A dataset of media realeases (Twitter, News and Comments, Youtube, Facebook) form Poland related to COVID-19 for open research

Andrzej Jarynowski, Monika Wójta-Kempa, Małgorzata Stochmal, Daniel Płatek, Martyna Mencel, Jan Maciejewski, Alexander Semenov, Vitaly Belik

Interdisciplinary Research Institute, Wroclaw, Poland; System Modeling Group, Institute of Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Germany; and many others

EOSC USE Case 03.09.2020

Outline

- ▶ Timeline of events in Poland
- ▶ The use of the traditional and social-content media on the Internet in Poland
- ▶ infodemics
- ▶ Data: Twitter, News and Comments, Youtube, Facebook
- ▶ Analysis of text records
- ▶ Analysis of networks

Internet media Poland

In this study, by quantitative analysis of "digital traces/footprints" on the Internet (social and traditional media), we concentrate on the number and nature of events such as information queries and information propagation.
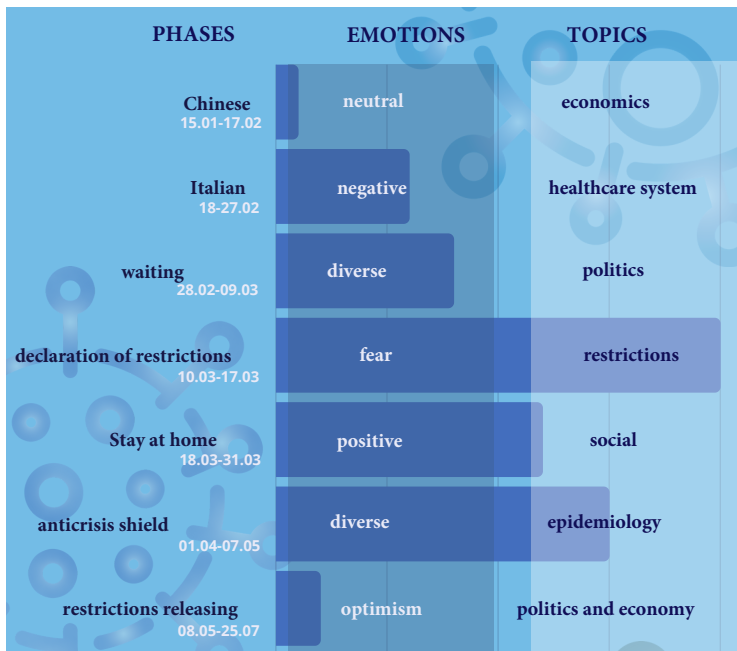
We cover ca. 28 millions Polish language speaking residents of Poland representing over 85% of the literate population. The Polish Internet users are younger (especially among content creators) with females prevailing (especially among content receivers) as compared to the average population. Our results do not describe the perception of the oldest groups, due to digital exclusions, health literacy limitations or consideration of social media as less informative
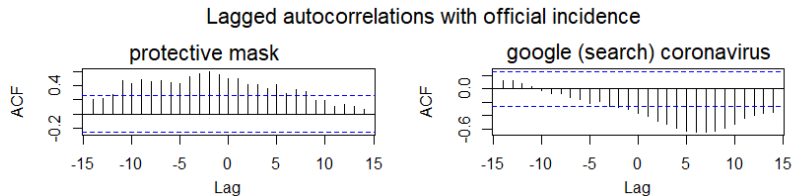
Interest in "Coronavirus" in Poland
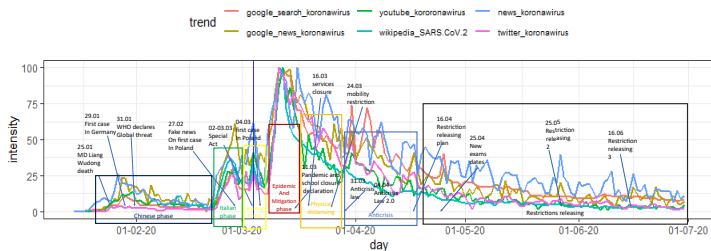
We have identified several temporal major clusters of interest on the topic COVID-19: 1) Chinese, 2) Italian, 3) Waiting, 4) Mitigations, 5) Social distancing and Lockdown, 6) Anti-crisis shield, 7) Restrictions releasing, 8) Incease of incidence.

