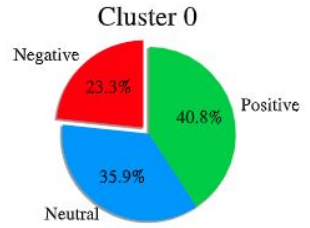


Cluster 0

Cluster 1

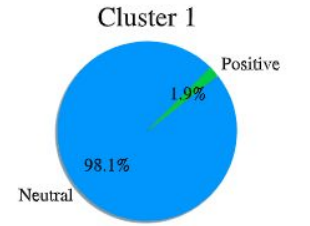
Cluster 2

Clusters sentiment polarity distribution and a sample tweet for each cluster



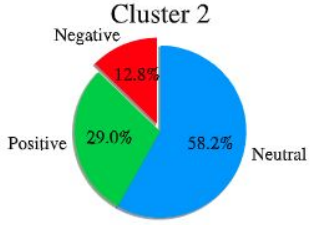
User
Another sinister night in Cali, an attack against the protesters who were in the Siloé roundabout. Even when? #CaliEnPerigro

User
Otra noche siniestra más en Cali, ataque contra lxs manifestantes que estaban en la glorieta de Siloé. Hasta cuándo? #CaliEnPeligro



User
@eu_eeas @EUAM_RCA #SOSCALICOLOMBIA @JosepBorrellF necesitamos una comisión de verificación urgente

User
@eu_eeas @EUAM_RCA #SOSCALICOLOMBIA @JosepBorrellF we need an urgent verification commission

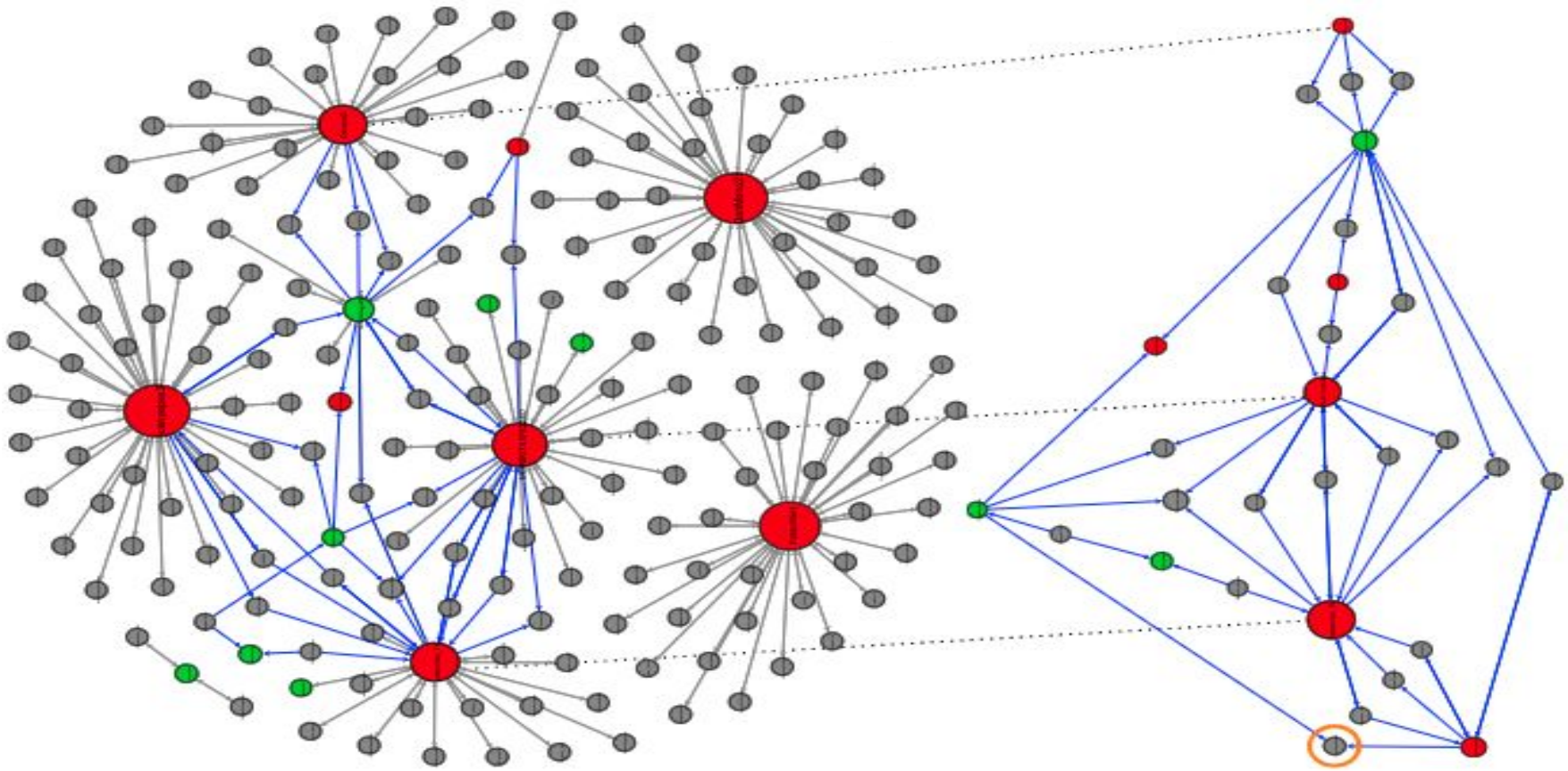


User
Los Policías de Colombia son los peores de #Latinoamérica (junto a la guardia de Venezuela). #ParoNacional9J

User
The Colombian Police are the worst in #LatinAmerica (along with the Venezuelan guard). # StopNational9J

Results (Violent protest on June 9th 2021 in Colombia

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Graph of nodes belonging to cluster 2 showing unique and commons followers

Universidad del Rosario

MACC
Matemáticas Aplicadas y Ciencias de la Computación

FCTNLP: An Architecture to fight Cyber Terrorism with Natural Language Processing

Who can be a customer of these kind of solutions?

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Government agencies with sensitive national security information or critical data

→ Department of Defense (DoD)
Federal Bureau of Investigation (FBI)
National Security Agency (NSA)
Department of Homeland Security (DHS)

Financial institutions that control investments

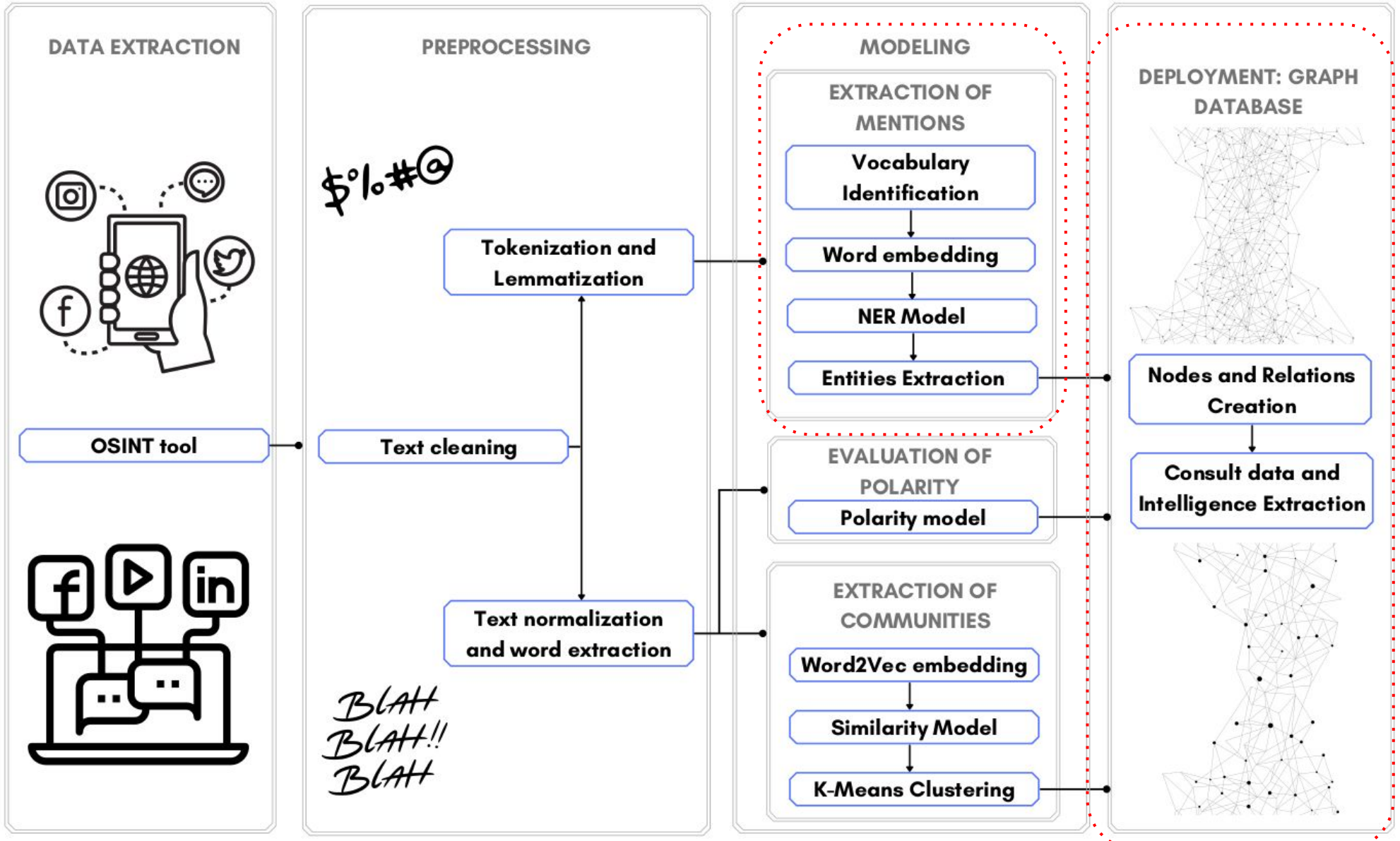
→ Banks and credit union, Investment firms, Insurance companies, Payment processing companies

