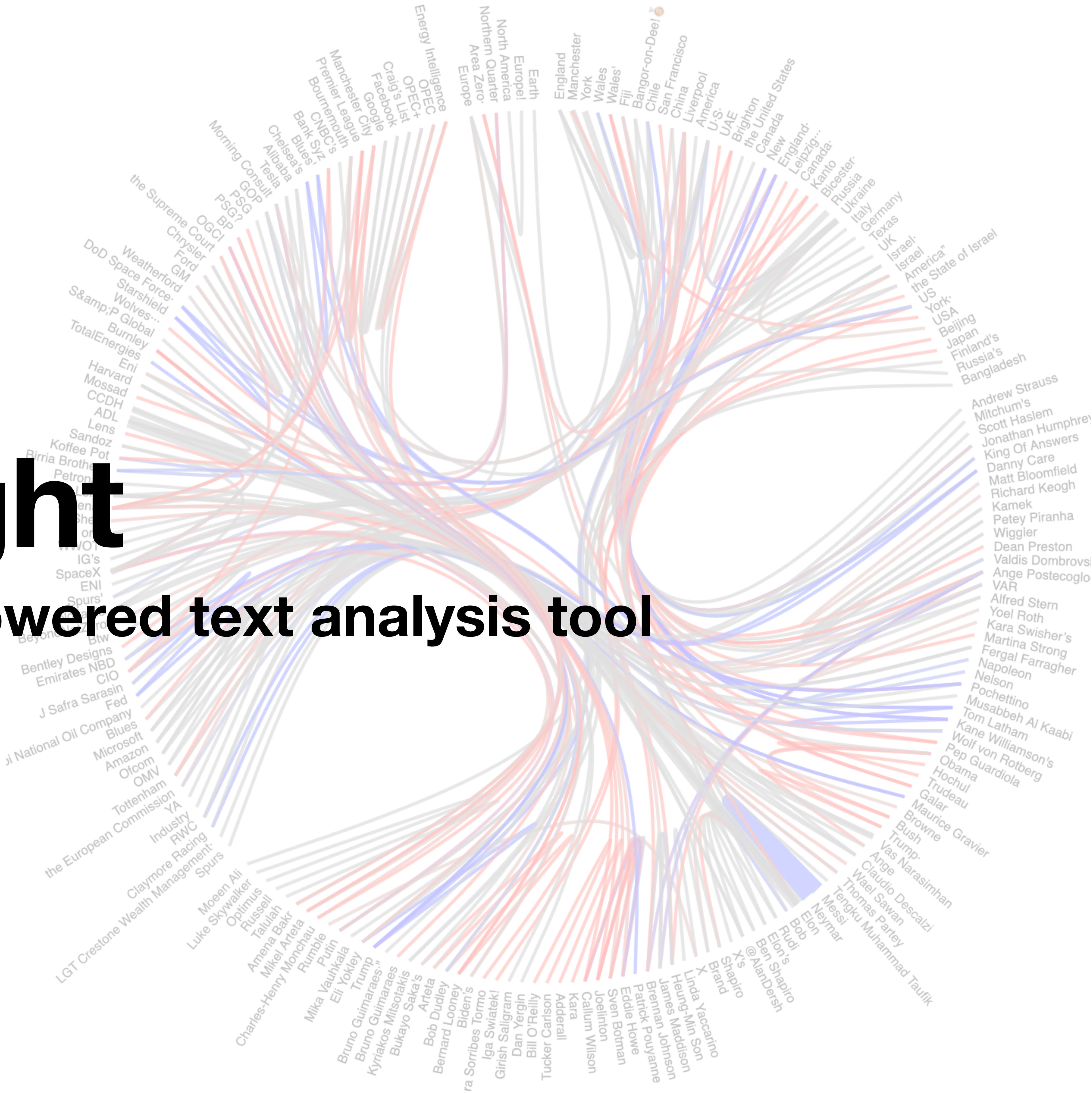
ARElight

AI & graph powered text analysis tool



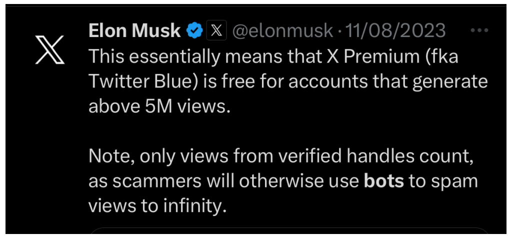
Nikolay R., **Maxim K.**

Structure

- The problem
- Description of ARElight
- Demo
- Conclusions

The problem

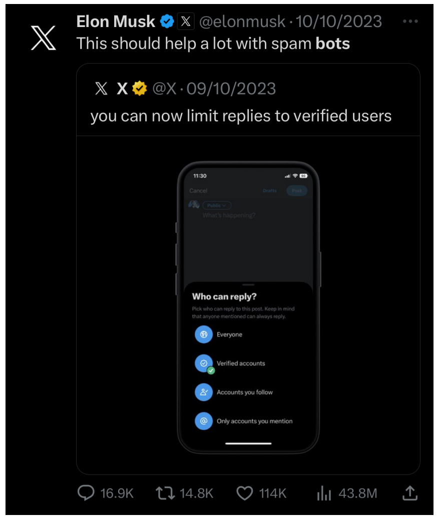
- When speaking about disinformation, there exist 2 walls of defence:
  - protection by social media company
  - protection by users themself



Elon Musk @elonmusk · 11/08/2023

This essentially means that X Premium (fka Twitter Blue) is free for accounts that generate above 5M views.

Note, only views from verified handles count, as scammers will otherwise use **bots** to spam views to infinity.



Elon Musk @elonmusk · 10/10/2023

This should help a lot with spam **bots**

X @X · 09/10/2023

you can now limit replies to verified users

11:30

Cancel Drafts Post

Public

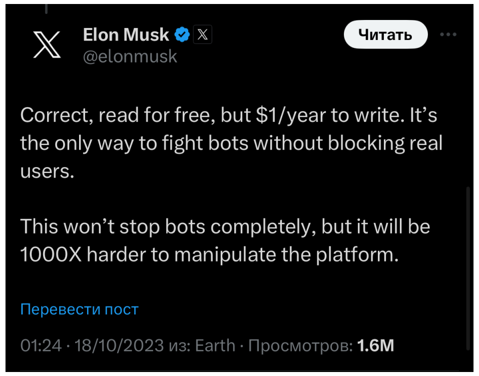
What's happening?

Who can reply?

Pick who can reply to this post. Keep in mind that anyone mentioned can always reply.

- Everyone
- Verified accounts
- Accounts you follow
- Only accounts you mention

16.9K 14.8K 114K 43.8M



Elon Musk @elonmusk

Читать

Correct, read for free, but $1/year to write. It's the only way to fight bots without blocking real users.

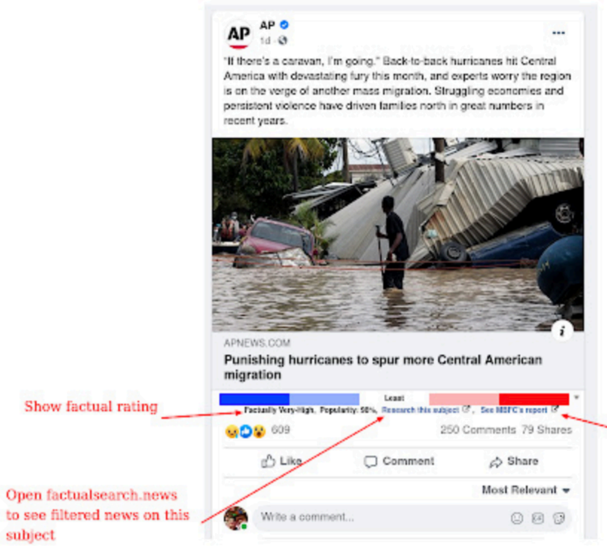
This won't stop bots completely, but it will be 1000X harder to manipulate the platform.

Перевести пост

01:24 · 18/10/2023 из: Earth · Просмотров: **1.6M**

disinfo

measures by social media



AP 1d

"If there's a caravan, I'm going." Back-to-back hurricanes hit Central America with devastating fury this month, and experts worry the region is on the verge of another mass migration. Struggling economies and persistent violence have driven families north in great numbers in recent years.

APNEWS.COM

Punishing hurricanes to spur more Central American migration

Least

Factually Very-High, Popularity: 98%, Research this subject, See MBFC's report

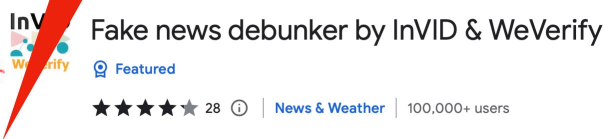
609 250 Comments 79 Shares

Show factual rating

Open factualsearch.news to see filtered news on this subject

Botometer®

An OSoMe project (bot•o•meter)



Fake news debunker by InVID & WeVerify

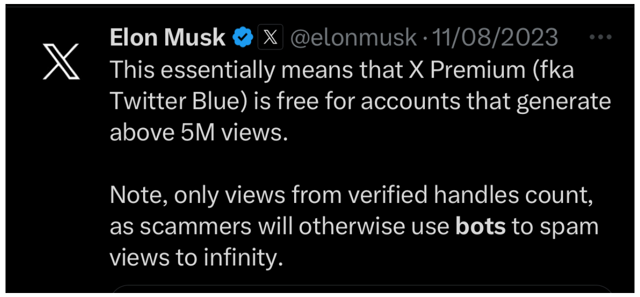
Featured

★★★★★ 28 | News & Weather | 100,000+ users

measures by users

The problem

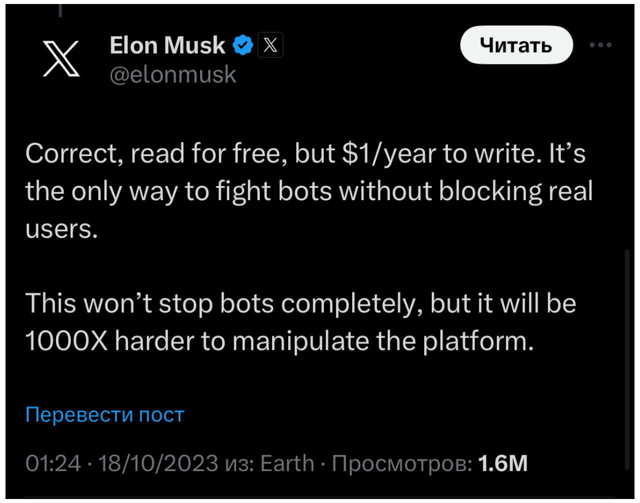
- Measures by SM examples:
  - Twitter/X efforts to decrease number of bots
  - Block/Shadowban of suspicious accounts
  - Tinder AI safety features



Elon Musk @elonmusk · 11/08/2023

This essentially means that X Premium (fka Twitter Blue) is free for accounts that generate above 5M views.