Phases

PHASES	EMOTIONS	TOPICS
Chinese 15.01-17.02	neutral	economics
Italian 18-27.02	negative	healthcare system
waiting 28.02-09.03	diverse	politics
declaration of restrictions 10.03-17.03	fear	restrictions
Stay at home 18.03-31.03	positive	social
anticrisis shield 01.04-07.05	diverse	epidemiology
restrictions releasing 08.05-25.07	optimism	politics and economy

Interest in "Coronavirus" did not reflect transmision dynamics

Lagged autocorrelations with official incidence

Moreover, "Are Online Searches for the Novel Coronavirus (COVID-19) Related to Media or Epidemiology? A Cross-sectional Study", published in International Journal of Infectious Diseases (doi: https://doi.org/10.1016/j.ijid.2020.06.028) presented the case important conclusions. that online searches for COVID-19 in Europe are not correlated with epidemiology.
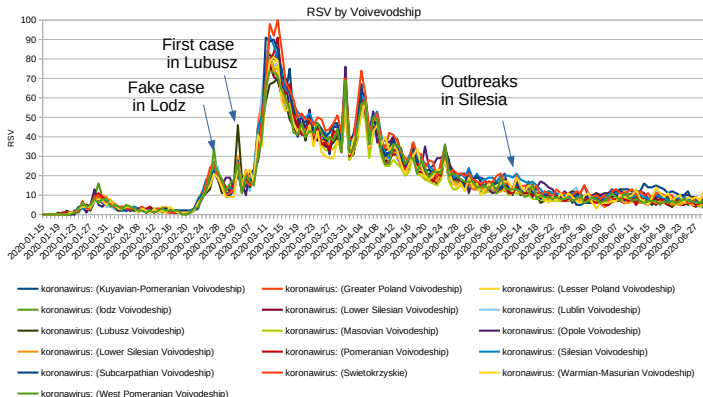
Interest in "Coronavirus" could reflect transmision dynamics in second wave
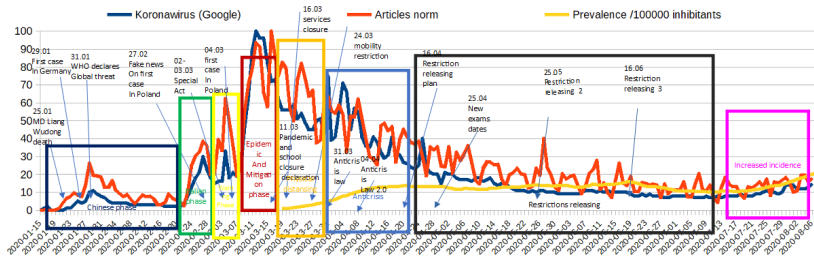
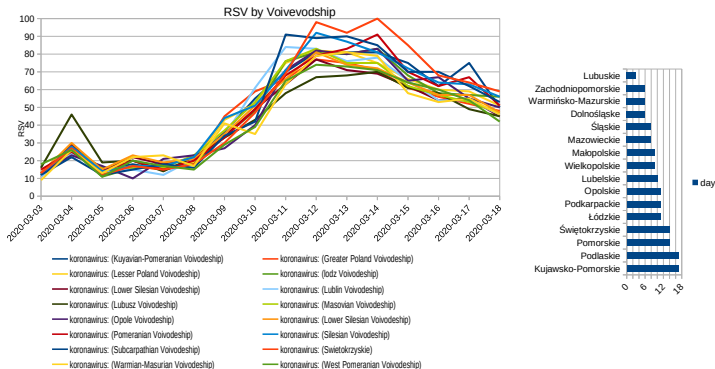Inceance of both incidence and interests in last weeks

Infodemic

- ▶ Traditional vs social-containment media
- ▶ a concept in the social medicine area: "lay-referral framing system" where opinions and beliefs of the general public are something different from medical knowledge.
- ▶ information ws widely treated as a trigger of consequences impacting daily-life security
- ▶ declarations of mitigation strategies by the prime minister or the minister of health gathered the highest attention to the COVID-19 pandemic of the Internet users in Poland. Thus enacted or in force events do not affect the interest to such extent. However, databases on mitigation strategies (WHO, ECDC, etc.) do not keep this in their records, although it would be of great importance.

