



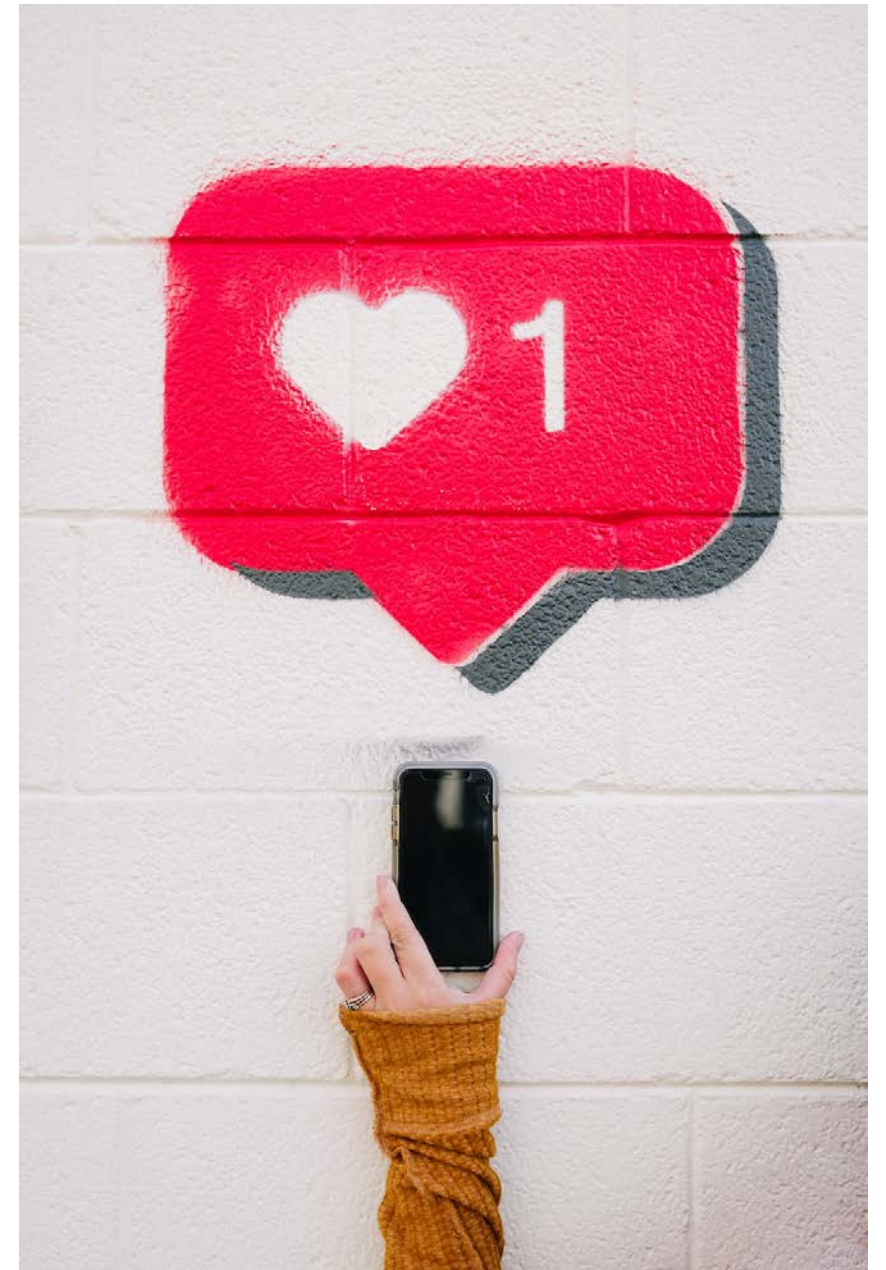
Telegram as an Alternative Platform for Marginalised Groups: Identification of Trending Topics

Justina MANDRAVICKAITĖ, Eglė KANKEVIČIŪTĖ, Bohdan ZHYHUN,
Danguolė KALINAUSKAITĖ and **Tomas KRILAVIČIUS**

- What are people talking about?
- What is public opinion?
- What are trending topics?

- Different channels: mainstream media, social media, ...
- Telegram – channel of the choice

- Tools
 - AI and language technologies
 - Visualization



Motivation

- Telegram
- 6 most active Lithuanian groups and news channels
 - Diskusijos ir Naujienos @Patriotai (Discussions and News @Patriots)
 - Geoengineering @Patriotai
 - Naujienos Lietuvoje (News in Lithuania)
 - Savas kampas LT (One's own corner LT)
 - Slepamos technologijos @Patriotai (Hidden Technologies @Patriots)
 - Žmonių žiniasklaida (People Media)
- ~10 000 messages from each channel (mixed time frames over 2020–2022) were extracted using Telethon
- Data from messages – only text data (message text and comments), all metadata was discarded
- Preprocessing: non-text data (HTML) removed, whitespaces removed, lowercased
- Lemmatization for keyword extraction and topic modelling



ous Message #1
...te, sveikiname prisijungus prie @Patriotai Sle...

1K Elusive Truth, 10:20

August 29

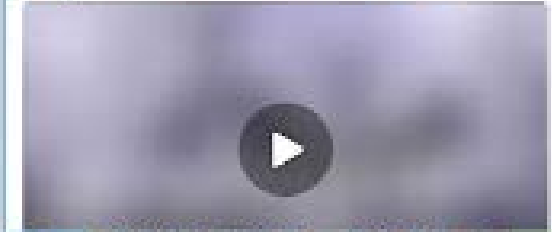
Nieko Nesakysiu
Deleted Account
O kaip cia susije su grupes tema? 5g ne...
5G technologija nera slepiama-taip, taciau yra slepiamas sios technologijos poveikis gyviems organizmams ir tikrasis sios technologijos tikslas. Tai tikrai nera daroma vardan greitesnio intiko jusu mobiliake....

5
08:28

Deleted Account
Nieko Nesakysiu
5G technologija nera slepiama-taip, taci...
Taip, sutinku, taciau jei sioj grupej vel bus keliami informacija apie skiepu poveiki, 5G poveiki tai jau nukryps nuo pvz Teslos isradimu ar free energy temos. Juk yra specialios grupes apie 5G poveiki temas, taip susiorientuojant kas prie ko🤔

Deleted Account
Nikolos Teslos isradimai, pakeite pasauli! Kuro neturintis generatorius, eteris ir nemokama laisvoji energija

zen.yandex.ru/id/622627b4cf9e.....