Note, only views from verified handles count, as scammers will otherwise use **bots** to spam views to infinity.



Elon Musk @elonmusk

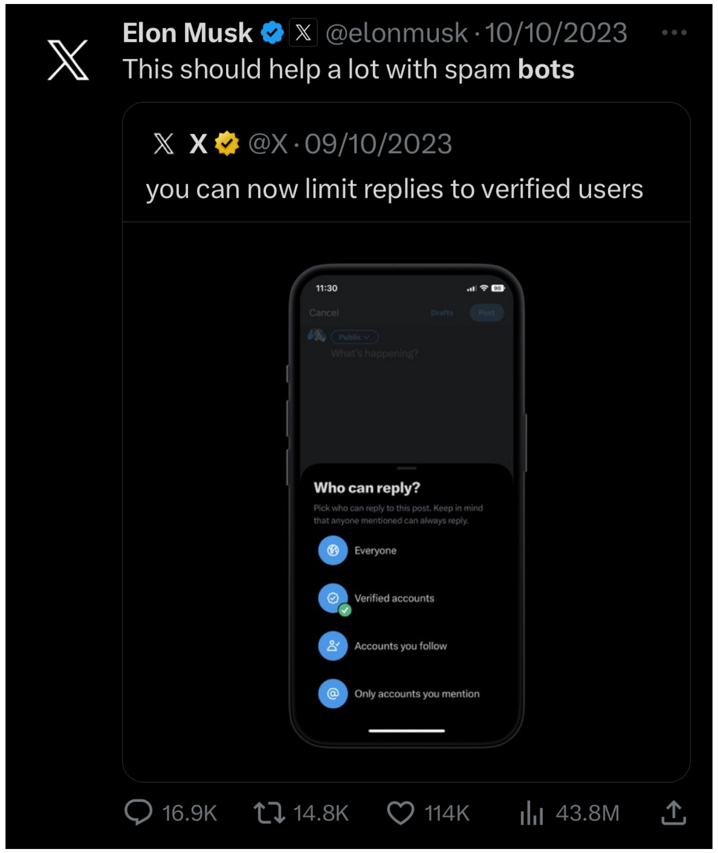
Читать

Correct, read for free, but $1/year to write. It's the only way to fight bots without blocking real users.

This won't stop bots completely, but it will be 1000X harder to manipulate the platform.

Перевести пост

01:24 · 18/10/2023 из: Earth · Просмотров: **1.6M**



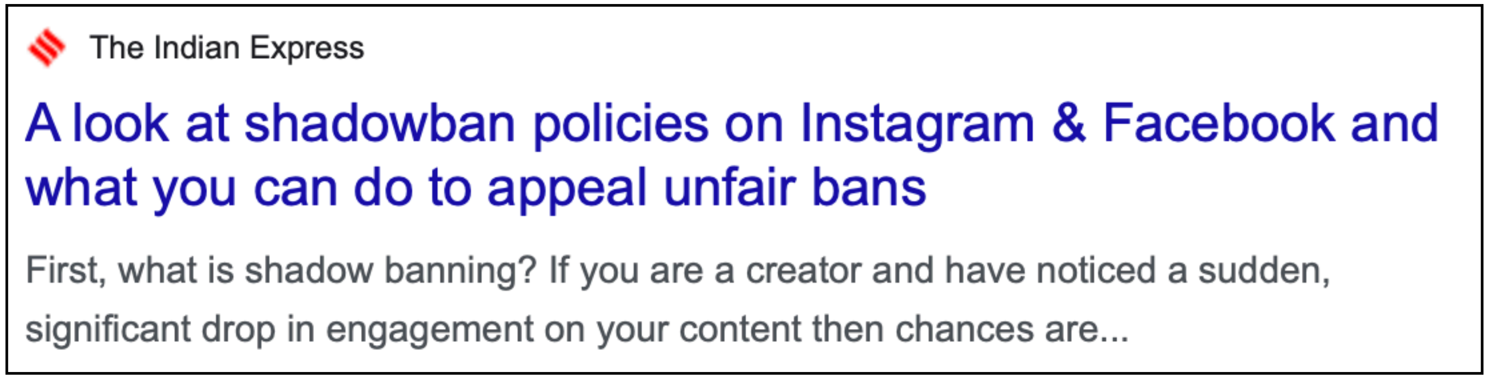
Elon Musk @elonmusk · 10/10/2023

This should help a lot with spam **bots**

X @X · 09/10/2023

you can now limit replies to verified users

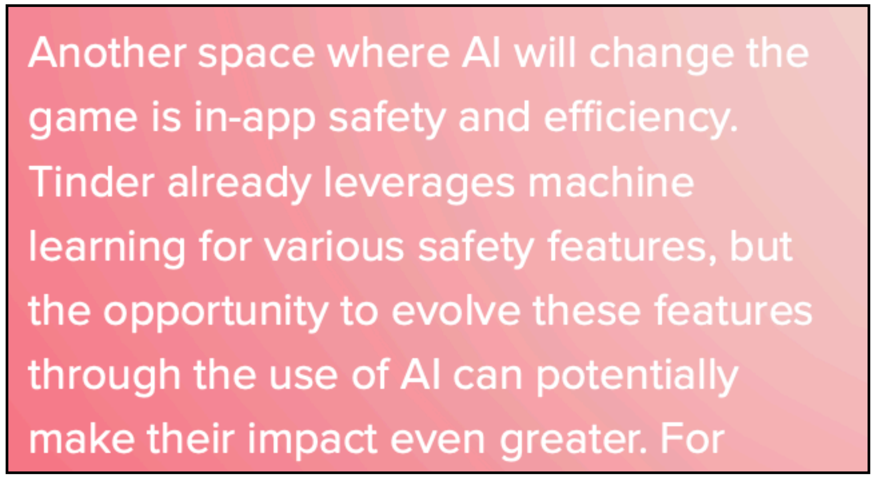
16.9K 14.8K 114K 43.8M



The Indian Express

A look at shadowban policies on Instagram & Facebook and what you can do to appeal unfair bans

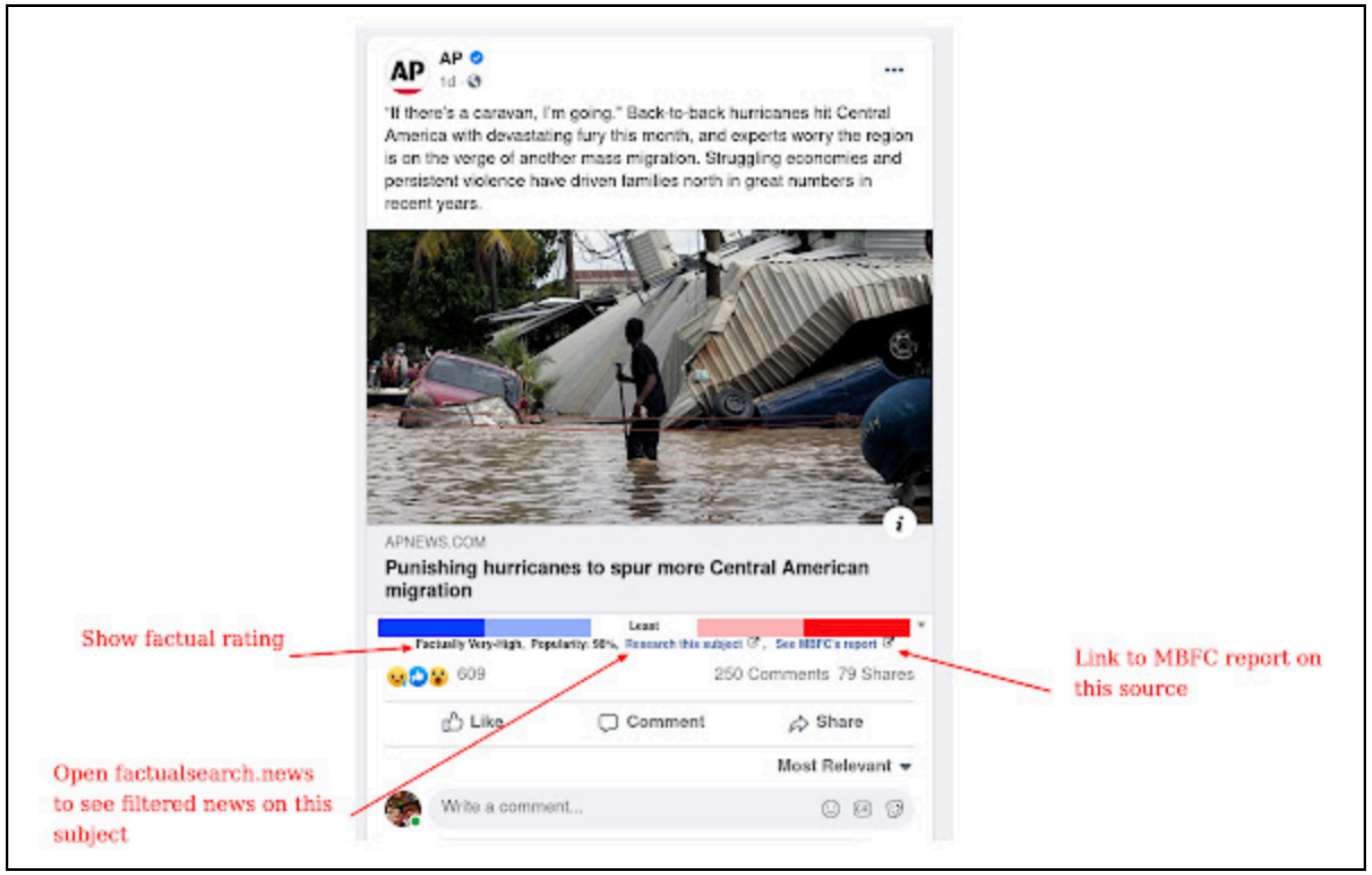
First, what is shadow banning? If you are a creator and have noticed a sudden, significant drop in engagement on your content then chances are...