The Power of Fake News

An effect of fake news of the first confirmed case

The Power of Information

RSV by Voivevodship

koronawirus: (Kuyavian-Pomeranian Voivodship) — koronawirus: (Greater Poland Voivodship)
koronawirus: (Lesser Poland Voivodship) — koronawirus: (łodz Voivodship)
koronawirus: (Lower Silesian Voivodship) — koronawirus: (Lublin Voivodship)
koronawirus: (Lubusz Voivodship) — koronawirus: (Masovian Voivodship)
koronawirus: (Opole Voivodship) — koronawirus: (Lower Silesian Voivodship)
koronawirus: (Pomeranian Voivodship) — koronawirus: (Silesian Voivodship)
koronawirus: (Subcarpathian Voivodship) — koronawirus: (Swietokrzyskie)
koronawirus: (Warmian-Masurian Voivodship) — koronawirus: (West Pomeranian Voivodship)

Introduction day as the main driving force

Datasets (deposited)

https://zenodo.org/record/3985807

- ▶ extracted 57,306 representative articles in Polish using Eventregitry.org tool in language Polish and topic "Coronavirus" in article body;
- ▶ extracted 1,015,199 and Tweets from #Koronawirus in language Polish usiing Twitter API.
- ▶ collected 1,574 videos with keyword: Koronawirus on YouTube and 247,575 comments on them using Google API;

Datasets (deposited) 2

- manually labeled 1,449 articles / Facebook posts from Lower Silesia and 111 texts from outside this region;
- manually labeled 1,500 Twitts / Youtube comments (for ML training);
- analysis of 244 social empirical studies till 25.05 on COVID-19 in Poland.

Techniques

Methodological and statistical analysis is a crucial point in this kind of study. We apply:

- ▶ the number and nature of social media events such as information queries (time series analysis);
- ▶ sentiment and conceptual fields analysis;
- ▶ social network analysis;
- ▶ topic modeling technique;
- ▶ comparative approach bewtween local and national-wide media;
- ▶ quasi-systematic review of studies.

Twitter

- We have collected $>1\ 000\ 000$ with #Koronawiurs
- Twitter in Poland has relatively low popularity (~3 million registered users or less than 8% of the country population) and is mainly used by expats, journalists and politicians
- For Twitter data we will deploy the temporal multilayer network analysis of users (using metadata on quotes, retweets, mentions, replies, and followings), because networks can be constructed based on the mentioned metadata types.

Retweet networks
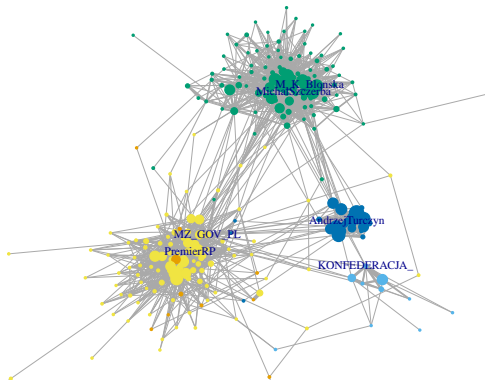
A network of retweets with the Koronawirus hashtag in Polish language during 28.01-11.03.2020 period (250 most central nodes according to weighted centrality, coloring according to Louvain community detection algorithm). Yellow – the ruling coalition, green – the mainstream opposition, dark blue – a Protestant minority, light blue – a far right party

Polarization
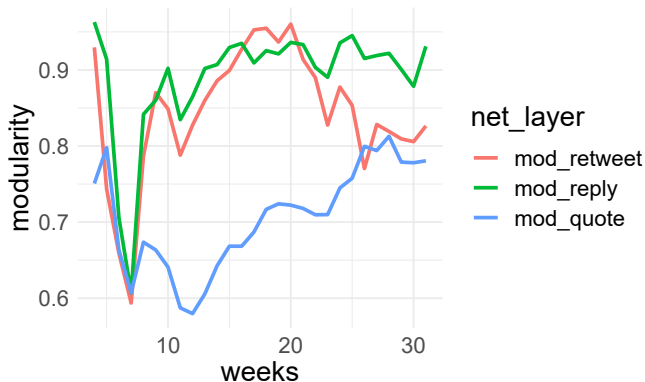
Modularity of the full networks (over 30 thousands accounts) and their weekly dynamics for different type of communication relations (retweeting, reply, quotation).