Nikola_Tesla_-_The_Fa
ntastic_Inventions...
20.3 MB PDF



it anyways.
Money does not represent such a value as men have placed upon it.

The Inventions Researches and Writing of Nikola Tesla.pdf

Blogai indoktrinuotas

Message

Pinned Message #3540
Pasipildzius @Patriotai moderatorių komandai - ati...

Delfis dar irgi neskambija 17:28

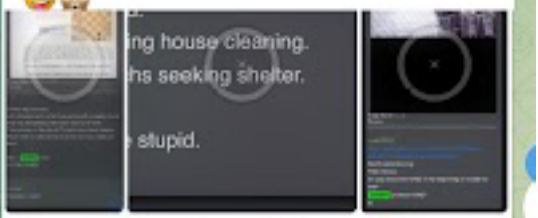
Light in me
Idomu kaip ten bus su testinumu. Nes jei niekas neprims sosto vadinasi visos okupuotos zemes defakto atgauna nepriklausomybe ir suverenuma

17:29

Deleted Account
Light in me
Photo

Sito nera Twitteryje. Cia tavo screenshot ar kazkur radai? 17:30

Light in me
Deleted Account
Delfis dar irgi neskambija
Reiktu pagal gematrija patikrint kokie skaiciai gaunasi. Zinant kad jie uzsiima okultizmu kurio net patys neslepia. Tai manau idomiu skaiciuku turetume pamatyt



Queen News Coming?

patraukta i priedanga ir gyvena sau nematoma. Kaip manai, galimas toks scenarijus?

17:33

Light in me

BBC REPTO QUEEN
BBC News



Queen under medical supervision at Balmoral
Prince Charles is there with his son Prince William and other senior royals.

INSTANT VIEW

Message

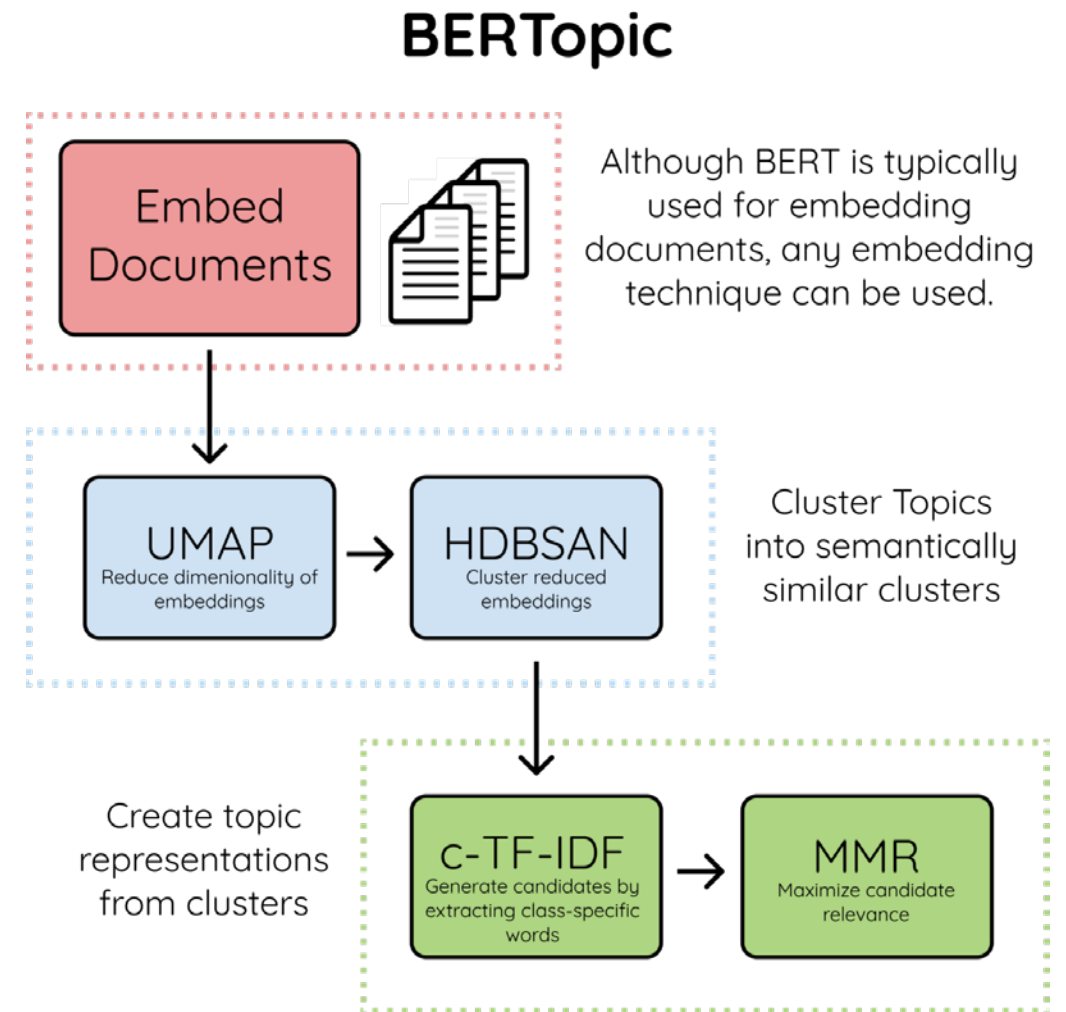
- Preprocessing: tokenization
- Feature extraction: a casing (uppercase vs lowercase), word position, frequency, relatedness to context, sentence – frequency
- Term weighting – a combination of features
- Candidate keywords generation – **word** and/or phrase



<https://towardsdatascience.com/keyword-extraction-methods-the-overview-35557350f8bb>

Methods: YAKE – Yet Another Keyword Extractor

- Class-based term frequency-inverse document frequency (TF-IDF) + clustering
- Embeddings using pre-trained language models
- Dimensionality reduction (document embeddings)
- Clustering – groups of semantically related texts
- Class-based version of TF-IDF to extract topic representation



<https://medium.com/geekculture/nlp-tutorial-topic-modeling-in-python-with-bertopic-da760e1d03aa>