Another space where AI will change the game is in-app safety and efficiency. Tinder already leverages machine learning for various safety features, but the opportunity to evolve these features through the use of AI can potentially make their impact even greater. For

The problem

- Measures by users examples:
  - Tools that spot fake news & bots
  - Turing tests for chats
  - Bot-helpers to manage group chats
  - Manual human verification



AP
1d

"If there's a caravan, I'm going." Back-to-back hurricanes hit Central America with devastating fury this month, and experts worry the region is on the verge of another mass migration. Struggling economies and persistent violence have driven families north in great numbers in recent years.

APNEWS.COM

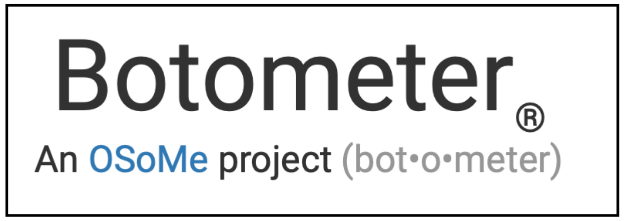
Punishing hurricanes to spur more Central American migration

Least
Factually Very-High, Popularity: 58%, Research this subject, See MBFC's report
609
250 Comments 79 Shares
Like
Comment
Share
Most Relevant
Write a comment...

Show factual rating

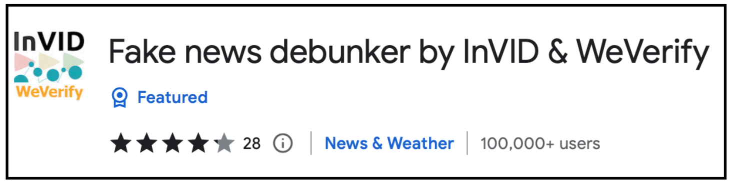
Link to MBFC report on this source

Open factualsearch.news to see filtered news on this subject

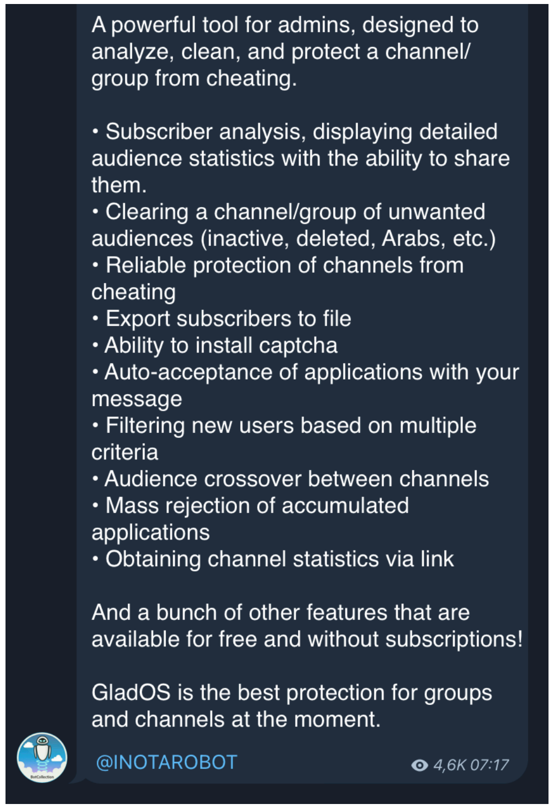


Botometer®

An OSoMe project (bot•o•meter)



Fake news debunker by InVID & WeVerify



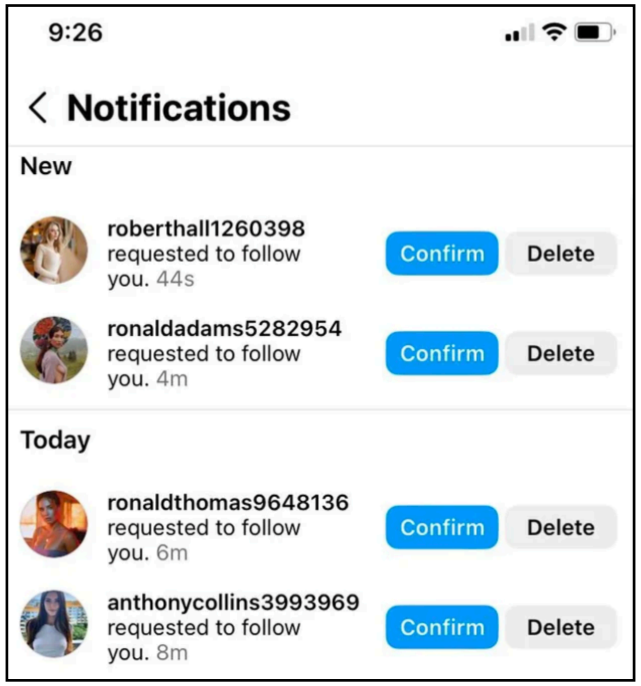
A powerful tool for admins, designed to analyze, clean, and protect a channel/group from cheating.

- Subscriber analysis, displaying detailed audience statistics with the ability to share them.
- Clearing a channel/group of unwanted audiences (inactive, deleted, Arabs, etc.)
- Reliable protection of channels from cheating
- Export subscribers to file
- Ability to install captcha
- Auto-acceptance of applications with your message
- Filtering new users based on multiple criteria
- Audience crossover between channels
- Mass rejection of accumulated applications
- Obtaining channel statistics via link

And a bunch of other features that are available for free and without subscriptions!

GladOS is the best protection for groups and channels at the moment.

@INOTAROBOT



Notifications

New

roberthall1260398 requested to follow you. 44s

ronaldadams5282954 requested to follow you. 4m

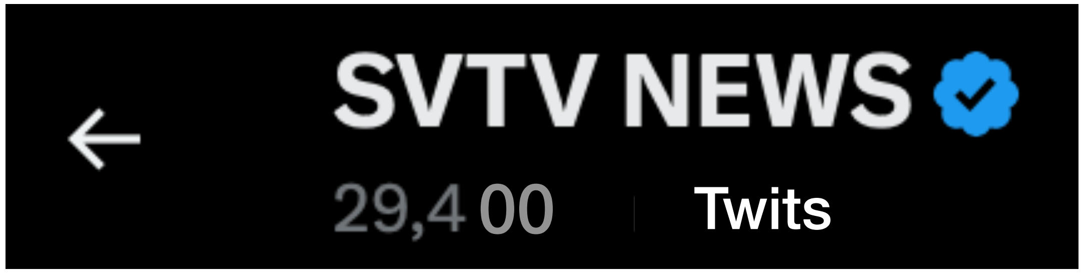
Today

ronaldthomas9648136 requested to follow you. 6m

anthonycollins3993969 requested to follow you. 8m

The problem

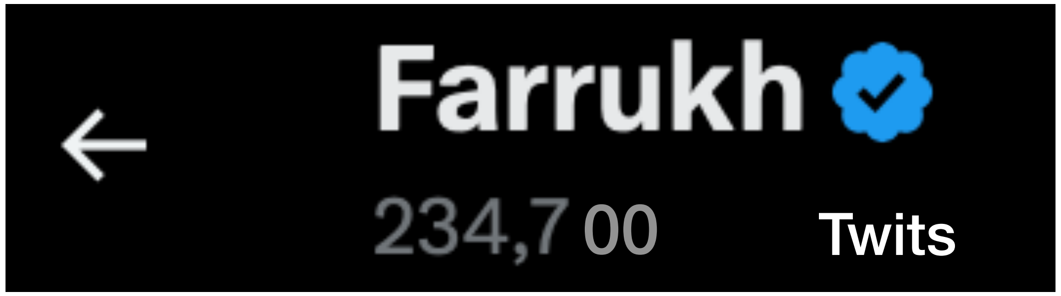
- For manual verification you need to read a lot of text - in fact, you need to scan previous activity of account