Actors
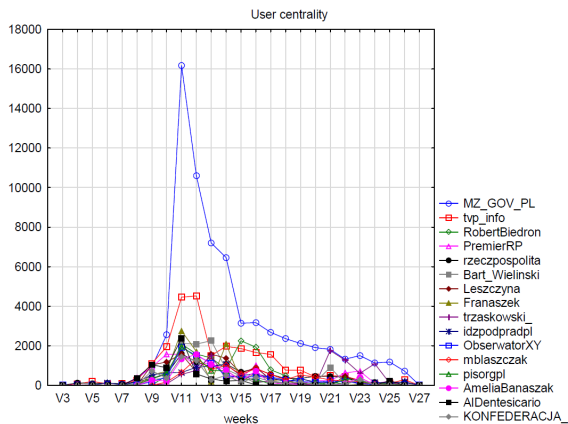
The most central users

News articles

- We have collected >60000 epresentative digital traditional media news (selected with EventRegistry tool) with stemmed word Koronawirus in title or short body.
- We choose EventRegistry (eventregistry.org) as a traditional media search engine because it has a large range of online news services representing a broad political spectrum. Besides, it gives priority to digital versions of other broadcasting channels, including television, radio or newspapers.
- For newss data we will deploy the temporal multilayer network analysis of aricles or news agency based on context similarity

Article sources
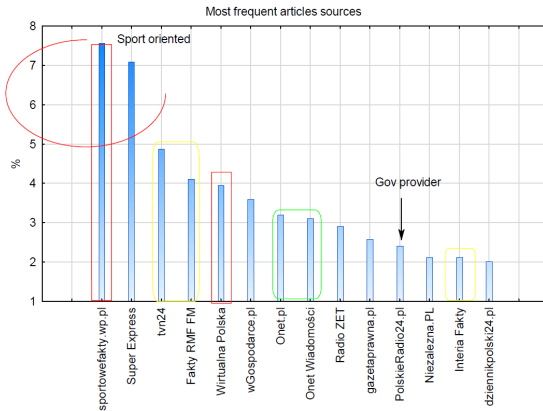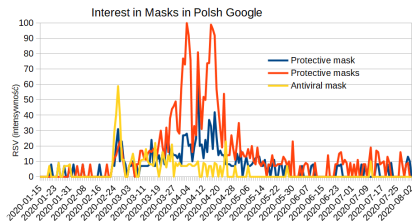