Cyber security companies that offer managed security services to other companies

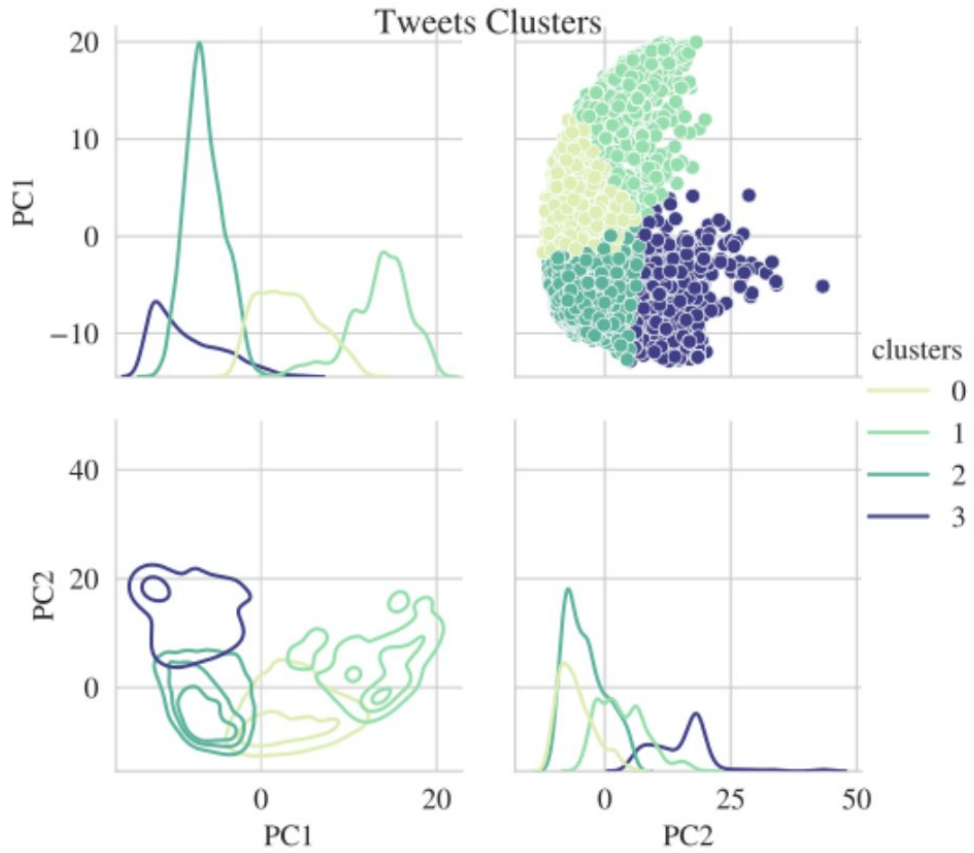
→ IBM Security, Symantec Corporation, FireEye, Inc., SecureWorks Corp., Cisco Systems, Inc., Accenture Security

Our proposal of architecture v.2.0

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Results (violent protest on Oct 26th 2021 in Ecuador)



Tweet collection, text cleaning, text normalization and word extraction

Embedding collected tweets with Google News Embeddings

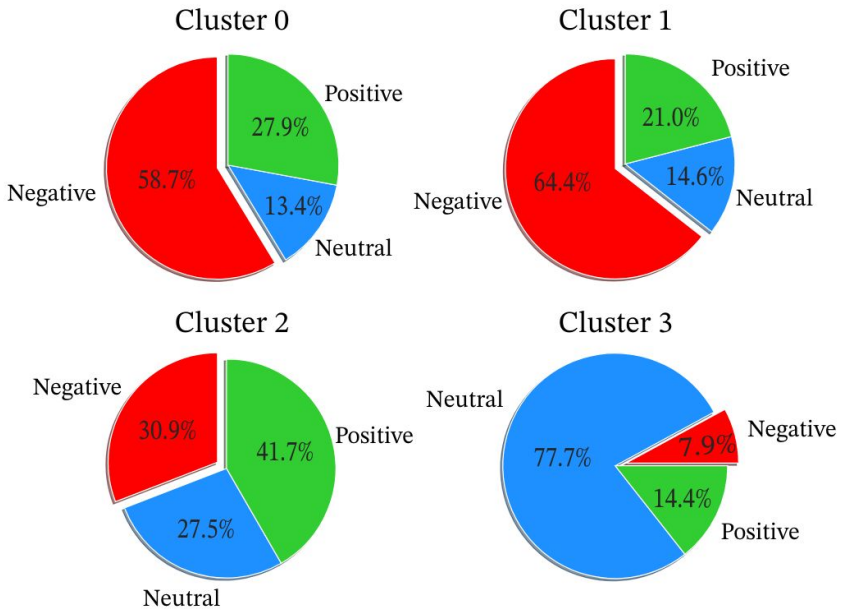
Similarity model

K-Means Clustering

Results (violent protest on Oct 26th 2021 in Ecuador)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Polarity for each detected cluster



Classification based on a Single Layer Perceptron supported by TextBlob Python library

NER (Name Entity Recognition) Predictions and percentage of accuracy

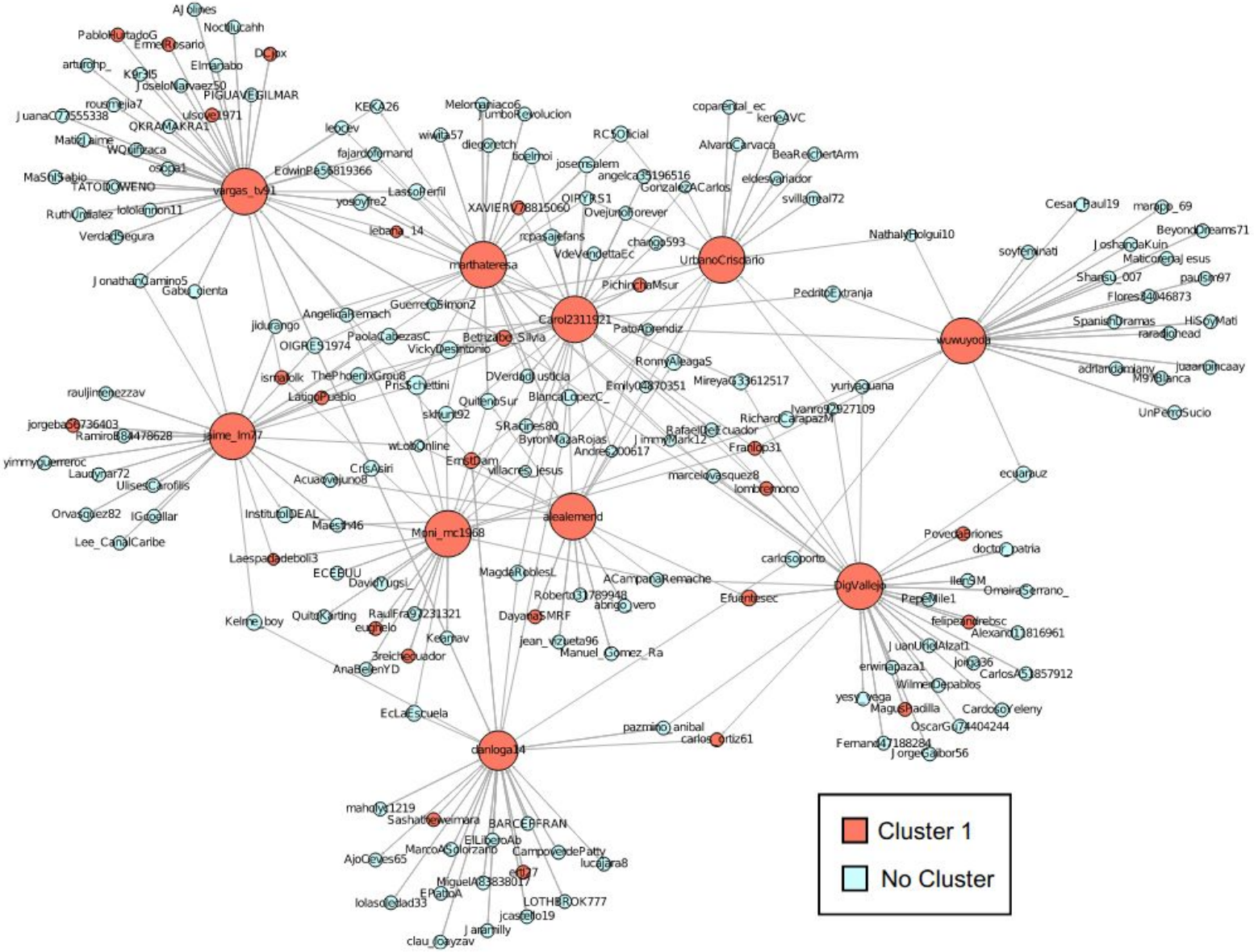
token	ORG	LOC	PER	MISC	Total	Percentage
Paro	3426	1801	196	74	5497	62.3% (ORG)
Lasso	353	3145	2125	264	5887	36.1% (PER)
Nacional	2661	2415	204	97	5377	49.5% (ORG)
Ecuador	724	2535	14	60	3333	76.1% (LOC)
Guillermo	21	36	1299	3	1359	95.6% (PER)
Fuera	115	1067	85	17	1284	1.3% (MISC)
Renuncia	10	593	56	3	662	0.5% (MISC)
Pandora	9	205	558	8	780	0.1% (MISC)