SVTV NEWS

29,4 00 Twits

Media outlet



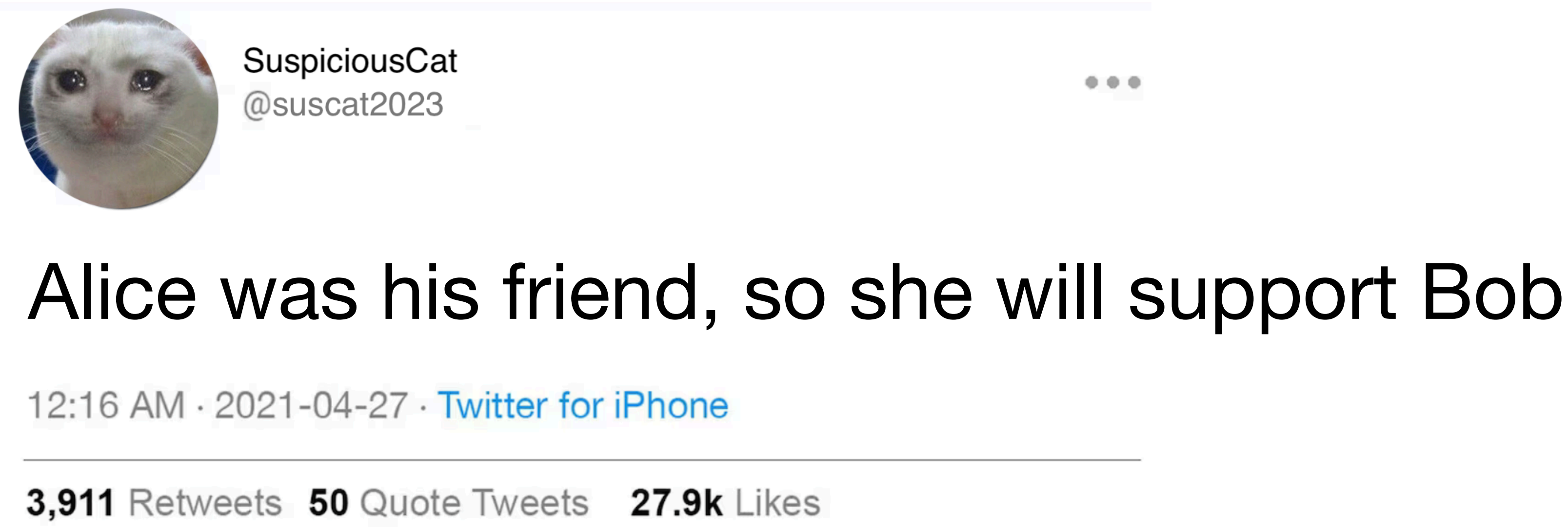
Farrukh

234,7 00 Twits

User

- For many accounts **you just can not do it manually**

ARElight

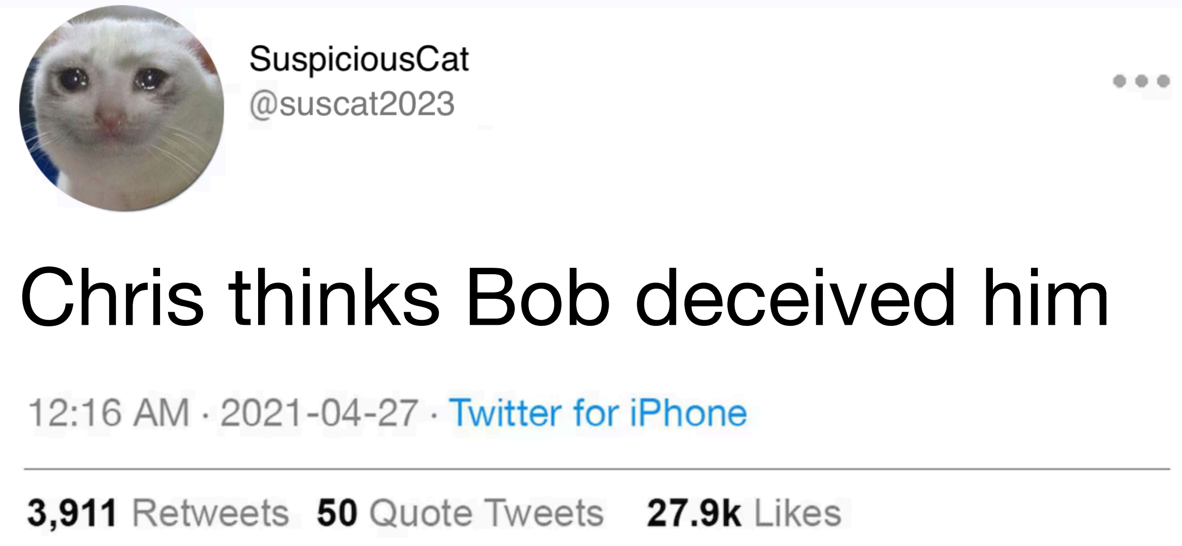


SuspiciousCat
@suscat2023

Alice was his friend, so she will support Bob

12:16 AM · 2021-04-27 · Twitter for iPhone

3,911 Retweets **50** Quote Tweets **27.9k** Likes

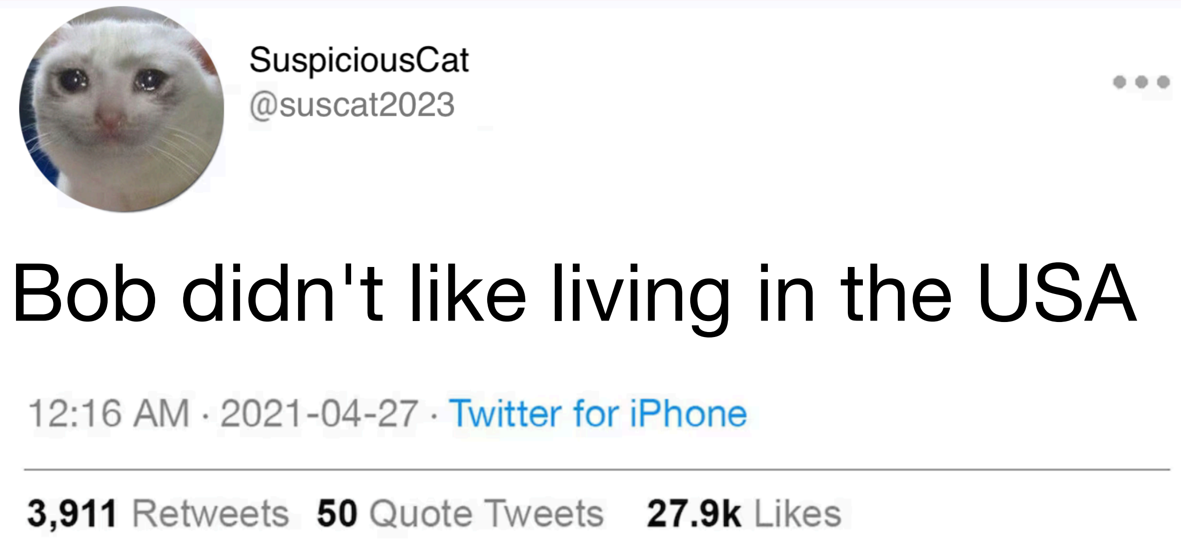


SuspiciousCat
@suscat2023

Chris thinks Bob deceived him

12:16 AM · 2021-04-27 · Twitter for iPhone

3,911 Retweets **50** Quote Tweets **27.9k** Likes



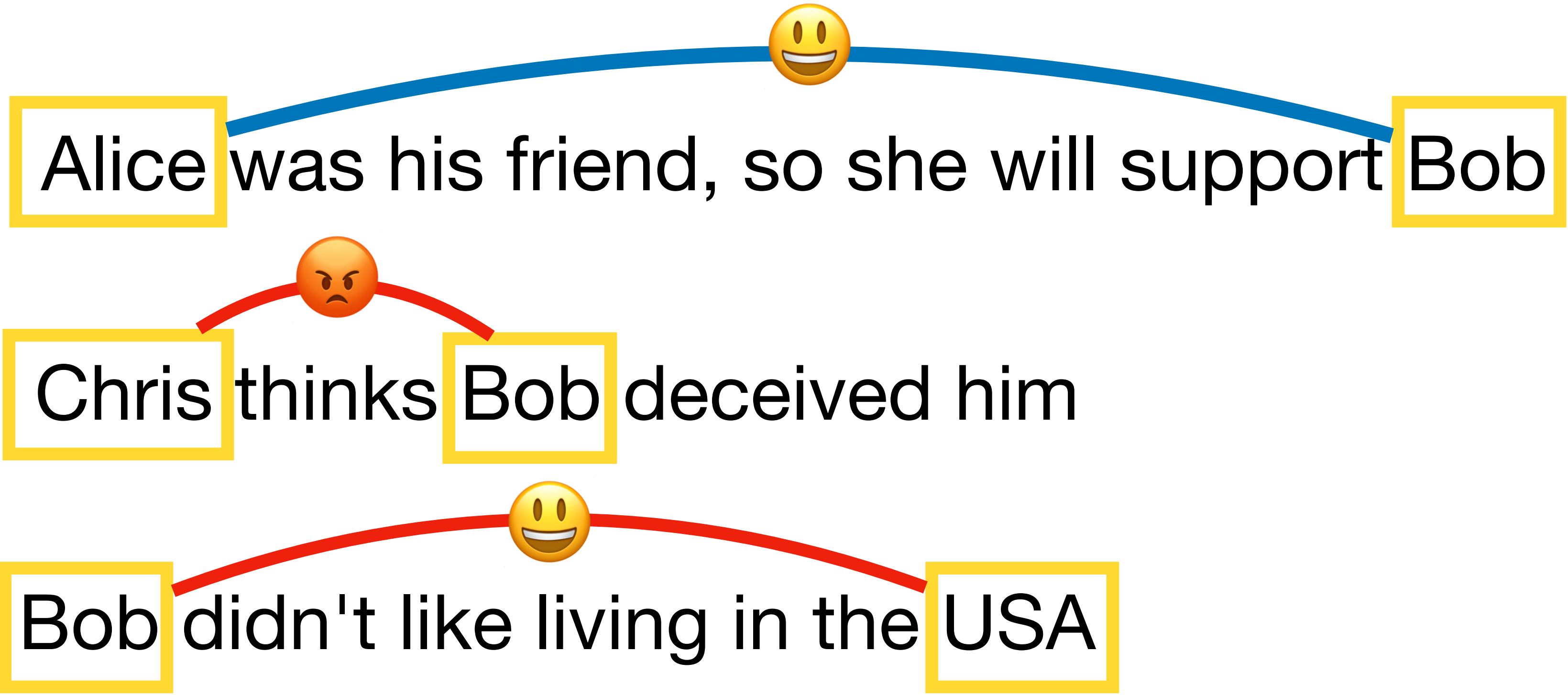
SuspiciousCat
@suscat2023

Bob didn't like living in the USA

12:16 AM · 2021-04-27 · Twitter for iPhone

3,911 Retweets **50** Quote Tweets **27.9k** Likes

ARElight



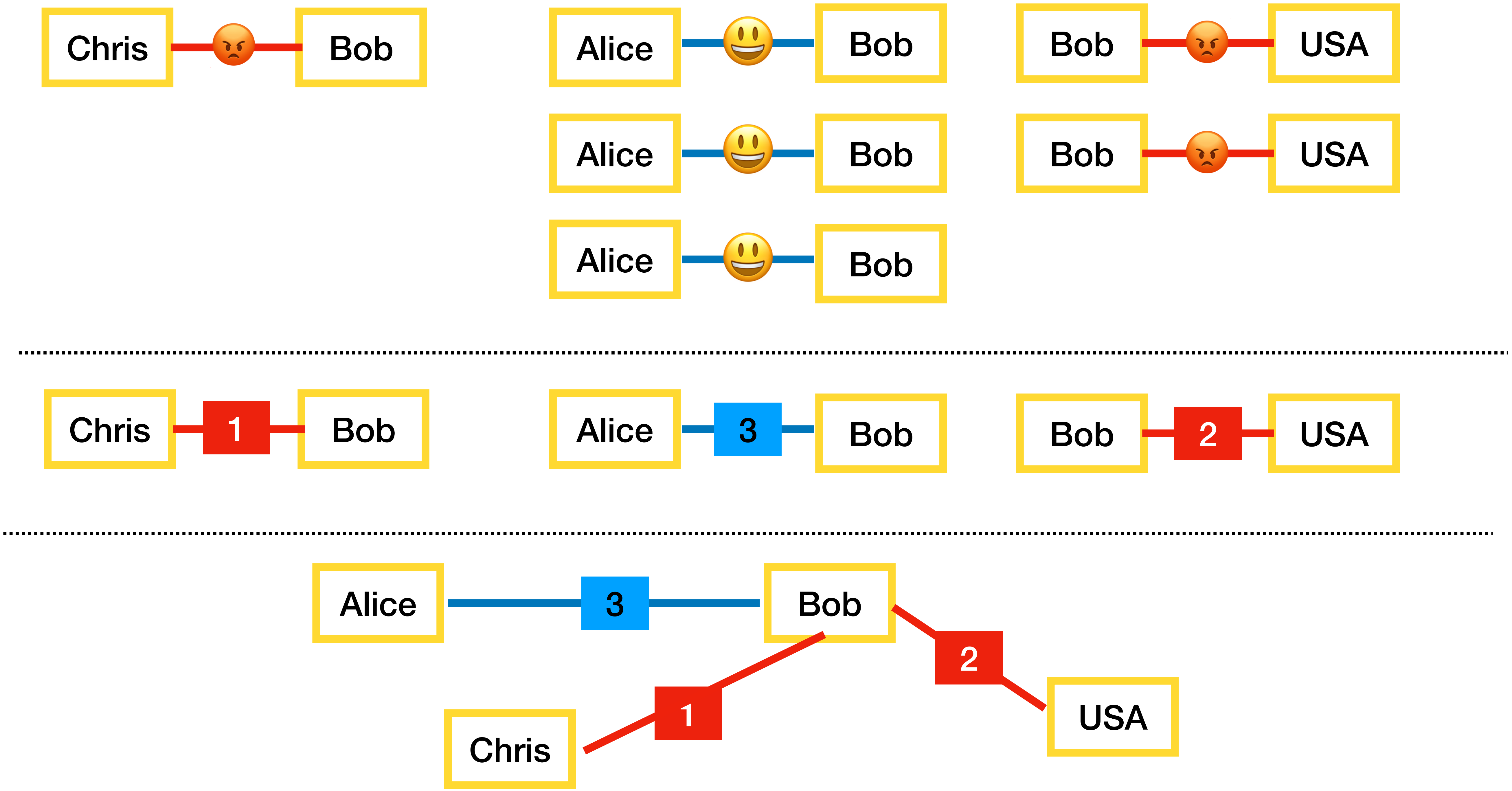
Alice was his friend, so she will support Bob

Chris thinks Bob deceived him

Bob didn't like living in the USA