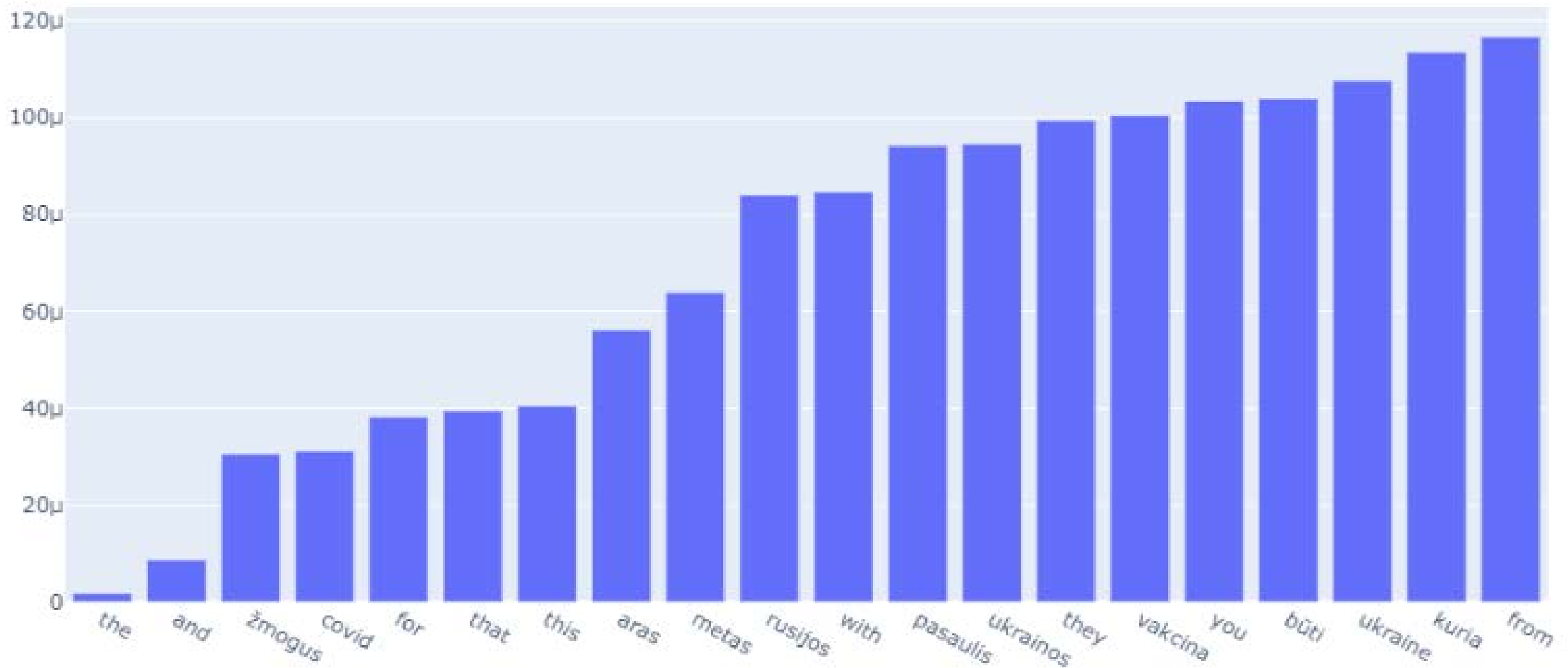
Methods: BERTopic for Topic Modelling

- Polyglot toolkit
- Polarity lexicon for Lithuanian
- Context (modifiers, negations) is ignored
- Basic tool