Results (violent protest on Oct 26th 2021 in Ecuador)

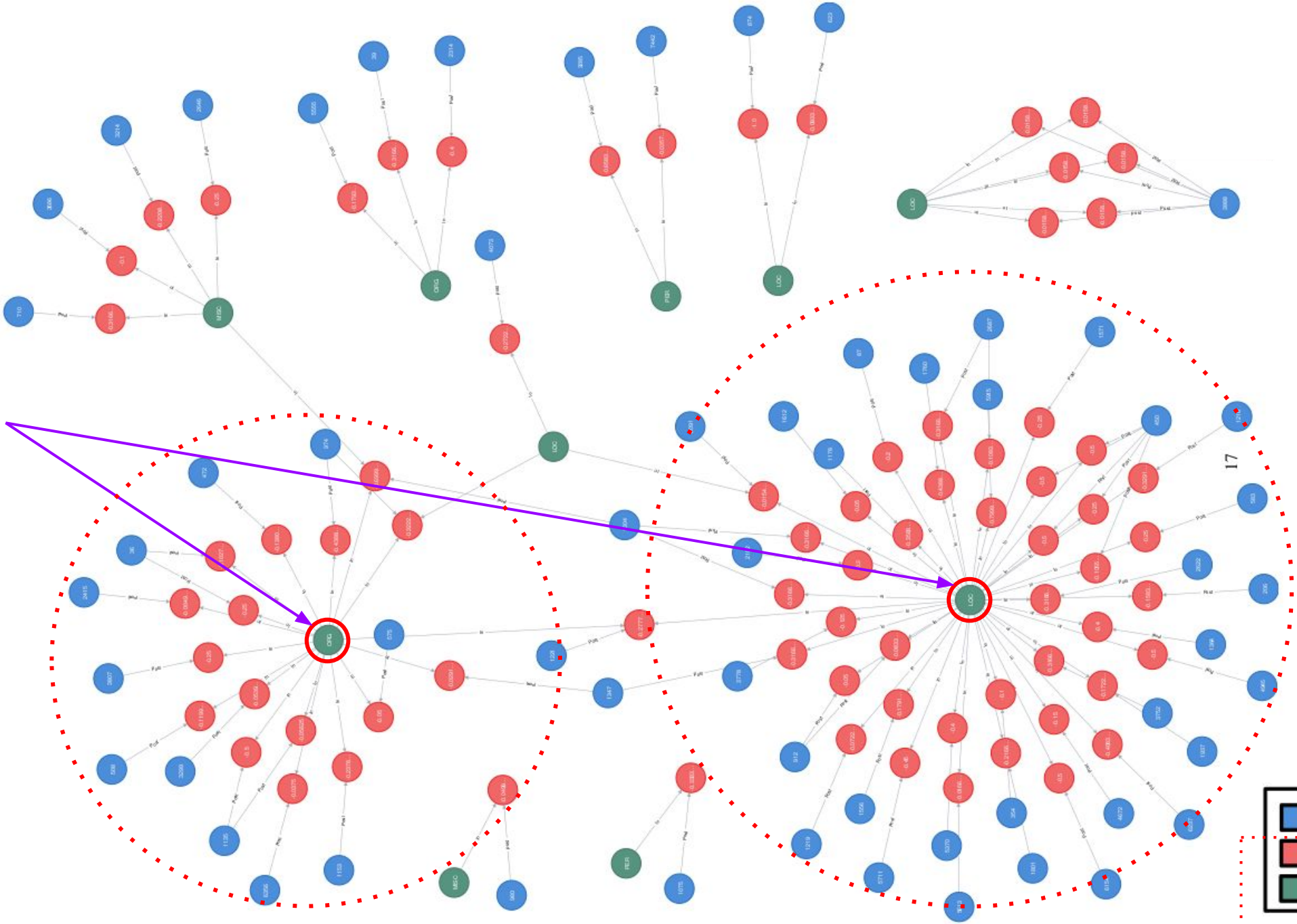
Universidad del Rosario

MACC Matemáticas Aplicadas y Ciencias de la Computación

Graph of nodes for the most violent cluster (Cluster 1)



Hostility detected around a specific ORG and LOC



What we have learnt until now?

1. OSINT generalities
2. Cyber intelligence generalities
3. Some OSINT tools: Tinfoleak, Gephy
4. Different NLP models usable for OSINT: Similarity, Sentiment, Polarity, NER
5. NLP allows monitor and detect suspicious activities expressed through human language: like hate promotion, violent speech or systematic terrorism
6. NLP + Cyber Intelligence allows to support LEA's in the fight against cybercrime

Next goal?

Next Goal

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

An NLP-based framework to detect hate speech and harassment on US congress elections 2022

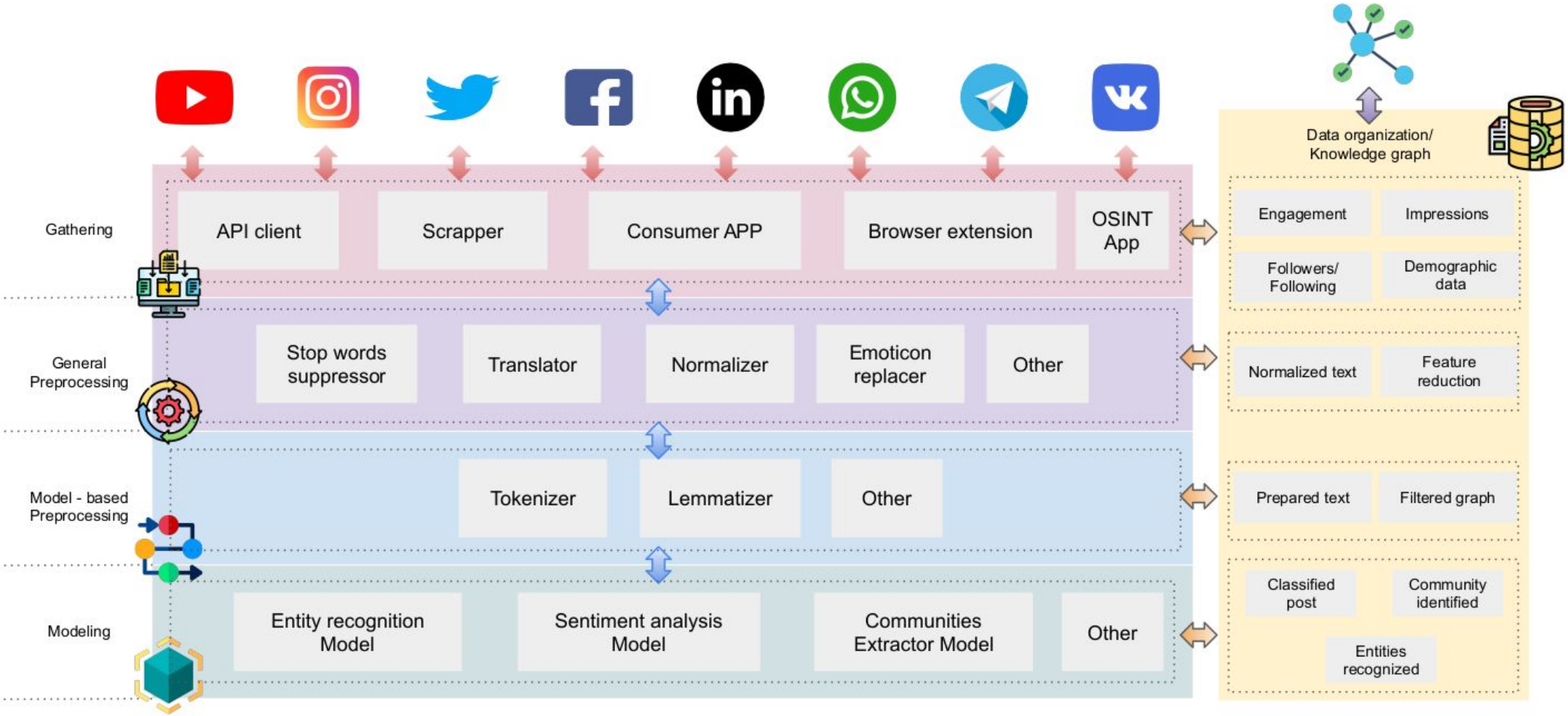
Election day:
Nov 8th, 2022
Monitoring period:
Oct 24th to Nov 17th, 2022

Strategy:
Monitor replies to Twitter accounts from Secretaries of States
Main method used:
get_tweets_from_user(user_id)