The most frequent sources

Article sources time dynamics

Sources in time

Youtube
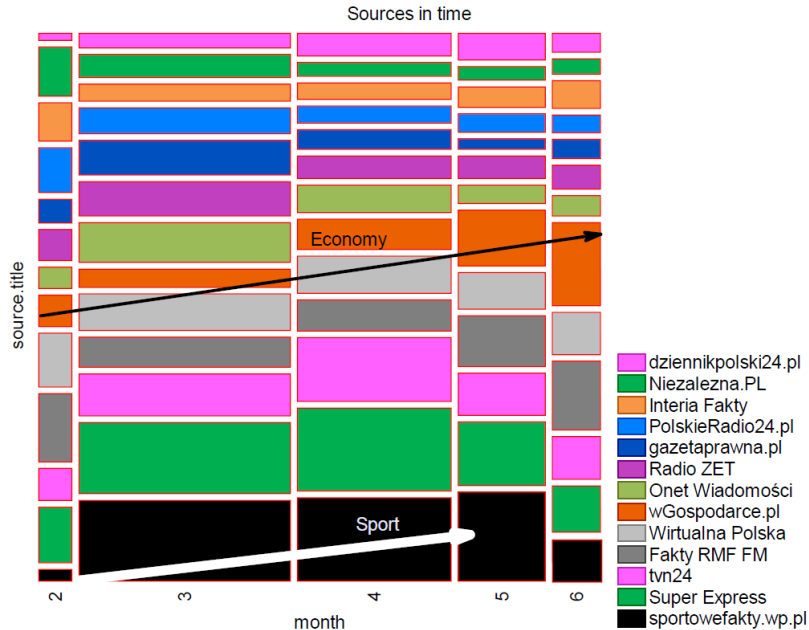
1	4493638	2020-03-11	Koronawirus: jak sie ochronic?	26	697988	2020-03-26	Koronawirus. Kiedy to sie skonczy?
2	2906744	2020-03-29	Andrzej Duda - Koronawirus (Sanah - Szampan PARODIA)	27	682970	2020-01-26	Tech Week #1: Koronawirus - to wcale nie jest smieszne
3	2243994	2020-05-22	Ewa Farna #hot16challenge2 (prod. Ochepovsky)	28	668031	2020-03-14	KORONAWIRUS - jak nie zachorowac? Stosuj "witamine" HSDI
4	2235800	2020-03-18	DISS NA KORONAWIRUS - MALCZYNSCY	29	661125	2020-01-31	Wirus z Chin. Analiza sytuacji.
5	2168029	2020-02-18	Chiny, koronawirus i tajemnice. Co ukrywa Xi Jinping?	30	656835	2020-02-28	KORONAWIRUS - realne zagrozenie, czy medialna panika? Eksperci odpowiadaja [LIVE]
6	1931949	2020-03-23	DODA - DZAGA <U+0001F602> PARODIA MATKI <U+0001F525> KORONAWIRUS EDITION	31	608529	2020-03-16	FAQ koronawirus - 10 pytan o pandemie i SARS-CoV-2
7	1904561	2020-03-09	Czy koronawirus moze zniszczyc ludzkosc?	32	576650	2020-04-06	Dr Jerzy Jaskowski (1/2): To co sie dzieje wokól pandemii koronawirusa to tylko i wylacznie oszustwo
8	1744599	2020-03-02	SPEJSONVLOG - KORONAWIRUS	33	561638	2020-02-19	Cejrowski o koronawirusie #Covid19 2020/2/17 Studio Dziki Zachód odc. 46 cz. 1
9	1703443	2020-03-14	Lil Narcyz - KORONAWIRUS ft. Kolba	34	541428	2020-04-23	Nie jestem lekarzem... ALE... - teorie spiskowe, koronawirus i ludzie, którzy wiedza lepiej...
10	1308960	2020-03-15	Szmitek mix KoronaWirus	35	539863	2020-04-02	Barbara Kwarc - KORONAWIRUS
11	1162461	2020-03-10	Chorowalismy na koronawirusa juz dawniej i nawet o tym nie wiedzielismy?	36	526812	2020-01-25	Koronawirus
12	1100925	2020-05-18	Kamil Bednarek #Hot16Challenge2	37	524967	2020-03-11	TYPOWY JANUSZ PODCZAS EPIDEMII KORONAWIRUSA
13	1096400	2020-04-08	Koronawirus interwencja Policji caly film	38	509416	2020-03-20	Koronawirus MeMes 2
14	1089199	2020-03-14	Jak wyglada Polska w czasach koronawirusa? Onet sprawdzil sytuacje	OnetNews	39	506601	2020-03-18	Koronawirus okiem naukowca - dr hab. Piotr Rzymski
15	1033772	2020-02-13	Zycie w SZANGHAJU podczas epidemii koronawirusa	40	491661	2020-03-13	KORONAWIRUS W JAPONII
16	957107	2020-03-01	Hubert Czerniak TV - KORONAWIRUS / Analiza lekarza / Najwazniejsze informacje / Profilaktyka	41	484484	2020-03-12	Szymon Holownia odpowiada na pytania: koronawirus, wiara, aborcja, ekologia	Impoderabilia #94
17	925832	2020-03-22	DISCO POLO - KORONAWIRUS?! "Piosenka"	42	467996	2020-03-10	"SYTUACJA JEST POWAZNA" - Oredzie Prezydenta Andrzeja Dudy ws. KORONAWIRUSA
18	897343	2020-02-11	Na naszym osiedlu wykryto koronowirusa.	43	465521	2020-03-19	Koronawirus. Czy jest grozny dla mlodych?
19	862133	2020-03-02	Wirusolog Wlodzimierz Gut: nie panikujmy z powodu koronawirusa	44	447278	2020-03-18	PJK Eastny - Jak zaczarowac
20	861045	2020-03-17	Koronawirus - co Cie czeka, gdy zachorujesz?	45	438422	2020-03-19	koronawirus (MYSZOJELEN - PARODIA)	Spetany
21	833495	2020-01-27	Jak wyglada zycie w WUHAN	Skad i KIEDY pojawil sie WIRUS?	46	423570	2020-04-02	Koronawirus. Co robi z Twoja psychika?
22	806170	2020-03-04	Sanepid w Slubicach. Pierwsze uderzenie.	47	421709	2020-03-04	KORONAWIRUS SONG (PRZEJMUJEMY JUTUBY - EKIPA x WIP BROS PARODIA) reprod. Krystjanin
23	744774	2020-03-01	OPANOWALISMY epidemie koronawirusa W CHINACH?	48	405960	2020-03-25	JANUSZE KORONAWIRUSA
24	741111	2020-03-29	KORONAWIRUS - 3 SCENARIUSZE NA PRZYSZLOSC	49	405283	2020-02-13	EPIDEMIA i GLÓD W CHINACH! Ide do sklepu. #Koronawirus #Chiny, #Epidemia, #DamianChen
25	731617	2020-03-26	W co gra Xi Jinping? Koronawirus rozlal sie po swiecie, a Chiny spiesza z pomoca - OKO NA SWIAT	50	384066	2020-01-27	KORONAWIRUS W AZJI !