- ARElight use AI to identify instances in text such as: Person, Organization, Location, Date
- Within small amount of text, identify relations and their sentiment (positive 😃, negative 😡, neutral 😐)

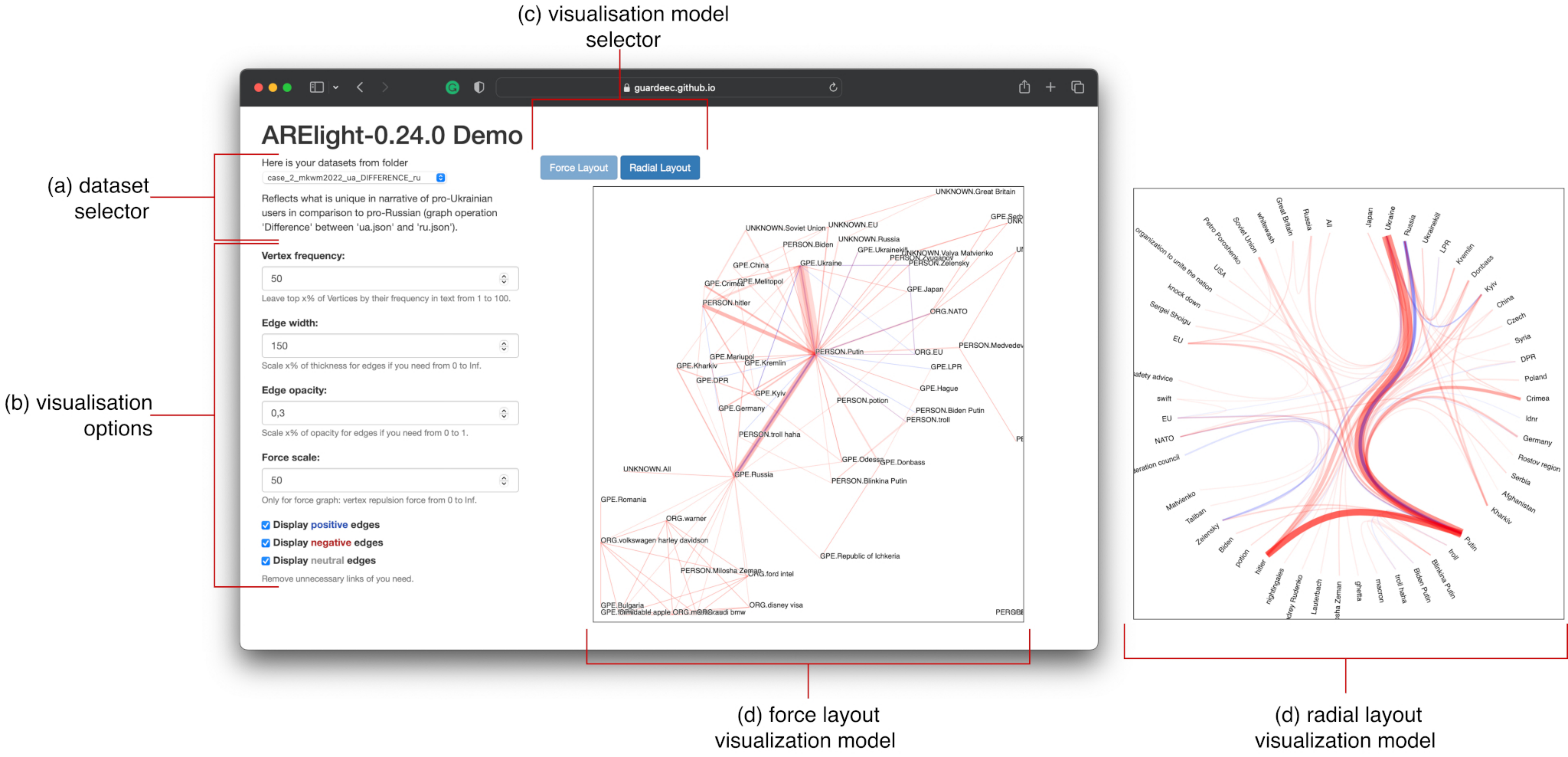
ARElight



- Small texts united into a single network, that represent the narrative of text

ARElight

- Demo: https://guardeec.github.io/arelight_demo/template.html



(c) visualisation model selector

(a) dataset selector

(b) visualisation options

(d) force layout visualization model

(d) radial layout visualization model

guardeec.github.io

ARElight-0.24.0 Demo

Here is your datasets from folder

case_2_mkwm2022_ua_DIFFERENCE_ru

Reflects what is unique in narrative of pro-Ukrainian users in comparison to pro-Russian (graph operation 'Difference' between 'ua.json' and 'ru.json').

Vertex frequency:

50

Leave top x% of Vertices by their frequency in text from 1 to 100.