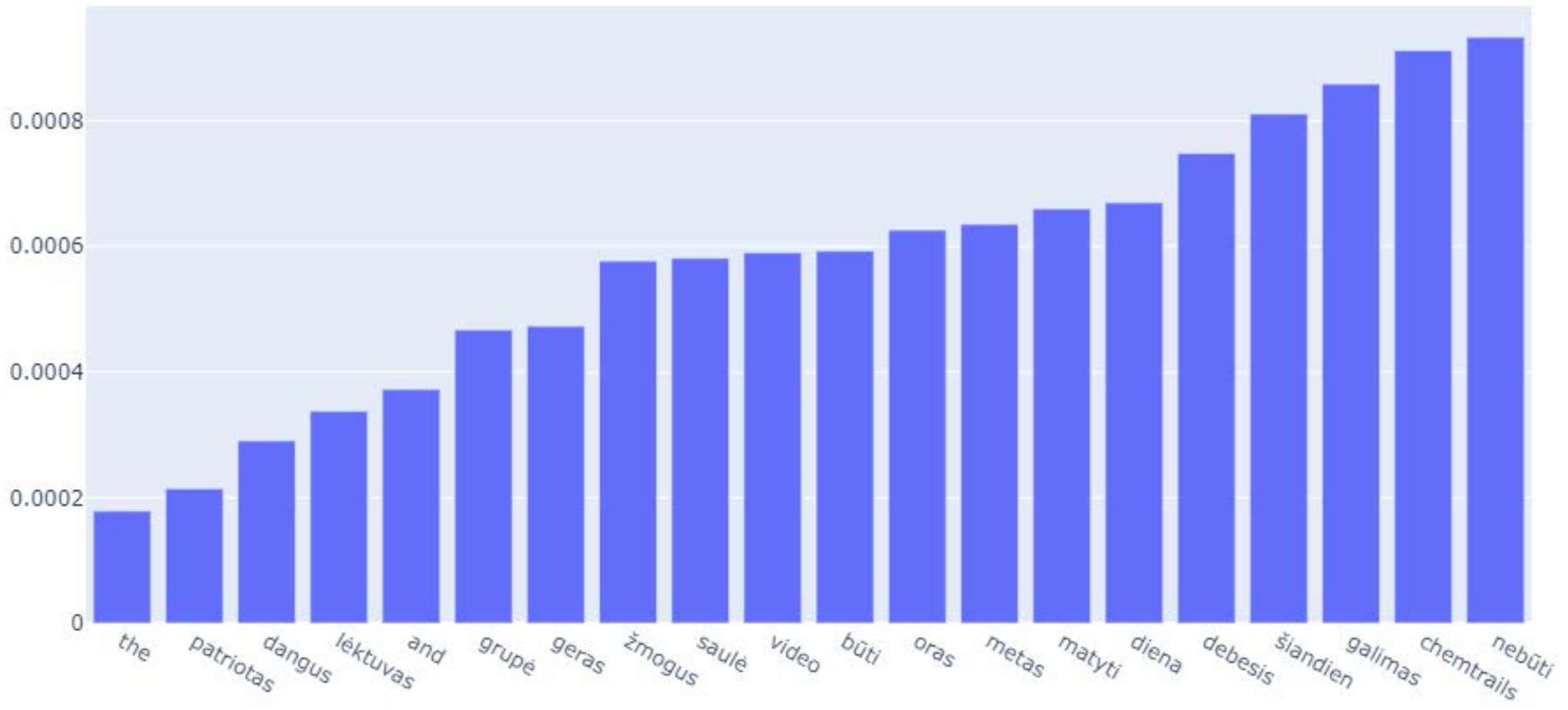


Results

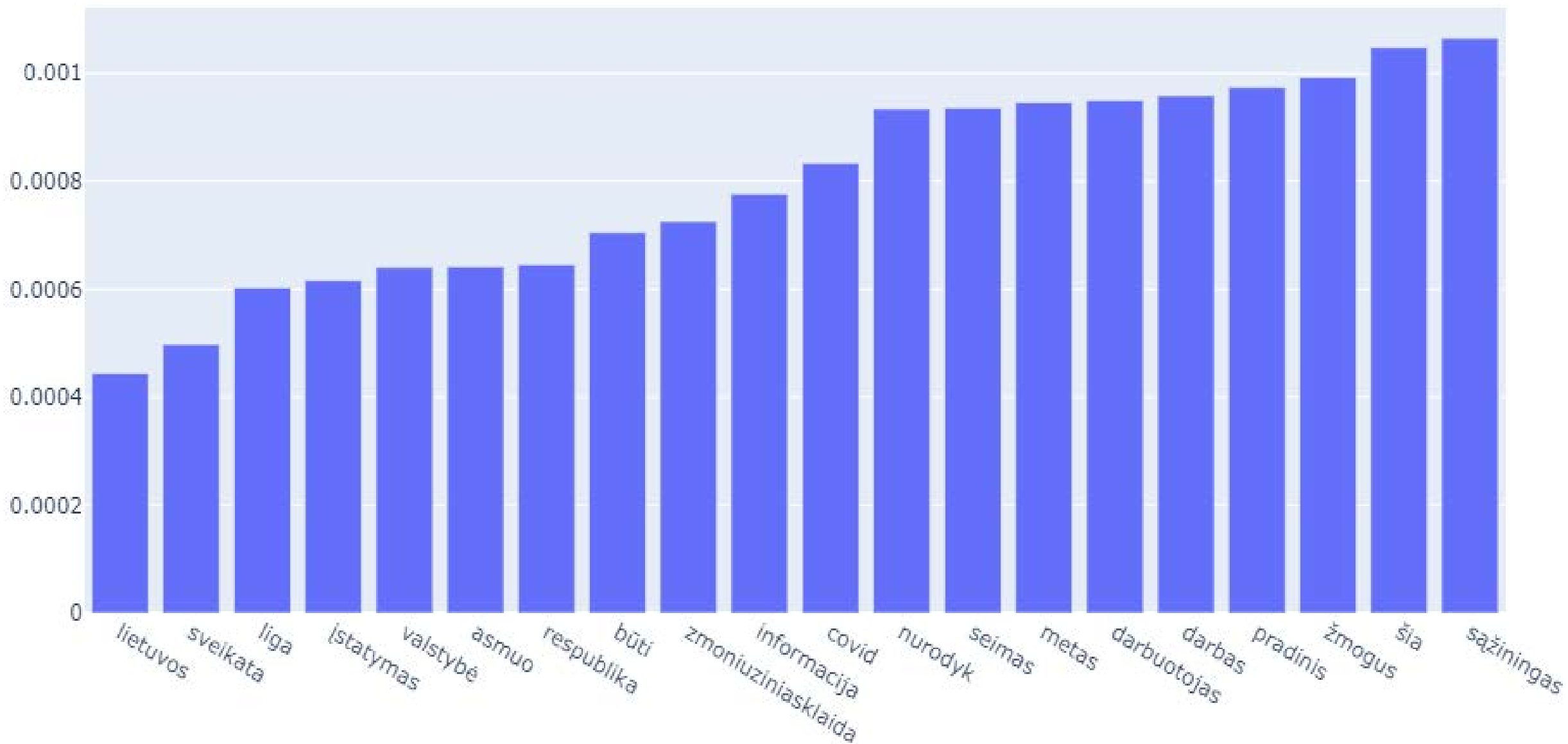




Keyword Extraction: Diskusijos ir naujienos @Patriotai, July of 2020, April of 2022



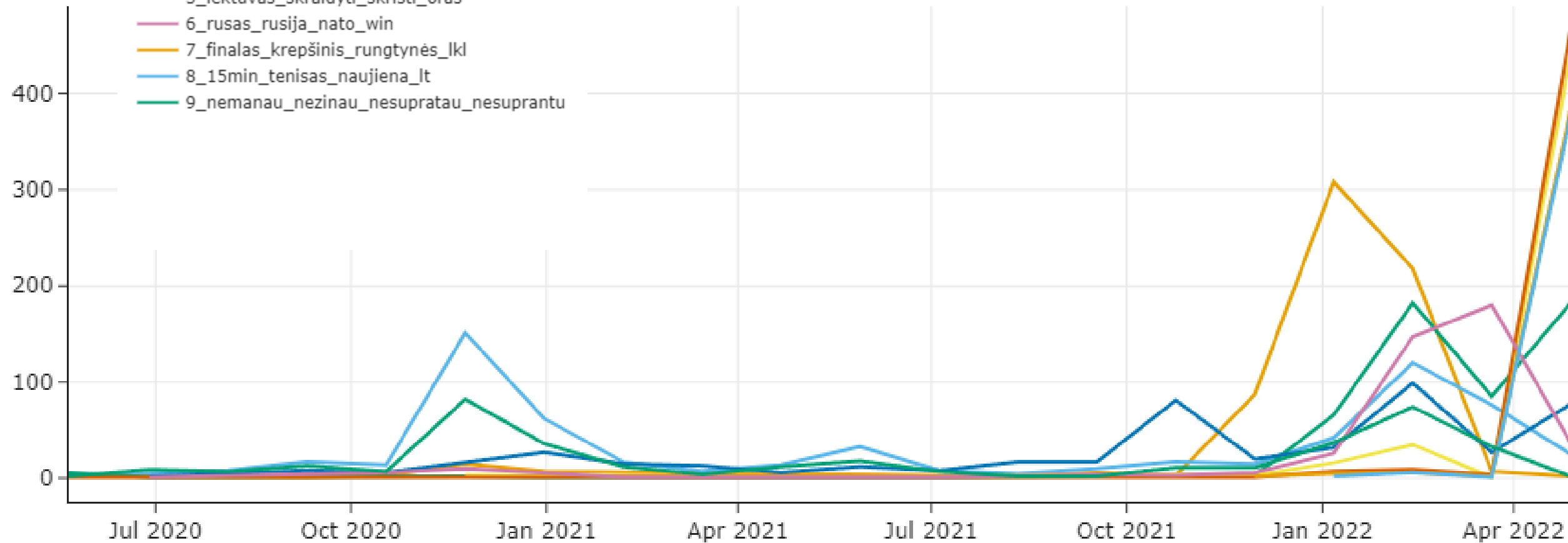
Keyword Extraction: Geoengineering @Patriotai, July of 2020, April of 2022