Good source for cultural effect of COVID and monitoring protest movements.
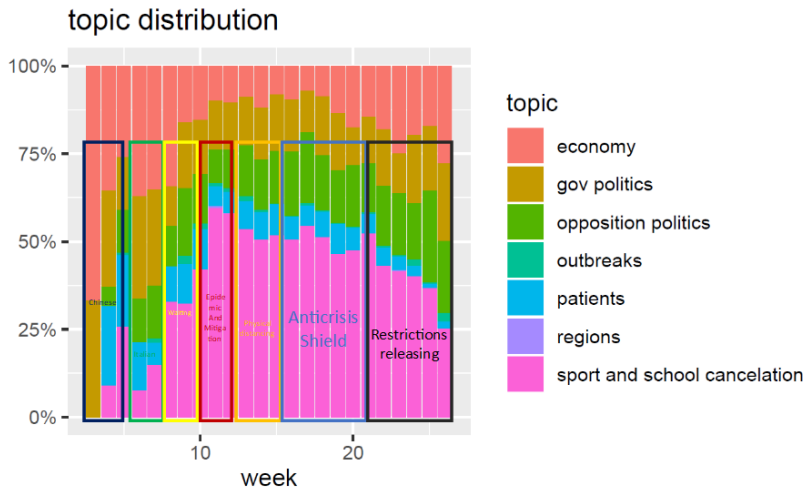
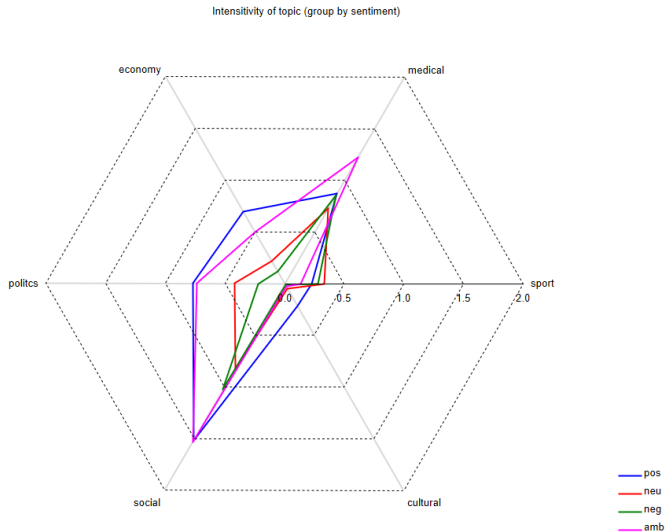# Sociolingistics

New Dictionary for Corpus

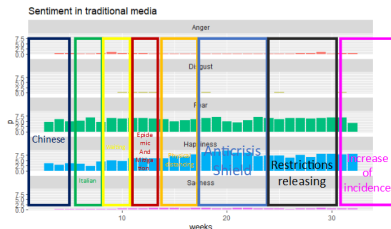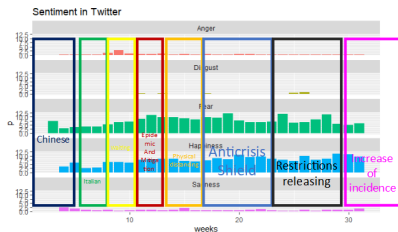## Intrest in masks

Topic Modelling

Weekly dynamics of media perception in Poland. Main Topics in over 50 thousand news articles (without garbage codes).