Edge width:

150

Scale x% of thickness for edges if you need from 0 to Inf.

Edge opacity:

0,3

Scale x% of opacity for edges if you need from 0 to 1.

Force scale:

50

Only for force graph: vertex repulsion force from 0 to Inf.

☑ Display positive edges

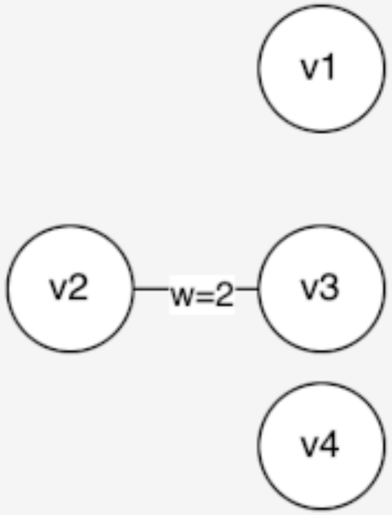
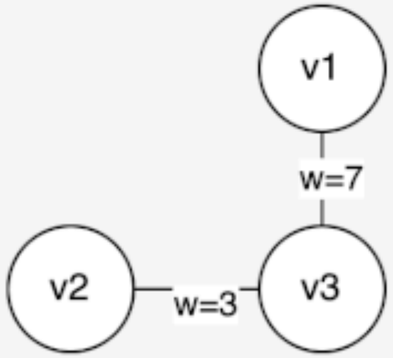
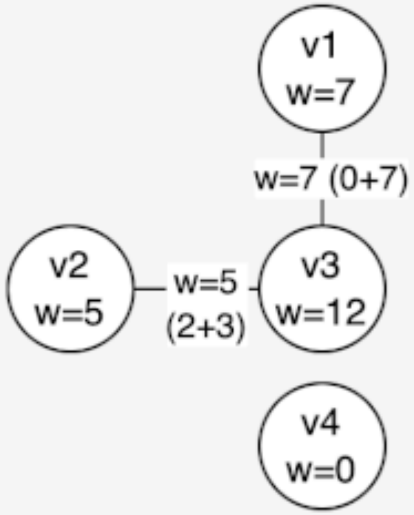
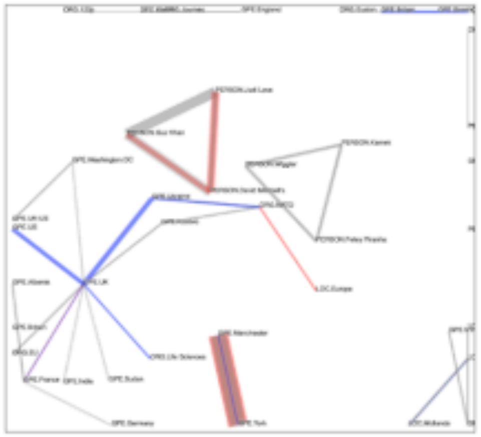
☑ Display negative edges

☑ Display neutral edges

Remove unnecessary links of you need.

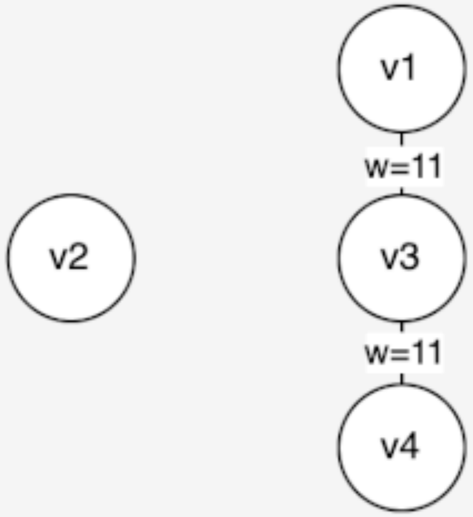
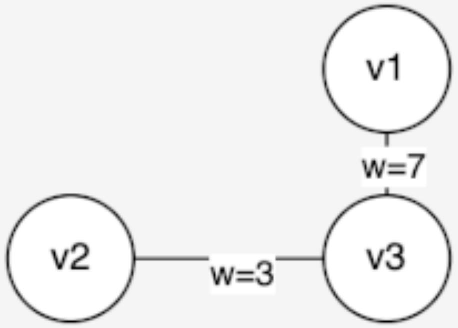
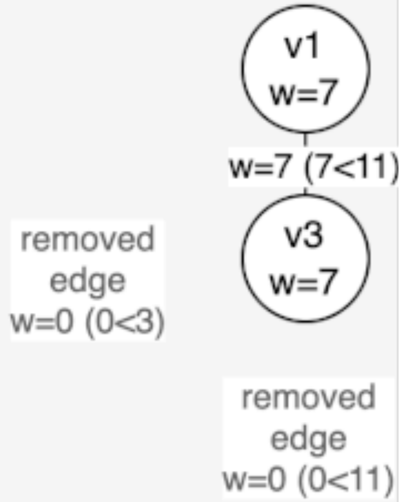
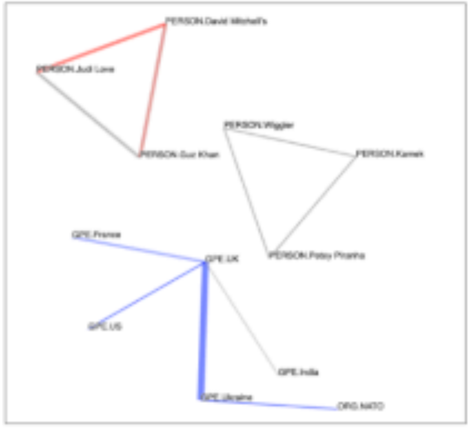
ARElight

- Operations - UNION
- Merge multiple texts' graphs together

	G_1		G_2		G
Union $W_e=W_{e1}+W_{e2}$ $Wv=\Sigma W_e(e)$ Summarizing multiple graphs, e.g. making graph of UK politicians from individual Twitter accounts		∪		=	 v1 w=7 w=7 (0+7) v2 w=5 w=5 (2+3) v3 w=12 v4 w=0
Example Create graph of **multiple politicians** using Twits of Boris Johnson and Rishi Sunak	 Boris	∪	 Rishi	=	 Boris & Rishi

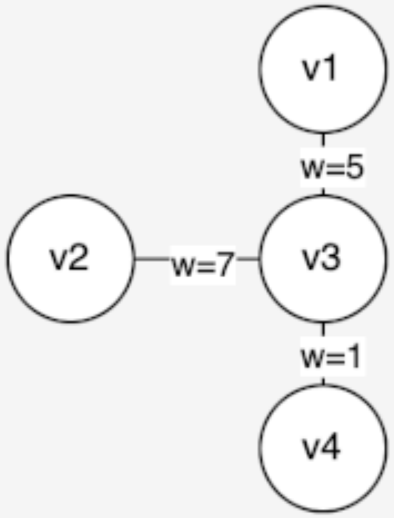
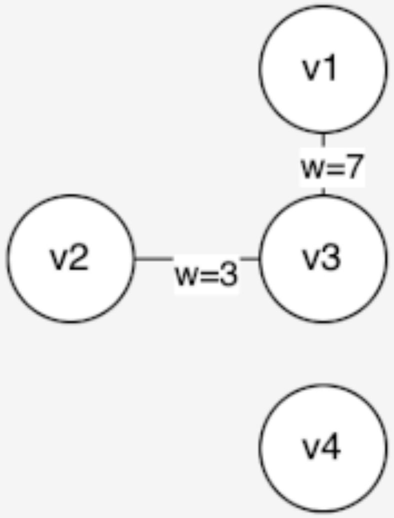
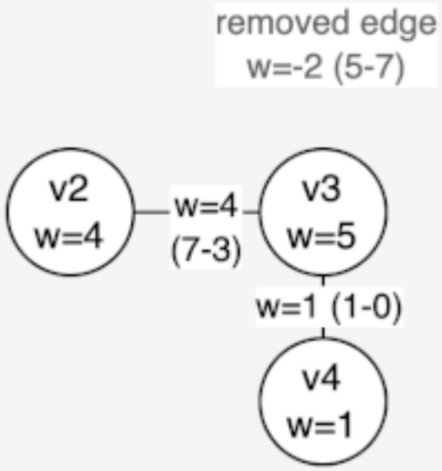
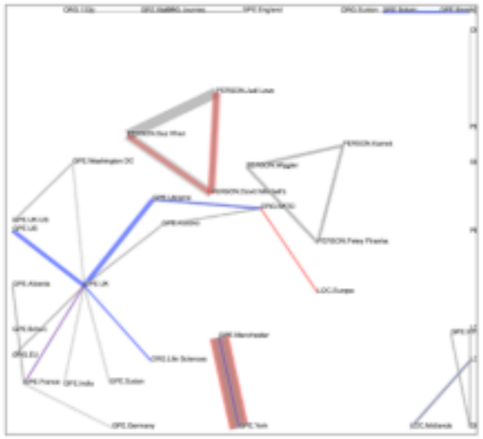
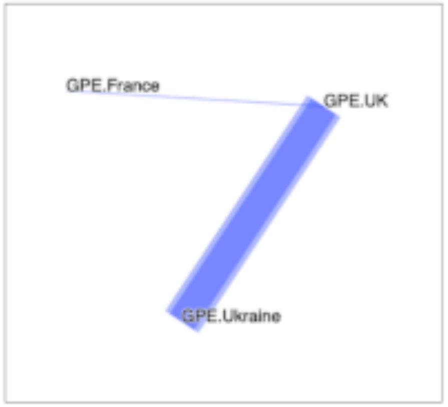
ARElight

- Operations - INTERSECTION
- What is similar between 2 texts?