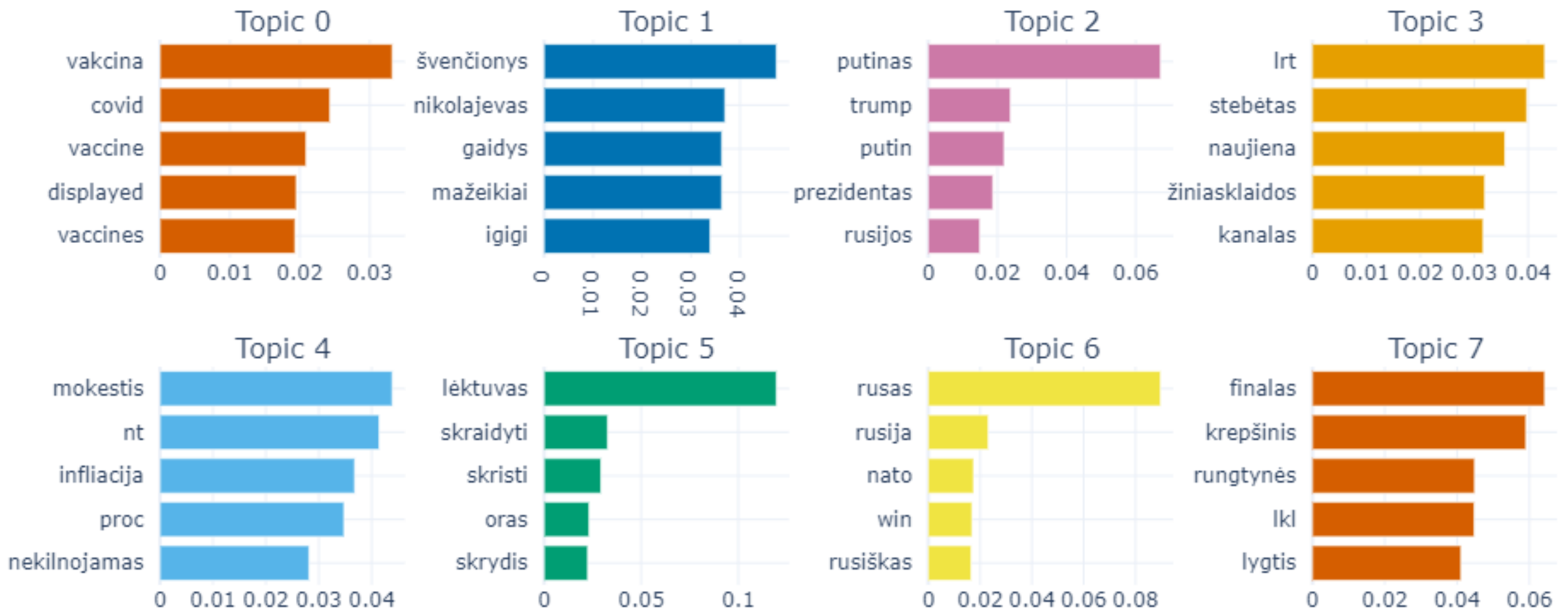
Keyword Extraction: Žmonių žiniasklaida @Patriotai, July of 2020, April of 2022

Global Topic Representation

- 0_vakcina_covid_vaccine_displayed
- 1_švenčionys_nikolajevas_gaidys_mažeikia...
- 2_putinas_trump_putin_prezidentas
- 3_lrt_stebėtas_naujiena_žiniasklaidos
- 4_mokestis_nt_infliacija_proc
- 5_lėktuvas_skraidyti_skristi_oras
- 6_rusas_rusija_nato_win
- 7_finalas_krepšinis_rungtynės_lkl
- 8_15min_tenisas_naujiena_lt
- 9_nemanau_nezinau_nesupratau_nesuprantu