NYU twitter FLASHPOINT

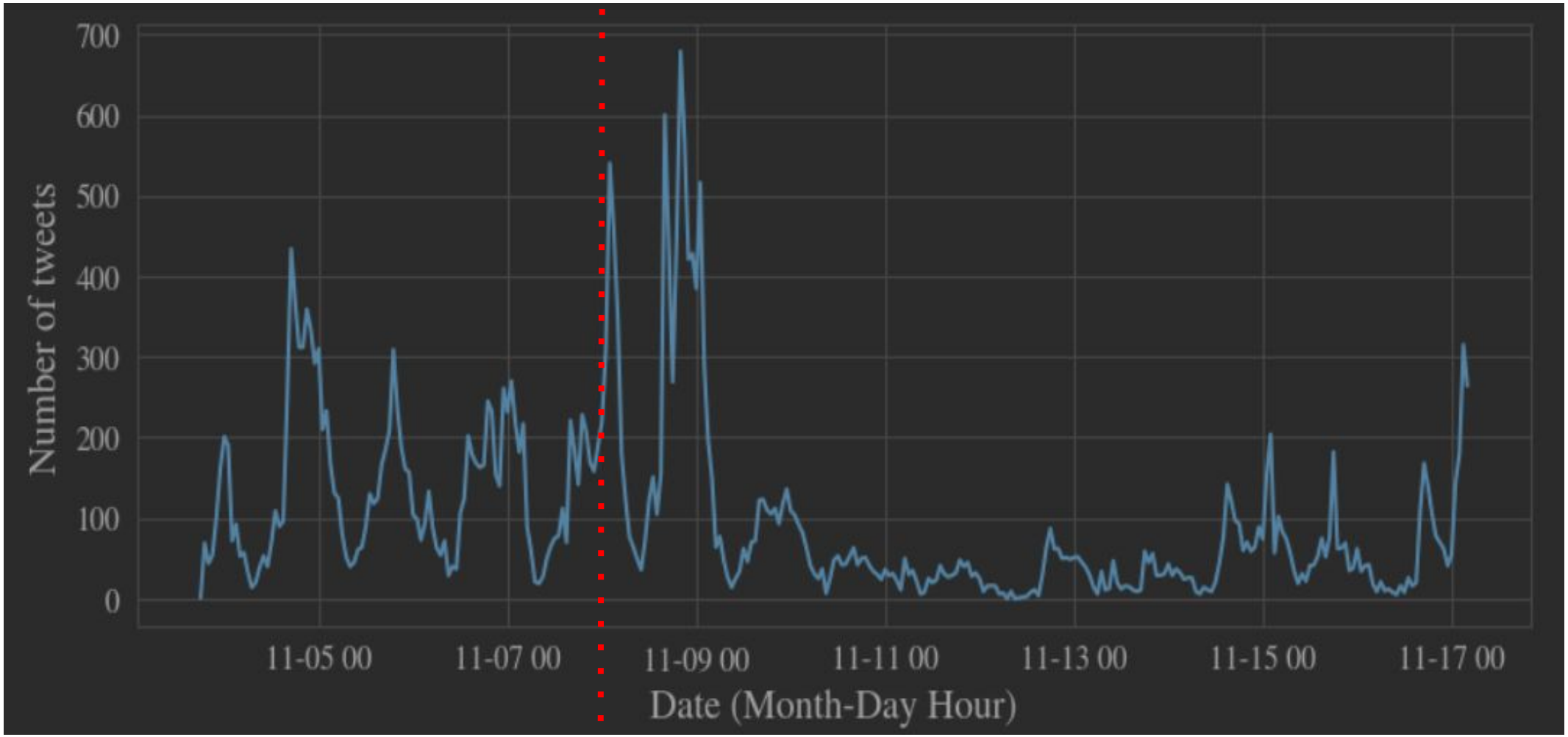
Our proposal of architecture v.3.0

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Results (US Elections on Nov 8th 2022)

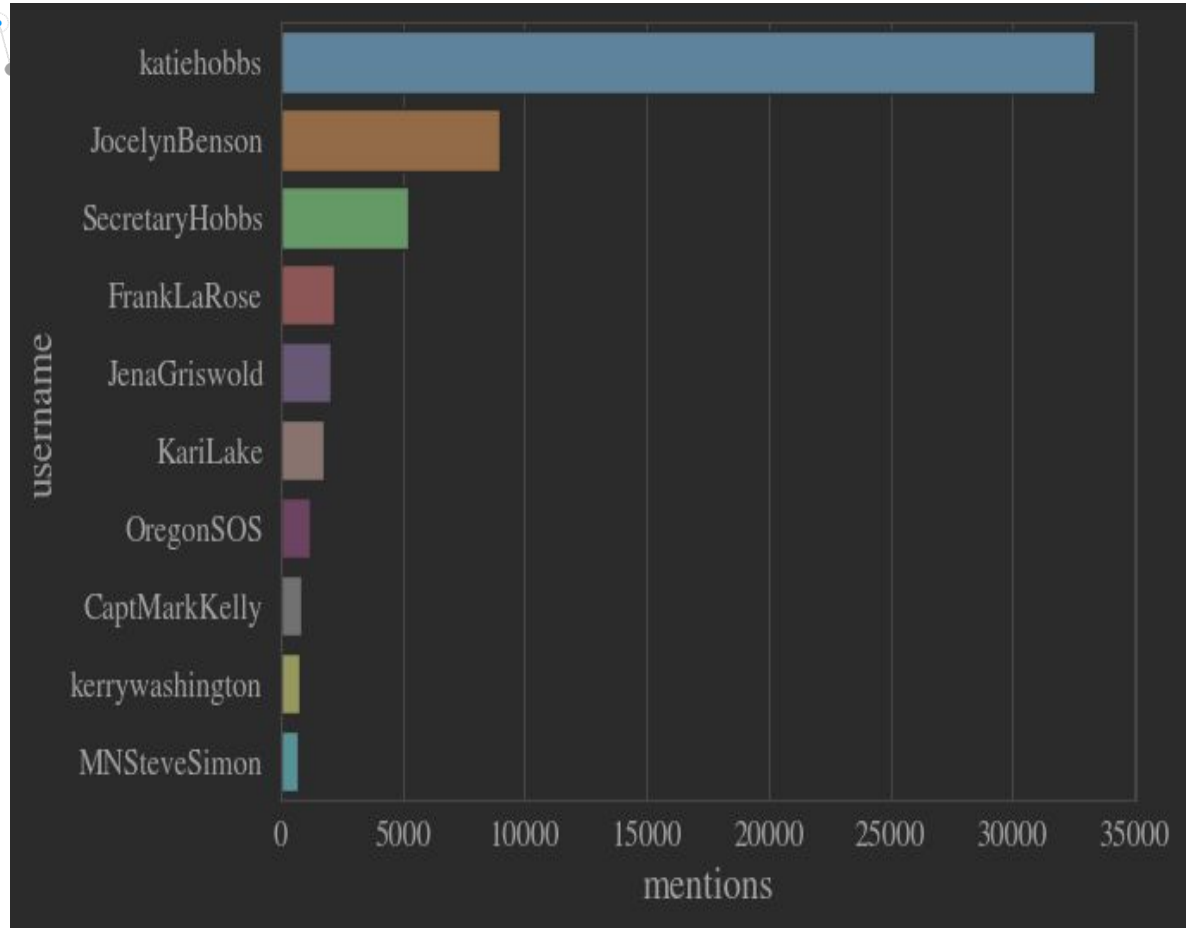
Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



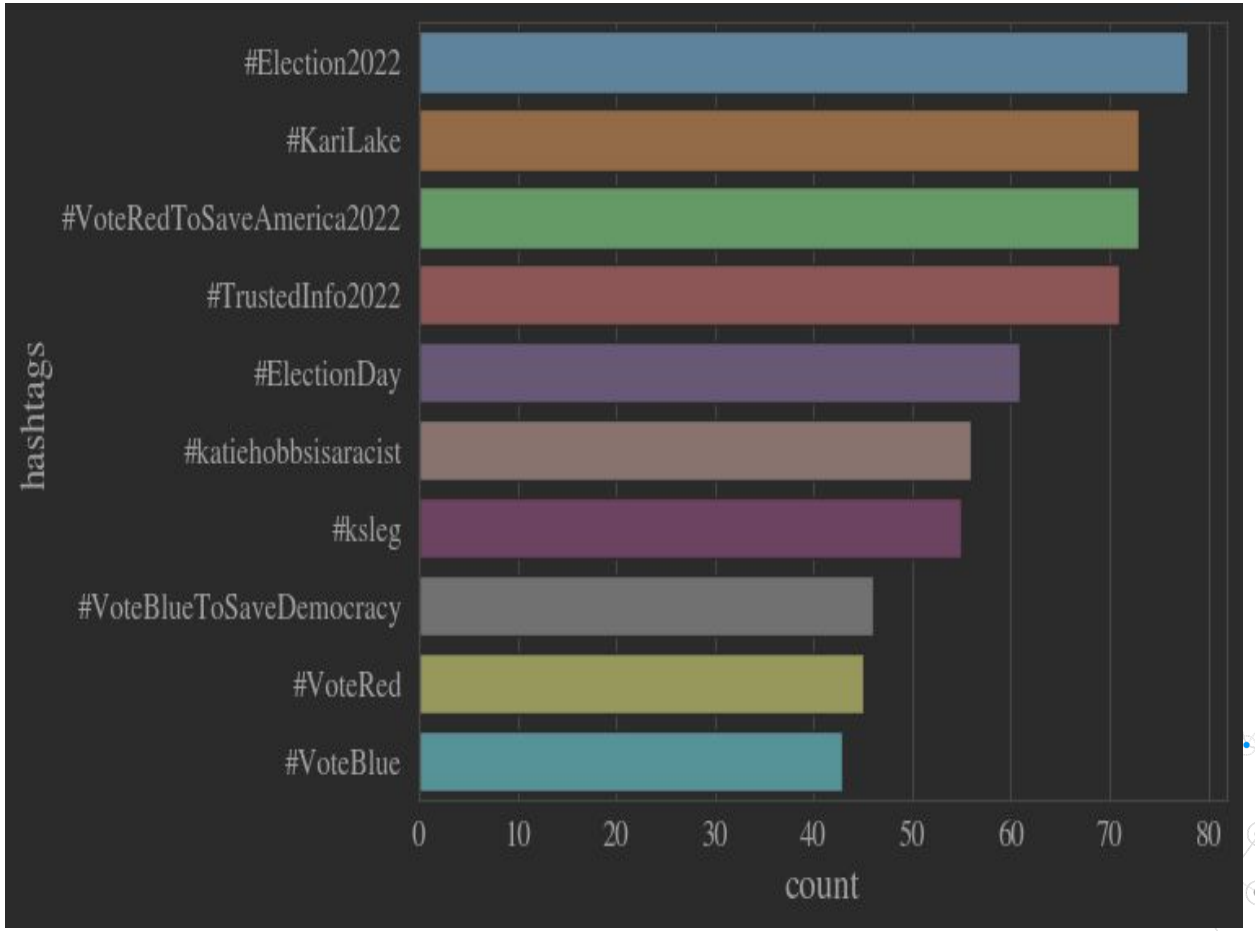
Histogram of number of tweets per day along the days **previous** and **after** the election day (Nov 8th)