Topics

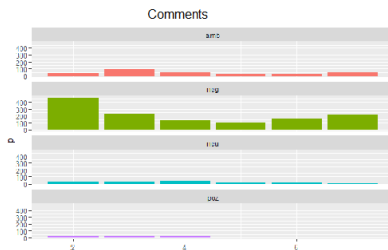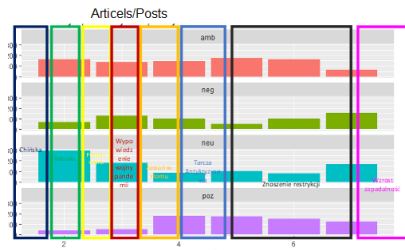Intensitivity of topic (group by sentiment)

Topics manually coded

Sentiment

Sentiment across ML labeled texts

Sentiment across manually labeled texts

Empirical Social Studies

- ▶ 180 projects carried out by scientific and commercial institutions in the period from January to May 2020.
- ▶ present a standard way of conducting empirical studies by social researchers who undertake the challenge of recognizing pandemic conditions and challenges in the future
- ▶ mitigating the polarization of social attitudes which are dynamically changing during the pandemic,
- ▶ practical solving / and not only diagnosing / problems revealed in covid reality
- ▶ supplementing the deficiencies of theoretical assumptions accompanying research projects.

Empirical Social Studies factor analysis

	Mobility patterns	Norms internalization	Negative emotions	Economy	Risk factors	Individual Risk Mitigation	Community Risk Mitigation	Attitudes towards institution	Reporting
Ograniczenie transportu publicznego	.736						.328		
Ograniczenie wizyt w centrach handlowych	.731								
Podróżowanie	.696								
Urlopy/wakacje	.551								.435
Ograniczenie uczęszczania do kościoła	.419						.360		
Kwarantanna		.811							
Izolacja fizyczna		.776							
Dystans społeczny		.697			.314				
Postawy wobec restrykcji		.675		.324					
Epidemia/pandemia		.463		.423					
Lęk			.973						
Strach			.716						
Gniew			.549						.438
Niepokój			.472						
Zahamowanie gospodarki				.750					
Straty finansowe				.709					
Bezpieczeństwo ekonomiczne				.674					
Ożywianie gospodarki				.537					
Czynniki ryzyka					.632				
Zapobieganie COVID19					.629	.409			
Percepcja zagrożenia		.326			.575				
Zarażenie zakażenie		.316			.543				
Mycie i dezynfekcja przedmiotów						.697			
Mycie i dezynfekcja rąk						.689			
Zdrowe odżywianie i nawadnianie organizmu						.531			.495
Maseczki ochronne						.509			.312
Zdalna praca i nauczenie							.803		
Odwołanie zajęć w szkolnictwie							.792		
Gromadzenie zapasów							.379		
Ocena działań grup dyspozycyjnych								.887	
Ocena działań służb rządowych				.380				.559	
Ocena działań służby zdrowia								.508	
Fakenews/dezinfomcja								.446	.327
Statystyki na temat korona wirus a Świat									.628
% \Variance explained	20.34	10.28	7.06	5.60	4.55	4.83	3.36	2.72	2.05

Main factors of social studies objects

Intervention schema
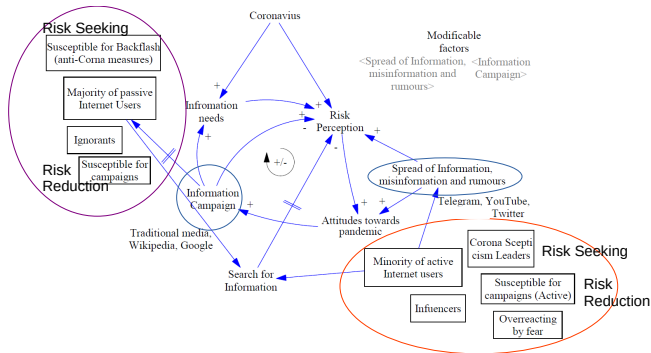
An attempt to describe main aspects of a Polish social system in Pandemic time. Rectangles are subpopulations, arrows mean main interactions, shadow variable are modifiable factors, signs represents positive or negative influence, crossing bars means interaction which are the most likely to disappear with time.

Conlusions

- ▶ The COVID-19 epidemic in Poland is primarily of a social, not medical, dimension;
- ▶ Local media are more reliably than national media, and Lower Silesians seem to be less susceptible to media manipulation and conspiracy theories;
- ▶ Campaigns should be targeted to sccific population and regions;
- ▶ Declarations of response strategies by the Polish prime minister or the minister of health gathered the highest attention of Internet users. So already enacted or in force events do not affect the interest to such an extent;
- ▶ The comments are significantly more likely to be negative.

www.infodemia-koronawirusa.pl