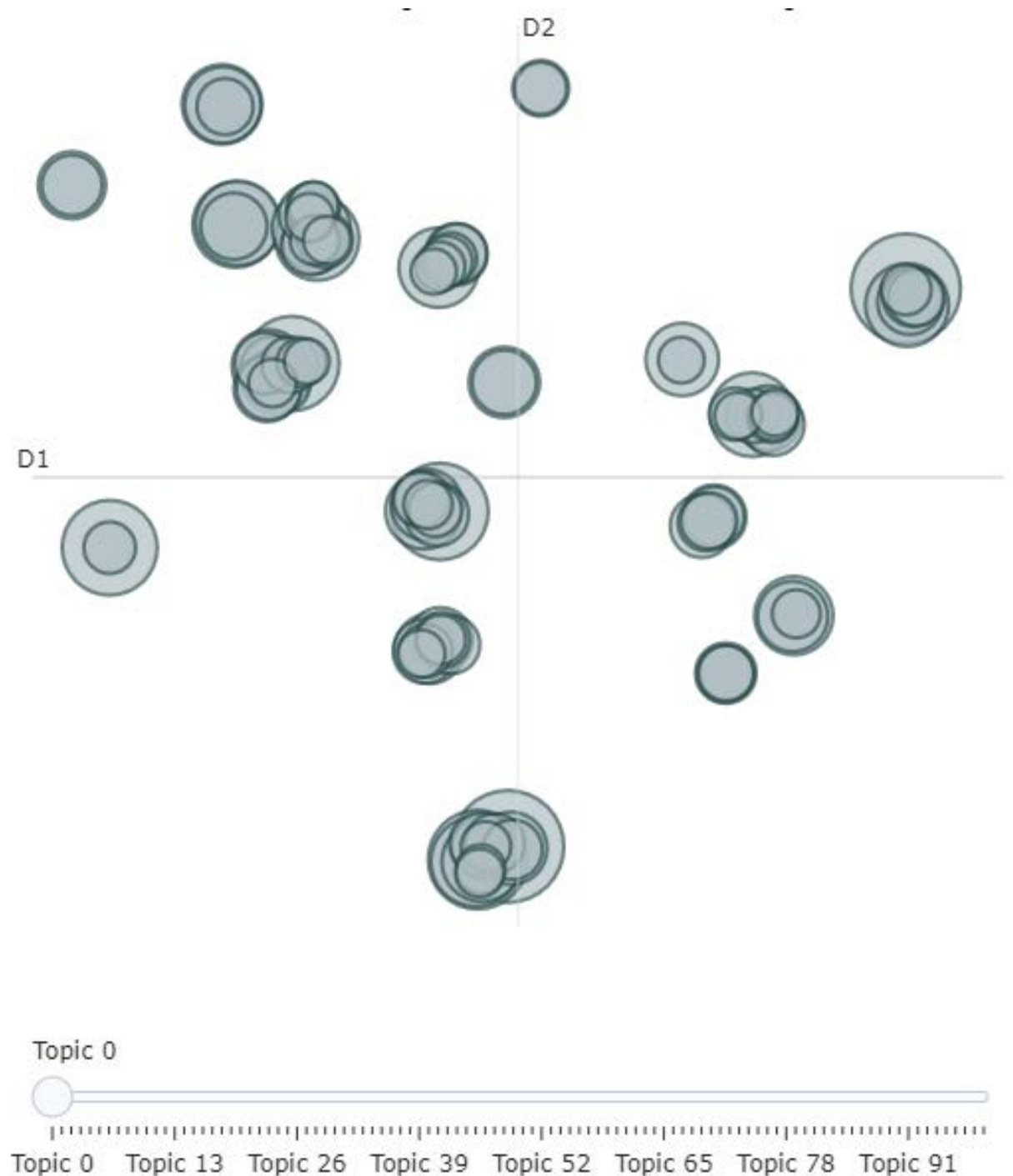


Telegram Topics over Time: July of 2020 – April of 2022

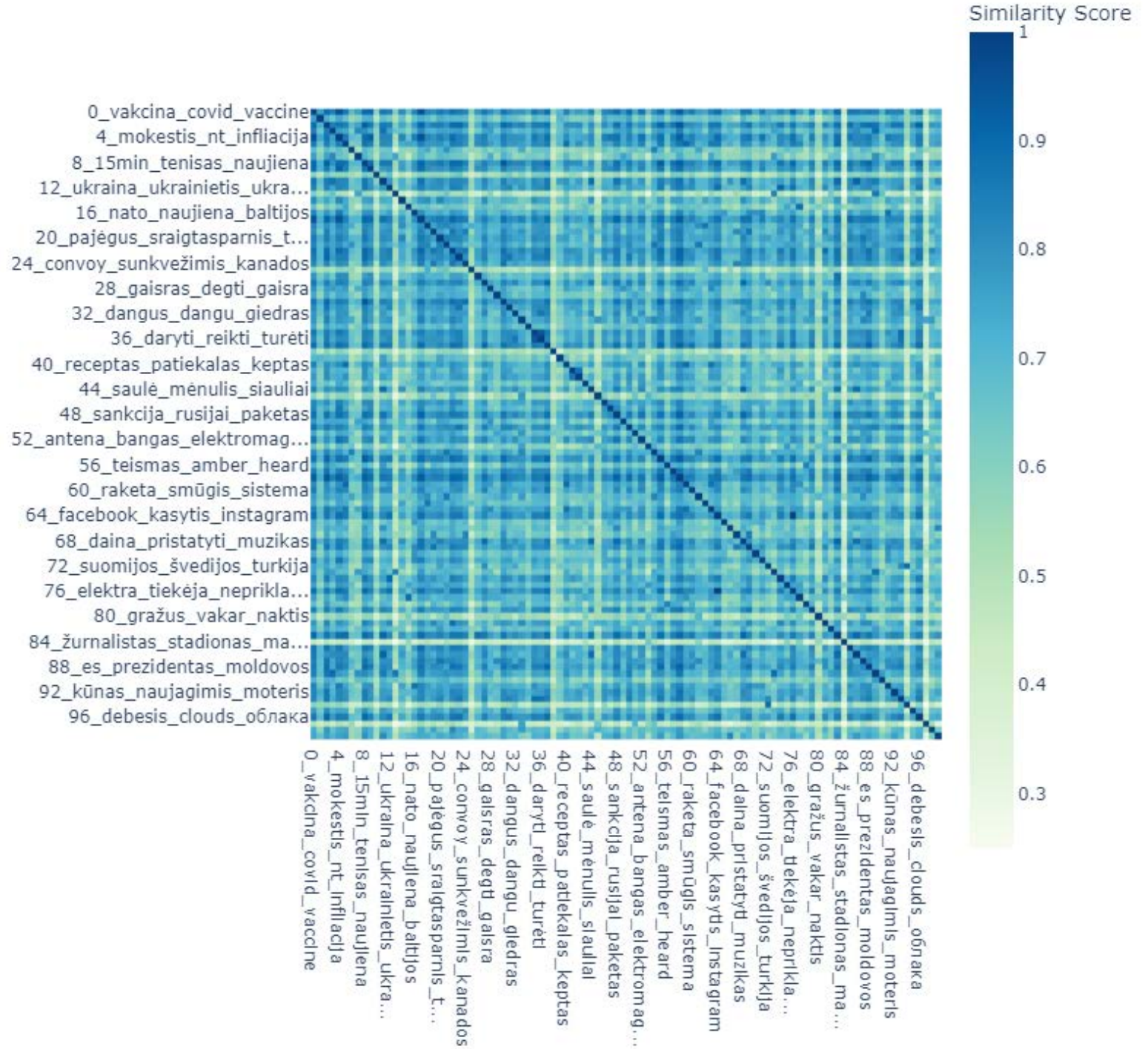


Word Scores for the First 8 Topics

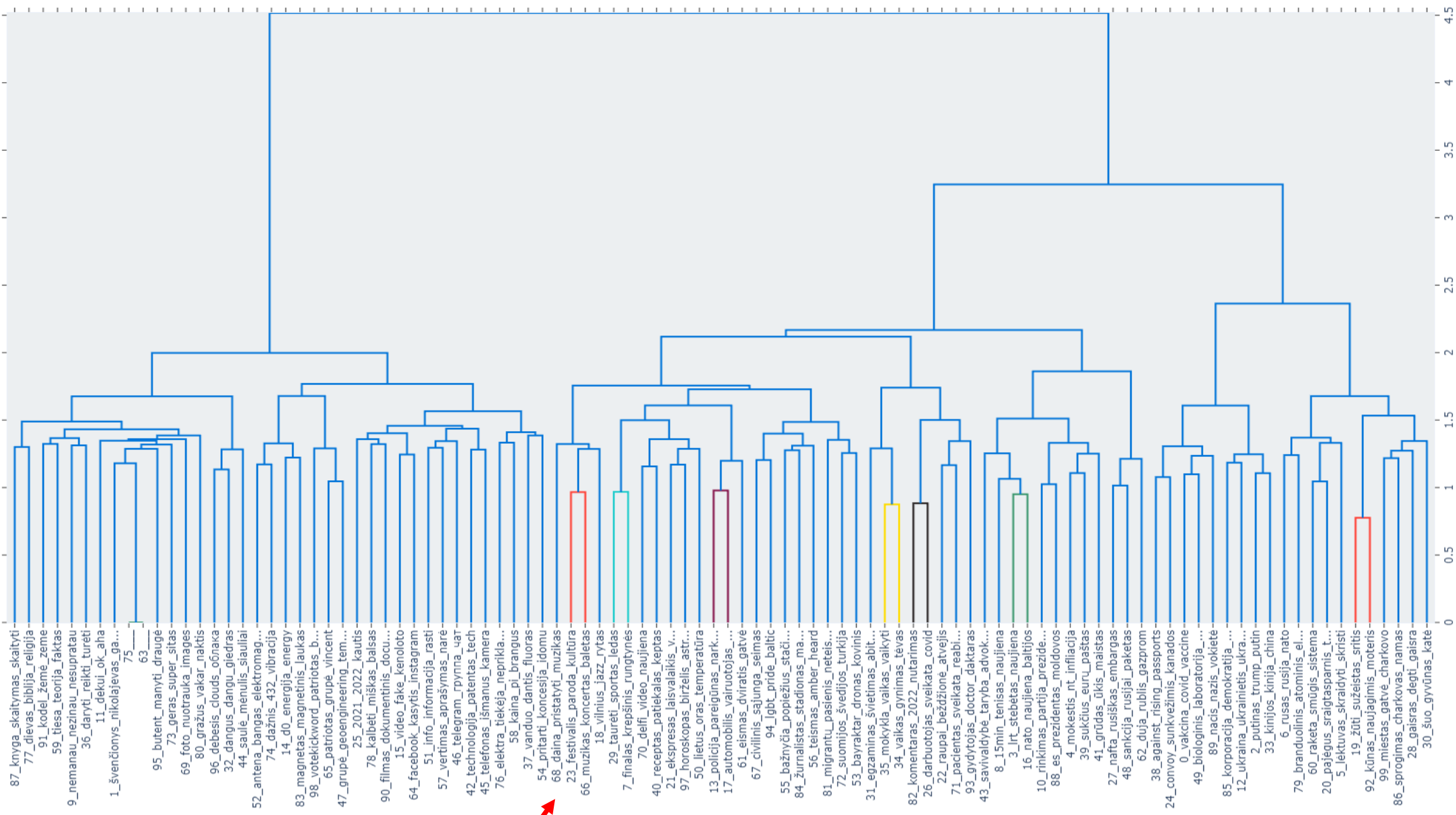
- When visualization is 2D and represents too many topics – not overly useful
- Shows strong overlap on the number of topics, but still at least 18 macro-clusters are visible (3D could change situation)