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

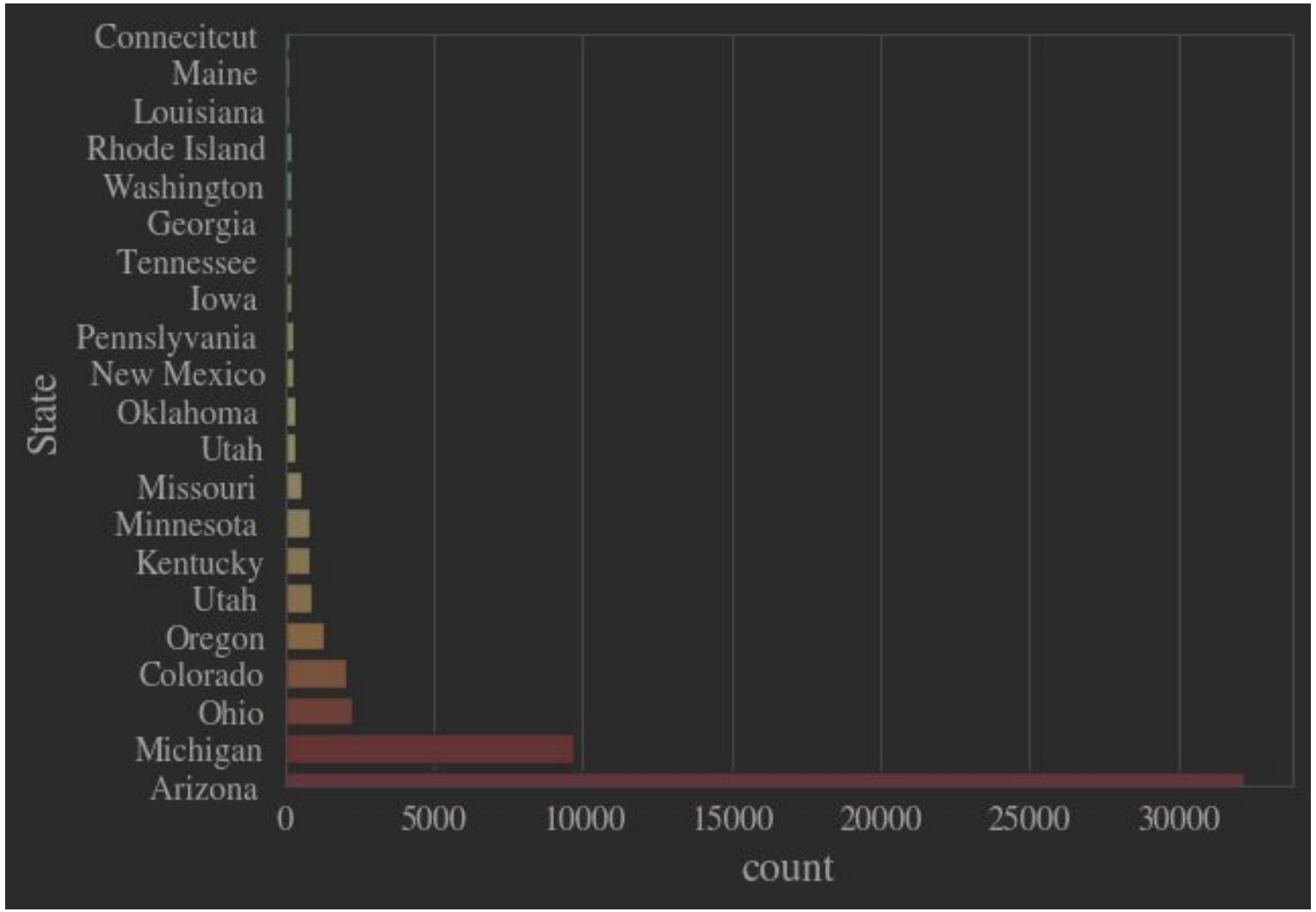


Number of mentions per usernames



Number of counts per hashtags

Results (US Elections on Nov 8th 2022)

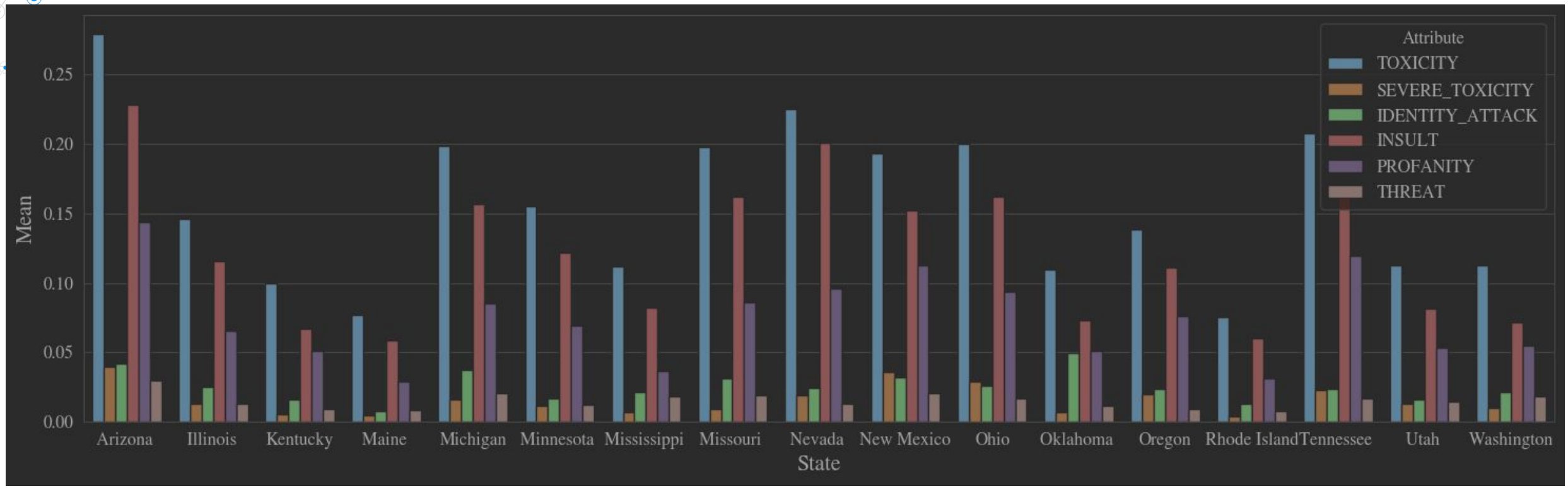


Number of tweet replies per accounts associated to Secretaries of State's accounts

Arizona is the state with the most amount of replies (almost 33K) to a Secretary of State followed by Michigan

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



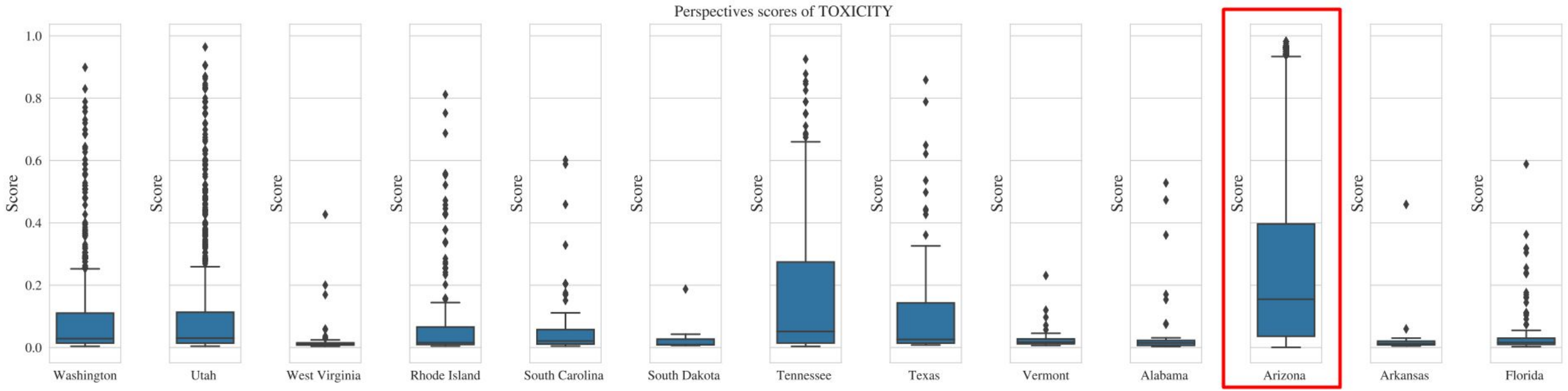
Results of the execution of Perspective API Engine over the data collected per state