	G_1		G_2		G
Similarity W_e=min(W_{e1},W_{e2}) Wv=$\Sigma W_e(e)$ Find similarity between graphs, e.g. common between graphs of Twits among 2 accounts		∩		=	 v1 w=7 w=7 (7<11) removed edge w=0 (0<3) v3 w=7 removed edge w=0 (0<11)
Example Create graph that describe **what is similar** between Twits of Boris Johnson and Rishi Sunak	 Boris	∩	 Rishi	=	 Their similarity

AREligt

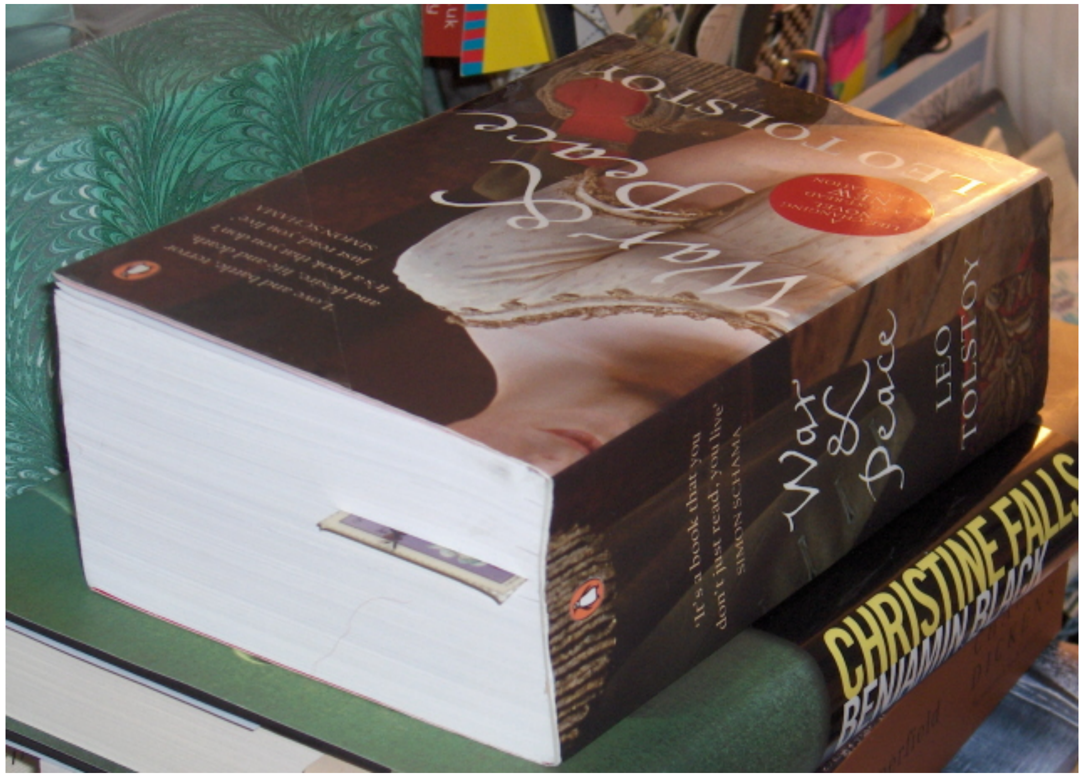
- Operations - DIFFERENCE (not commutative)
- What is unique in one text in comparison to another?

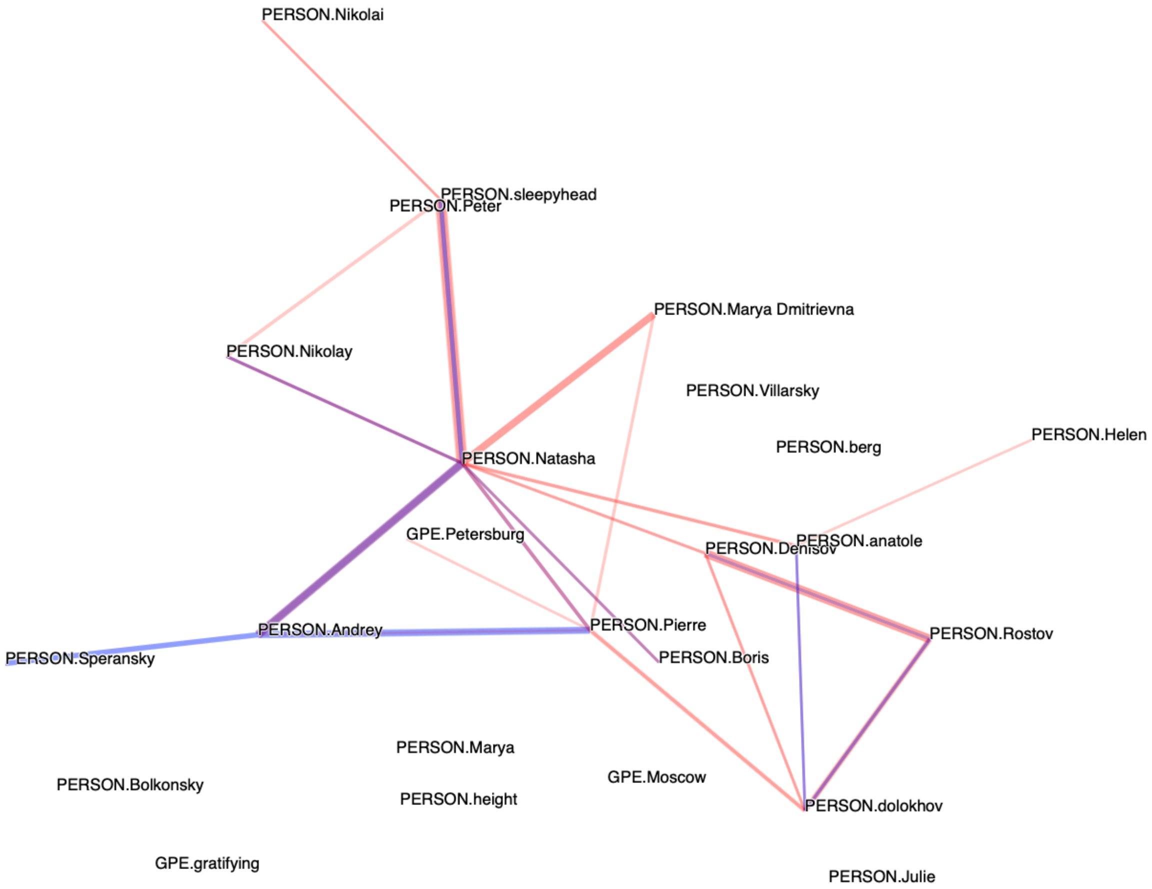
	G_1		G_2		G
Difference $W_e=W_{e1}-W_{e2}$ $Wv=\Sigma W_e(e)$ Find difference between graphs, e.g. what is unique in Twits of user$_1$ in comparison to user$_2$		-		=	 removed edge w=-2 (5-7) v2 w=4 w=4 (7-3) v3 w=5 w=1 (1-0) v4 w=1
Example Create graph that describe **what is unique** in Twits of Boris Johnson in comparison to Rishi Sunak	 Boris	-	 Rishi	=	 Their difference

Demo

- Examples:

- War & Peace book

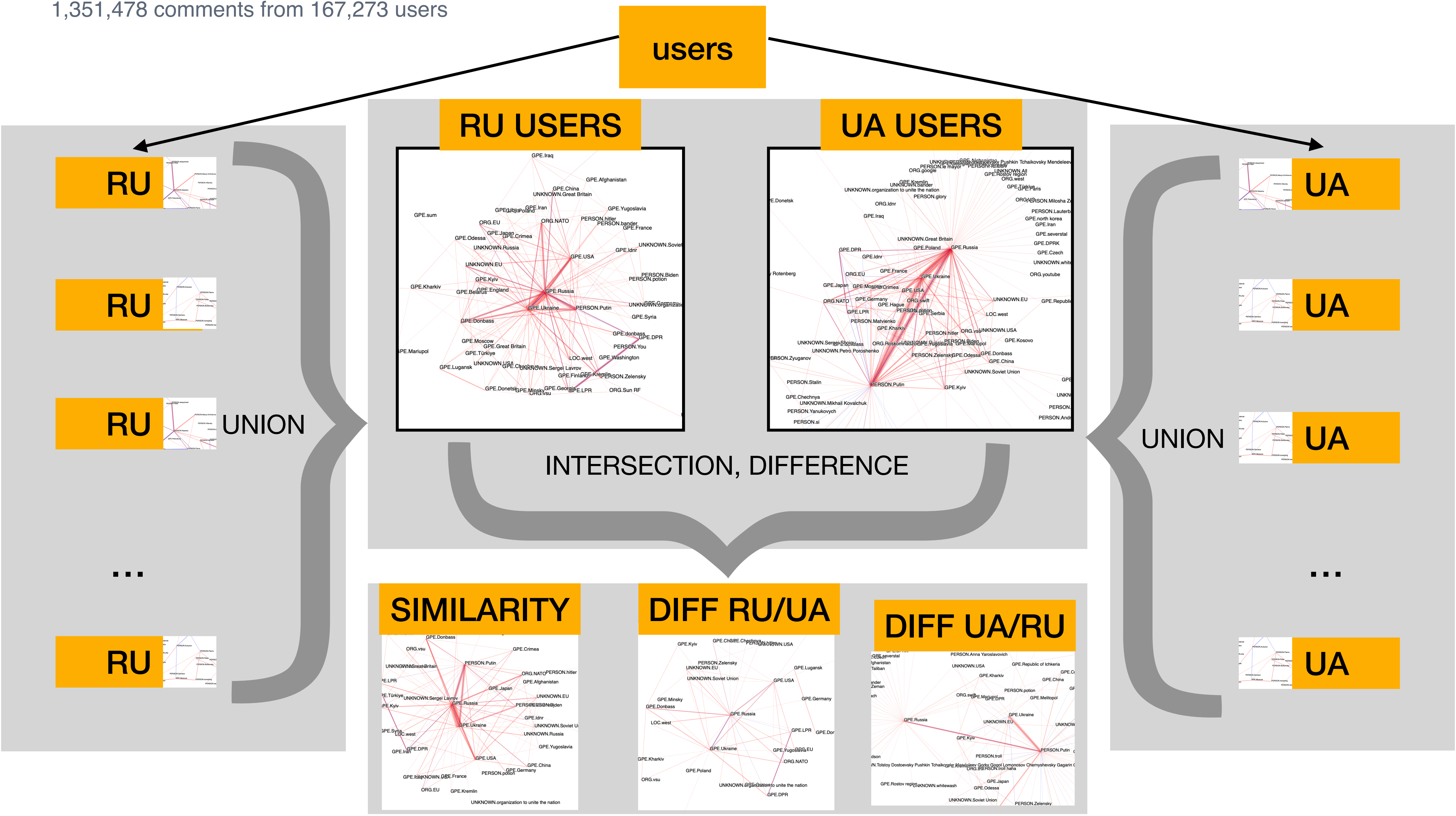




Demo

- Examples: RU/UA comments

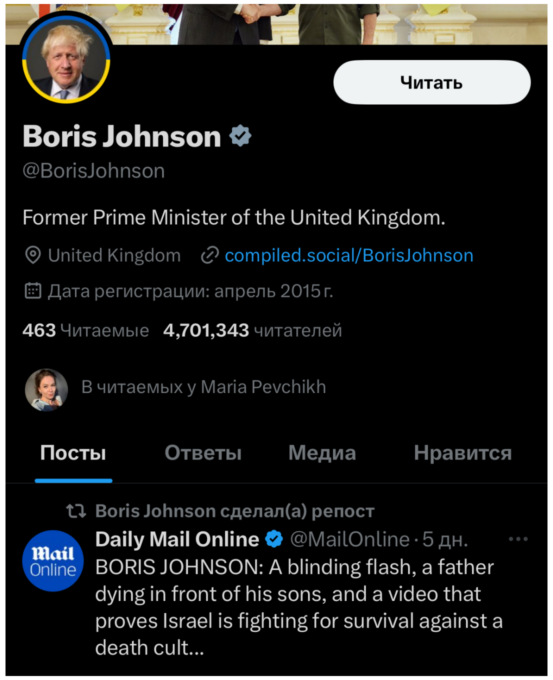
1,351,478 comments from 167,273 users



Demo

- Examples: Boris Johnson VS Rishi Sunak

1000 last Twits



Читать

Boris Johnson

@BorisJohnson

Former Prime Minister of the United Kingdom.

United Kingdom compiled.social/BorisJohnson

Дата регистрации: апрель 2015 г.

463 Читаемые **4,701,343** читателей

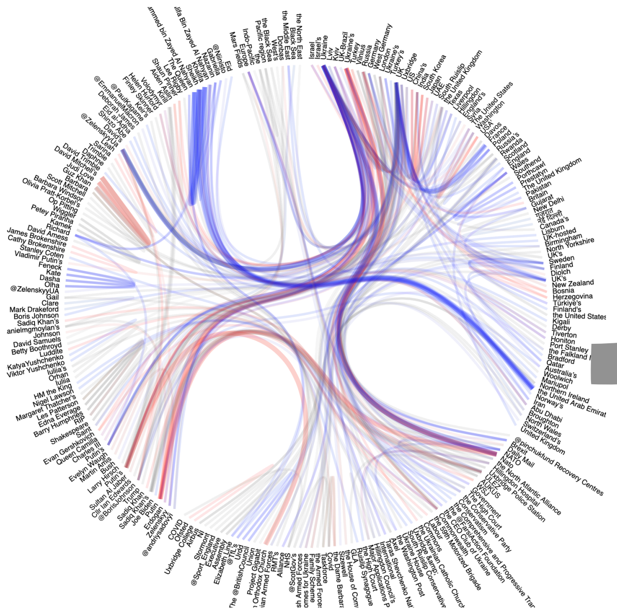
В читаемых у Maria Pevchikh

Посты Ответы Медиа Нравится

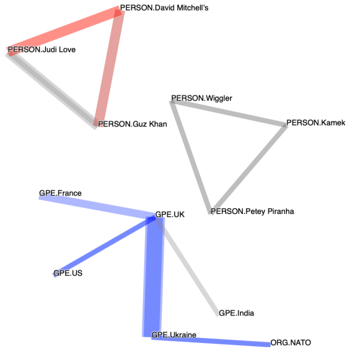
Boris Johnson сделал(а) репост

Daily Mail Online @MailOnline · 5 дн.

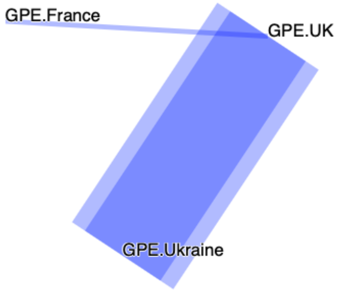
BORIS JOHNSON: A blinding flash, a father dying in front of his sons, and a video that proves Israel is fighting for survival against a death cult...



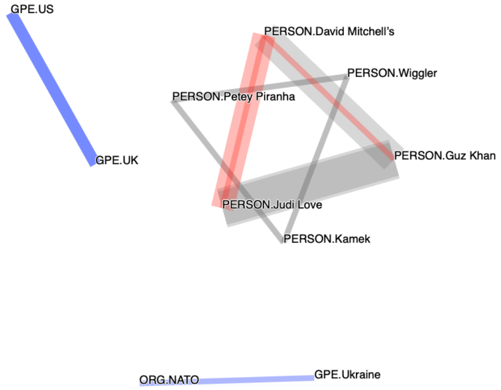
INTERSECTION (What is similar?)

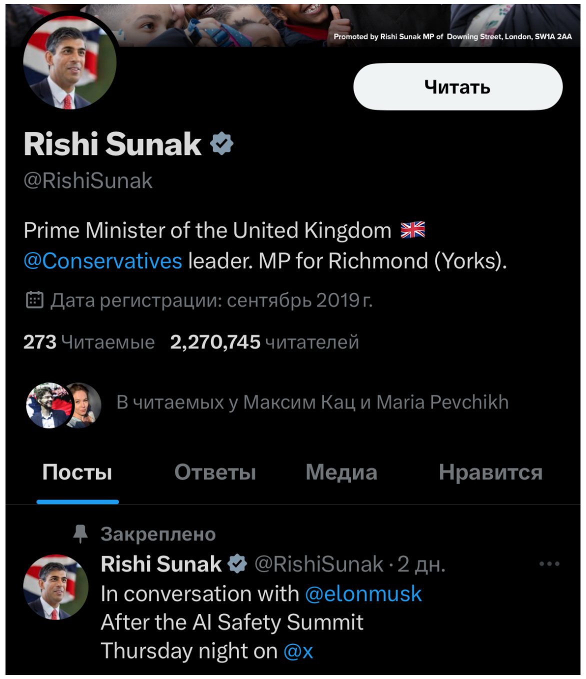


DIFFERENCE (What is unique in Boris?)



DIFFERENCE (What is unique in Rishi?)





Promoted by Rishi Sunak MP of Downing Street, London, SW1A 2AA

Читать

Rishi Sunak

@RishiSunak

Prime Minister of the United Kingdom 🇬🇧

@Conservatives leader. MP for Richmond (Yorks).

Дата регистрации: сентябрь 2019 г.

273 Читаемые **2,270,745** читателей

В читаемых у Максим Кац и Maria Pevchikh

Посты Ответы Медиа Нравится

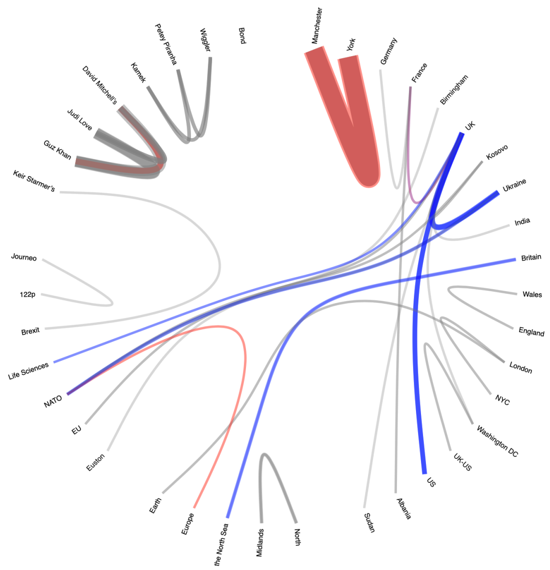
Закреплено

Rishi Sunak @RishiSunak · 2 дн.

In conversation with @elonmusk

After the AI Safety Summit

Thursday night on @x



Overview

- ARElight **helps to extract/analyse narrative** and represent it in understandable manner for:
    - ➡users - for accounts verification and media checks
    - ➡researchers - extract/compare narratives of accounts/disinformation
    - ➡STEM - present texts in form of graphs for ML/automation/etc.
- ARElight is a tool in several versions:
    - ➡Python Library & Demo [complete] - for STEM
    - ➡Web Service [planned] - for researchers
    - ➡Twitter/X browser extension [planned] - for users
- We are ready to help if you would like to use it in your research.

- Python Library: https://github.com/nicolay-r/ARElight/
- Demo: https://guardeec.github.io/arelight_demo/template.html

Nikolay R. (rusnicolay@gmail.com), **Maxim K.** (maksim.kalameyets@newcastle.ac.uk)