Intertopic Distances

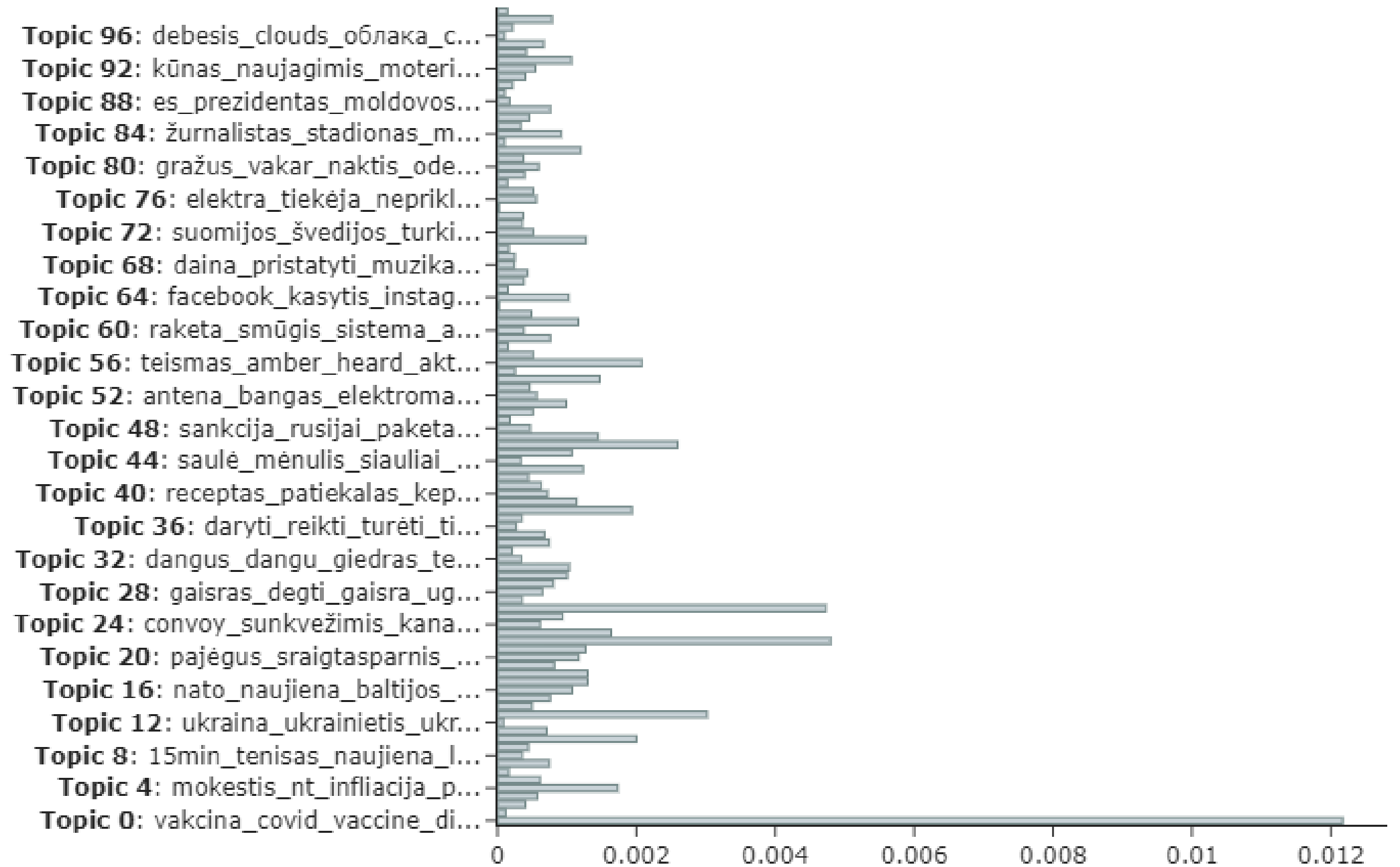


Similarity of the Topics



- 87_knyga_skaitymas_skaityti
- 77_dievas_bibilija_religija
- 91_kodel_žemė_zeme
- 59_tiesa_teorija_faktas
- 9_nemanaui_nezinai_nesupratau
- 36_daryti_reikti_turėti
- 11_dekui_ok_aha
- 1_švenčionys_nikolajevs_ga...
- 75_---
- 63_---
- 95_butent_manyti_draugė
- 73_geras_super_sitas
- 69_foto_nuotrauka_images
- 80_gražus_vakar_naktis
- 96_debesis_clouds_ožnaka
- 32_danguo_dangu_giedras
- 44_saulė_mėnulis_siauliai
- 52_antena_bangos_elektromag...
- 74_dažnis_432_vibracija
- 14_d0_energija_energy
- 83_magnetas_magnetinis_laukas
- 98_votekickword_patriotas_b...
- 65_patriotas_grupė_vincent
- 47_grupė_geoengineering_tem...
- 25_2021_2022_kautis
- 78_kalibėti_miškas_balsas
- 90_filmas_dokumentinis_docu...
- 15_video_fake_kenoloto
- 64_facebook_kasytis_instagram
- 51_info_informacija_rasti
- 57_vertimas_aprašymas_narė
- 46_telegram_rpyrna_uat
- 42_tehnologija_patentas_tech
- 45_telefonas_išmanus_kamera
- 76_elektra_tiekėja_neprikla...
- 58_kaina_pi_brangus
- 37_vanduo_dantis_fluoras
- 54_pritarti_koncesija_idomu
- 68_daina_pristatyti_muzikas
- 23_festivalis_paroda_kultura
- 18_vilnius_jazz_rytas
- 29_taurėti_sportas_ledas
- 7_finalas_krepšinis_rungtynės
- 70_delfi_video_naujiena
- 40_receptas_patiekalas_keptas
- 21_ekspressas_laisvalaikis_v...
- 97_horoskopas_birželis_astri...
- 50_lietus_oras_temperatūra
- 13_policija_pareigūnas_nark...
- 17_automobilis_vairuotojas_...
- 61_eismas_dviratis_gatvė
- 67_civilinis_sąjunga_seimas
- 94_lgbt_pride_baltic
- 55_bažnyčia_popiežius_staci...
- 84_žurnalistas_stadionas_ma...
- 56_teismas_amber_heard
- 81_migrantu_pasienis_netels...
- 72_suomijos_švedijos_turkija
- 53_bayraktar_dronas_kovinis
- 31_egzaminas_švietimas_abit...
- 35_mokykla_vaikas_vaikyti
- 34_vaikas_gynimas_tevas
- 82_komentaras_2022_nutarimas
- 26_darbuotojas_sveikata_covid
- 22_raupai_beždžionė_atvejis
- 71_pacientas_sveikata_reabi...
- 93_gydytojas_doctor_daktaras
- 43_savivaldybė_taryba_advok...
- 8_15min_tenisas_naujiena
- 3_lrt_stebėtās_naujiena
- 16_nato_naujiena_baltijos
- 10_rinkimas_partija_prezide...
- 88_es_prezidentas_moldovos
- 4_mokestis_nt_infliacija
- 39_sukčius_euru_paistas
- 41_grūdas_ūkis_maistas
- 27_nafta_rusiškas_embargas
- 62_duja_rublis_gazprom
- 48_sankcija_rusijai_paketas
- 38_against_rising_passports
- 24_convoy_sunkvežimis_kanados
- 0_vakcina_covid_vaccine
- 49_biologinis_laboratorija_...
- 89_nadis_nazis_vokiete
- 85_korporacija_demokratija_...
- 12_ukraina_ukrainietis_ukra...
- 2_putinas_trump_putin
- 33_kinijos_kinija_china
- 6_rusas_rusija_nato
- 79_branduolinis_atominis_el...
- 60_raketa_smūgis_sistema
- 20_pajėgus_sraigtašparnis_t...
- 5_lektuvas_skraidyti_skristi
- 19_žūti_sužeistas_sritis
- 92_kūnas_naujagimis_moteris
- 99_miestas_gatvė_charkovo
- 86_sprogimas_čarkovas_namas
- 28_gaisras_degti_gaisra
- 30_šuo_gyvūnas_kate





Topic Probability Distribution

Probability

Sentiment

0.000
-0.100
-0.200
-0.300
-0.400
-0.500
-0.600
-0.700
-0.800
-0.900



Sentiments

- Telegram is a trending and very interesting source for public (special groups) analysis
- Ready-made toolsets and libraries (e.g., BERTopic, Polyglot, YAKE) allow performing analysis of informal language even for less popular languages, such as Lithuanian
- AI, NLP, and Visual Analytics are great tools for analysts, but an interpretation of the results may require deeper analysis and discussions



Conclusions

CARD



VYTAUTAS
MAGNUS
UNIVERSITY

Faculty of
Informatics

PASSION LED US HERE