Perspective

https://perspectiveapi.com/

JIGSAW

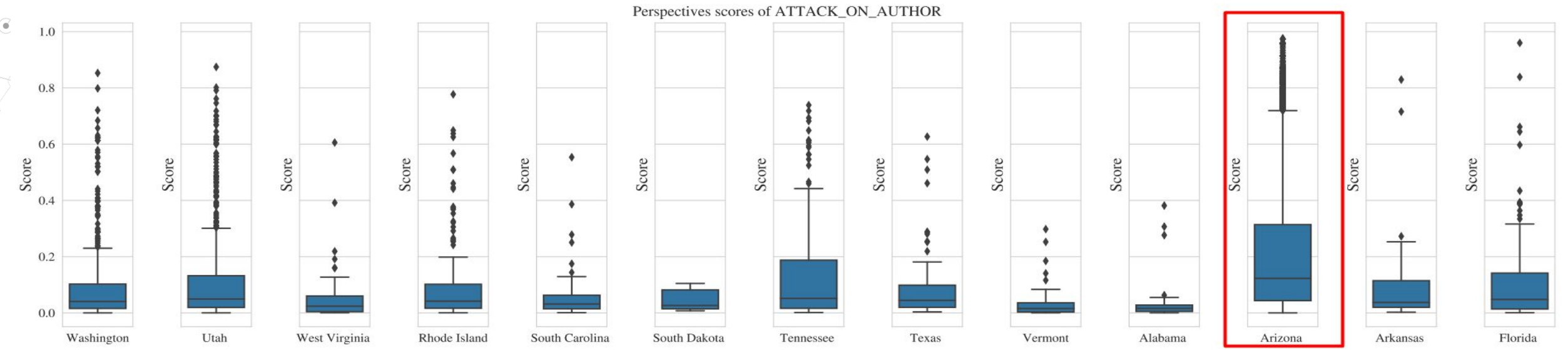
Results (US Elections on Nov 8th 2022)

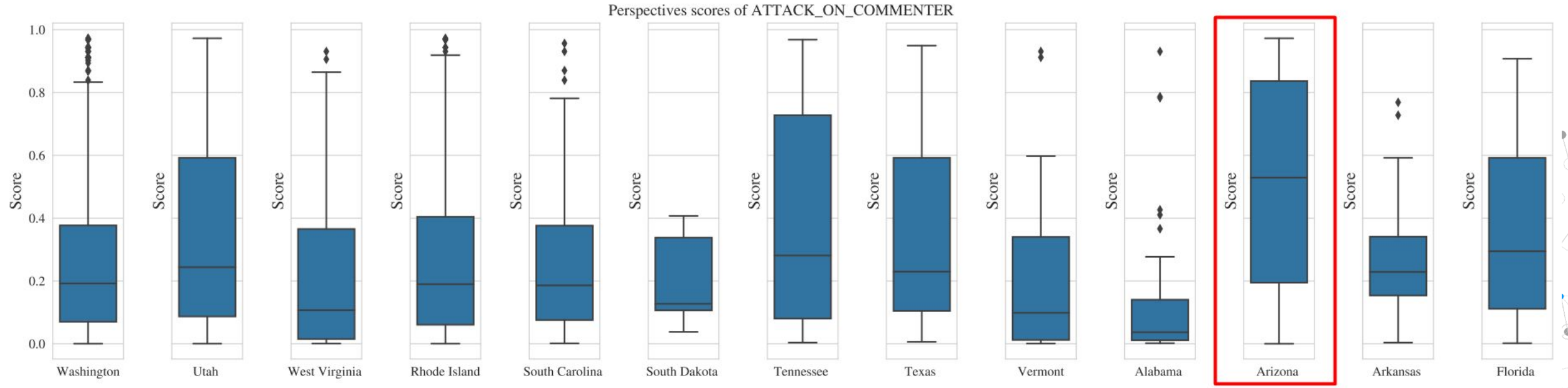


Boxplot “Toxicity” per State

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



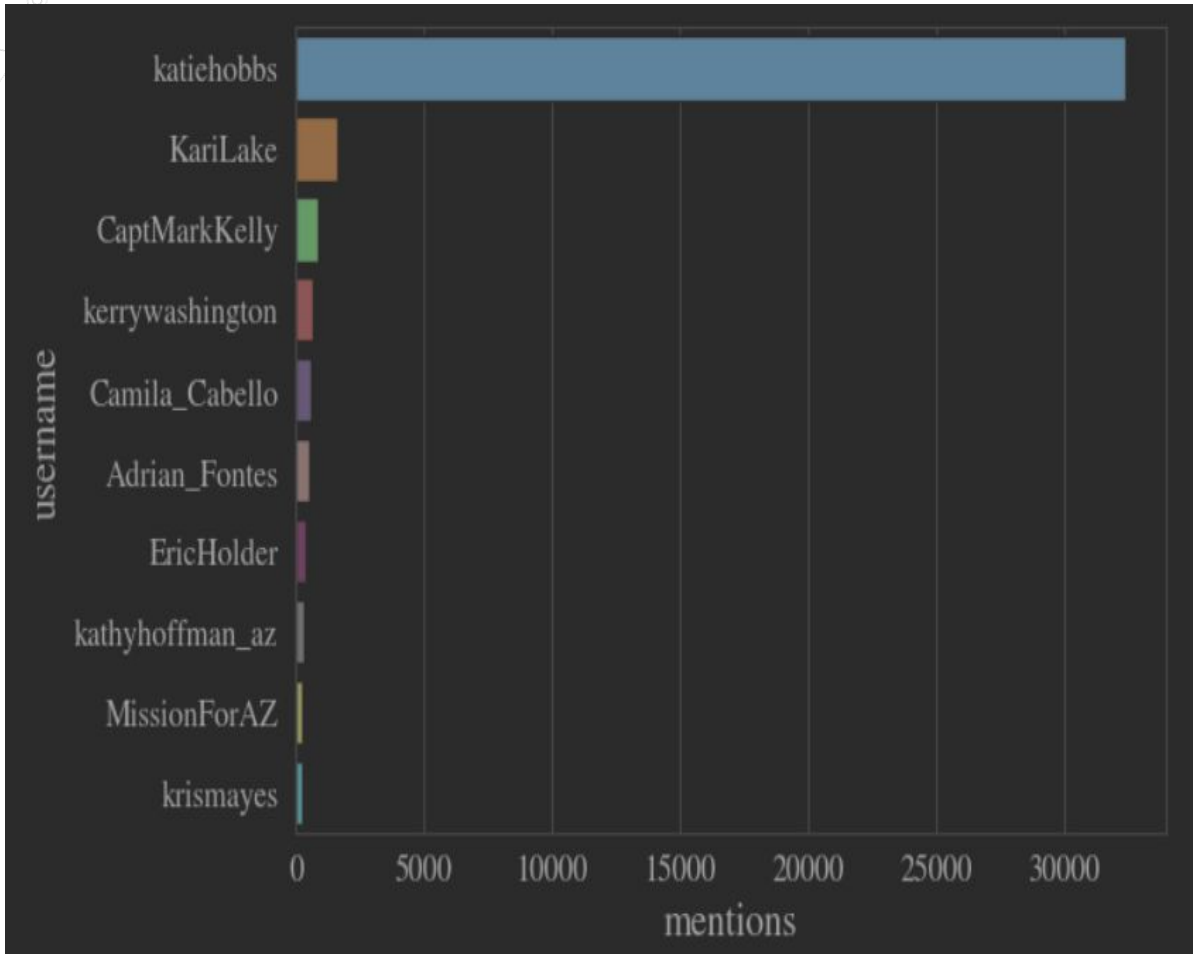


Results (US Elections on Nov 8th 2022)

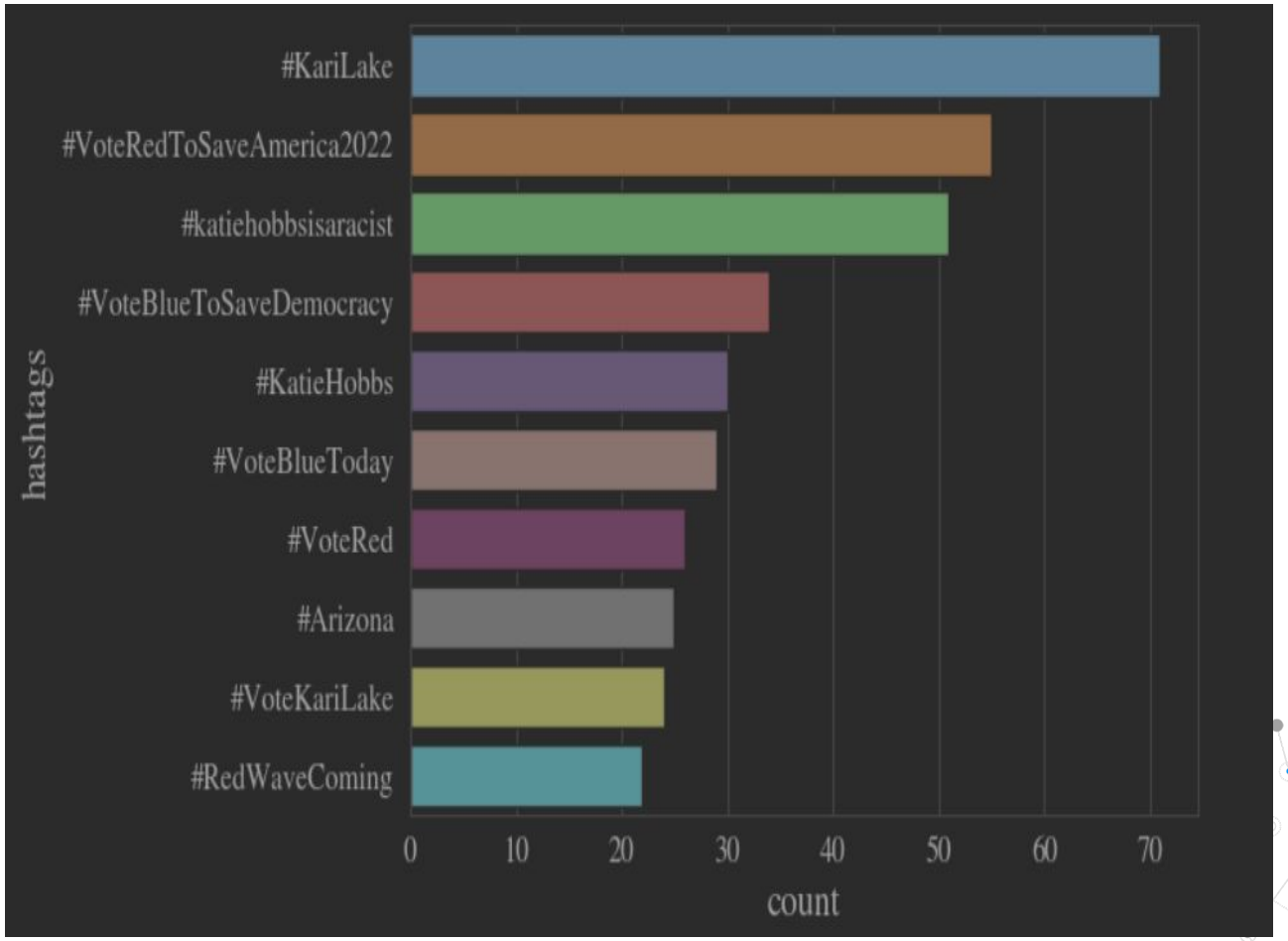
The statistics show that the state of Arizona has higher levels of toxicity, insults, and profanity

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

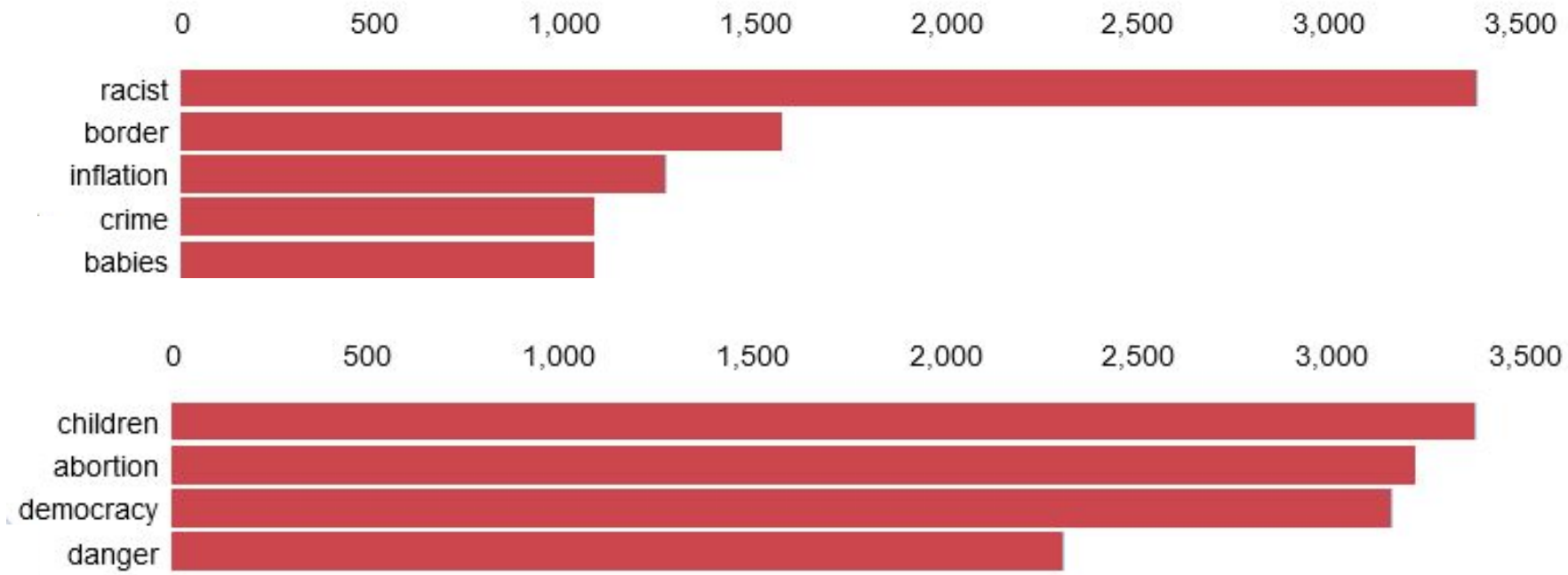


Number of mentions per usernames in Arizona



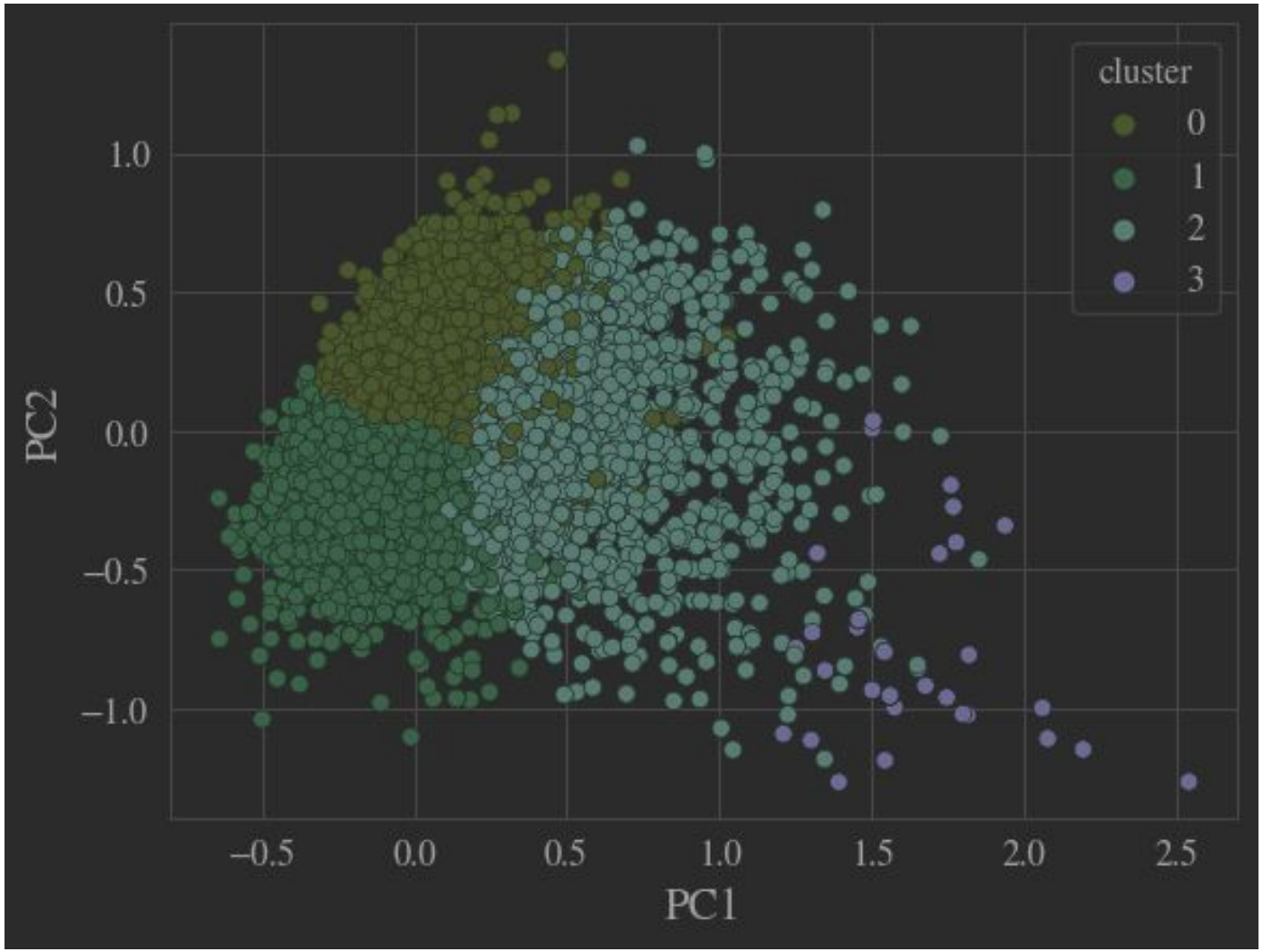
Number of counts per hashtags identified in Arizona

Results (US Elections on Nov 8th 2022)



The most highlighted topics in Arizona are inflation, crime, children, abortion, and democracy.

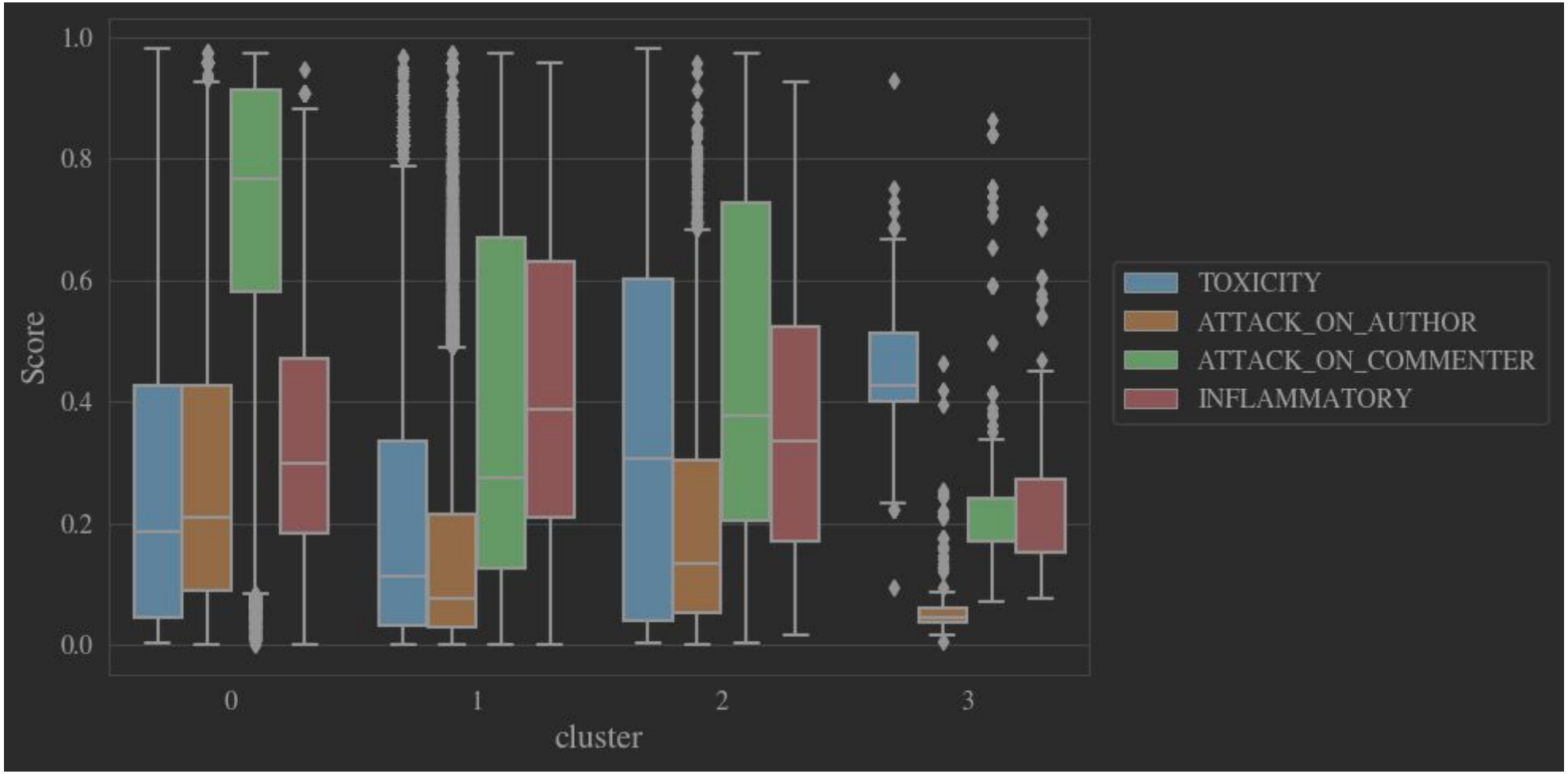
Results (US Elections on Nov 8th 2022)



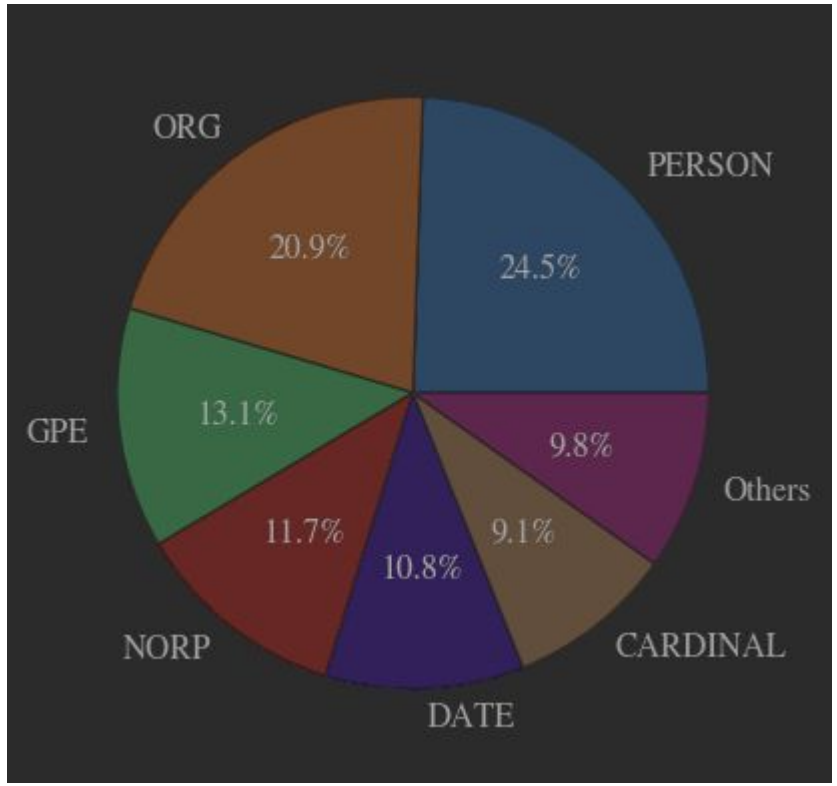
Clustering results over the Arizona's tweets using Google News Word2Vec embeddings

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Results (US Elections on Nov 8th 2022)



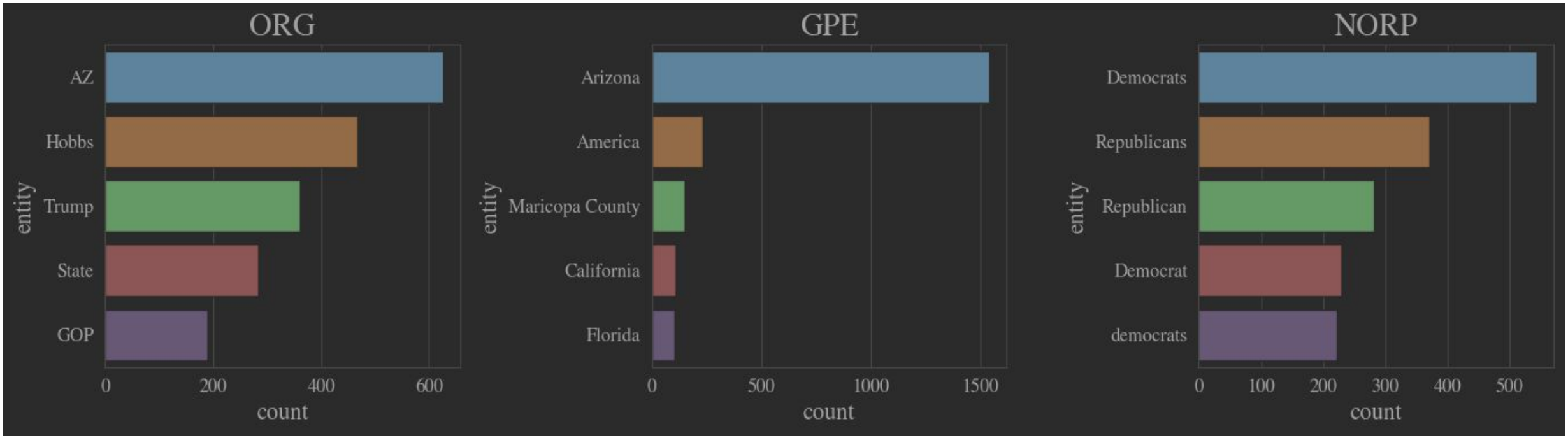
	entity	cluster	TOXICITY
0	KAtIe	2.0	0.956375
1	Mark Kelley	2.0	0.950486
2	Communist Party	2.0	0.939145
3	Denial	0.0	0.928801
4	Lane	2.0	0.928801

	entity	cluster	ATTACK_ON_AUTHOR
0	M.O.	0.0	0.973236
1	Paulo Freire	0.0	0.960000
2	Peoria School	0.0	0.960000
3	Stephanie Clark	0.0	0.960000
4	Good day	1.0	0.957282

	entity	cluster	INFLAMMATORY
0	Tyrants	0.0	0.946667
1	Big Labor Union Donors	1.0	0.927948
2	Killing Babies	1.0	0.927948
3	Rethugs	1.0	0.927948
4	Joseph Goebbels	2.0	0.926956

Results (US Elections on Nov 8th 2022)

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación



Some of our publications

- A. Zapata, **D. Díaz-López**, J. Pastor-Galindo, and F. Gómez Mármol, "FCTNLP: An Architecture To Fight Cyberterrorism With Natural Language Processing," in VII National Conference on Cybersecurity Research (JNIC), Bilbao, Spain, vol. 1, pp. 42–49, 2022.

- J. Ramirez Sánchez, A. Campo-Archbold, A. Zapata Rozo, **D. Díaz-López**, J. Pastor-Galindo, F. Gómez Mármol, J. Aponte Díaz, "On The Power Of Social Networks To Analyze Threatening Trends", IEEE Internet Computing (Q1), vol. 2022, 9 pages, 2022, doi: 10.1109/MIC.2022.3154712.

- J. Ramirez Sánchez, A. Campo-Archbold, A. Zapata Rozo, **D. Díaz-López**, J. Pastor-Galindo, F. Gómez Mármol, J. Aponte Díaz, "Uncovering Cybercrimes In Social Media Through Natural Language Processing", Complexity (Q1), vol. 2021, Article ID 7955637, 15 pages, 2021

- J. Ibañez, S. Rocha, **D. Díaz-López**, J. Pastor-Galindo, F. Gómez. "C3-Sex: A Conversational Agent To Detect Online Sex Offenders". Electronics, SI Advanced Cybersecurity Services Design", Electronics (Q2), October 2020, 9(11).

- J. Murcia, S. Moreno, **D. Díaz-López**, and F. Gómez Mármol . "C3-Sex: A Chatbot To Chase Cyber Perverts". The 4th IEEE Cyber Science and Technology Congress. Fukuoka, Japan, 2019.

An Inspiring Ending Phrase …

"Unlock the power of knowledge, stay one step ahead of the game: With Cyber Intelligence, you have the ability to detect, prevent, and respond to cyber threats, ensuring the safety and security of your organization's valuable assets."

Universidad del Rosario | MACC Matemáticas Aplicadas y Ciencias de la Computación

Thanks!