

Social dimension of COVID-19 pandemic in Poland on the Internet

Andrzej Jarynowski^{1,2} Alexander Semenov^{3,4} Monika
Wójta-Kempa⁵ Vitaly Belik¹

¹System Modeling Group, Institute of Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Germany

²Interdisciplinary Research Institute, Wrocław, Poland

³Department of Industrial and Systems Engineering, University of Florida, USA

⁴Centre for Econometrics and Business Analytics, Saint Petersburg State University, Russia

⁵Department of Public Health, Wrocław Medical University, Poland

Jul 10, 2020

Internet in Poland

COVID-19 represent a major disruptive event affecting the whole society. Internet provides an easy tool to assess data on the COVID-19 related interest and societal dynamics.

Internet covers 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 probably do not describe the perception of the oldest groups, due to digital exclusions, health literacy limitations or consideration of social media as less informative sources.

Interest in “Coronavirus” in Poland

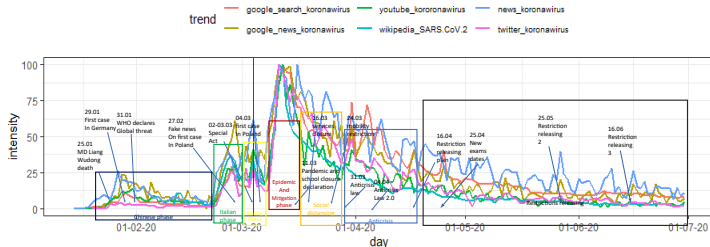


Figure 1: Several temporal major clusters of interest on the topic COVID-19: 1) Chinese, 2) Italian, 3) Waiting, 4) Mitigations, 5) Social distancing and Lockdown, 6) Anti-crisis shield, 7) Restrictions releasing.

Interest in “Coronavirus” and COVID-19 confirmed incidence in Poland

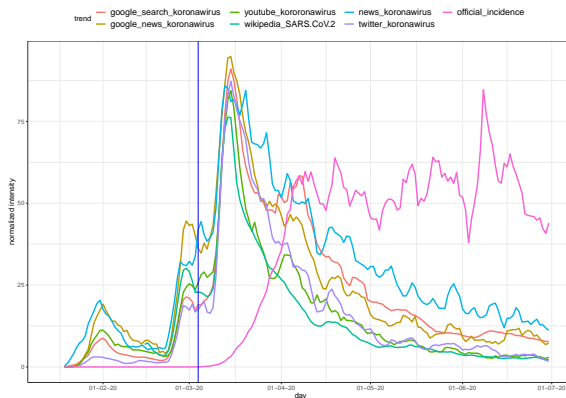


Figure 2: Interest on the Internet and COVID-19 incidence.

- ▶ Information is vastly considered as a trigger of implications for daily-life security
- ▶ Traditional vs social media
- ▶ 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.

Correlation of Internet queries

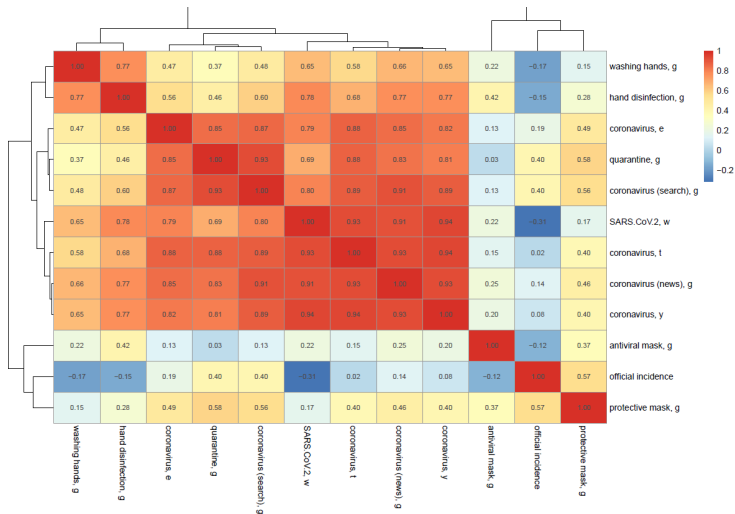


Figure 3: Heatmap of Pearson correlation of COVID-19 related Internet queries.

Twitter I

- ▶ We have collected 930319 tweets with #Koronawirus during 15.01-30.06.2020
- ▶ Twitter in Poland has a 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 investigate i) a (temporal) network of users linked by retweets and ii) user network based on the topic similarity of their tweets.

Twitter II

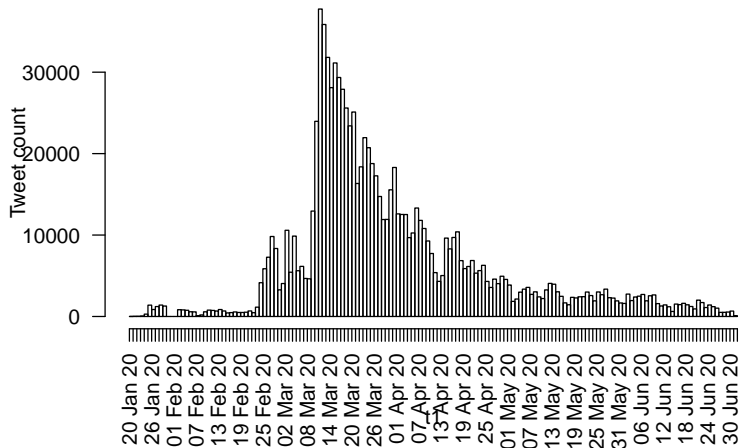


Figure 4: Total number of tweets including retweets with Coronavirus hashtag.

Fraction of original tweets

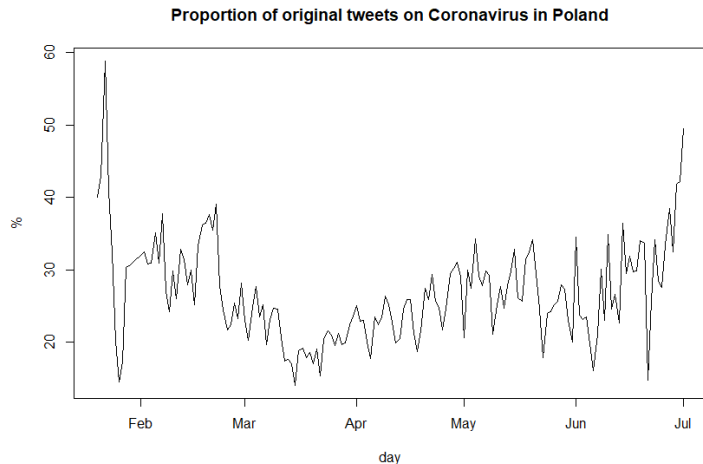


Figure 5: Fraction of original tweets vs time.

Frequency distribution of twitter users

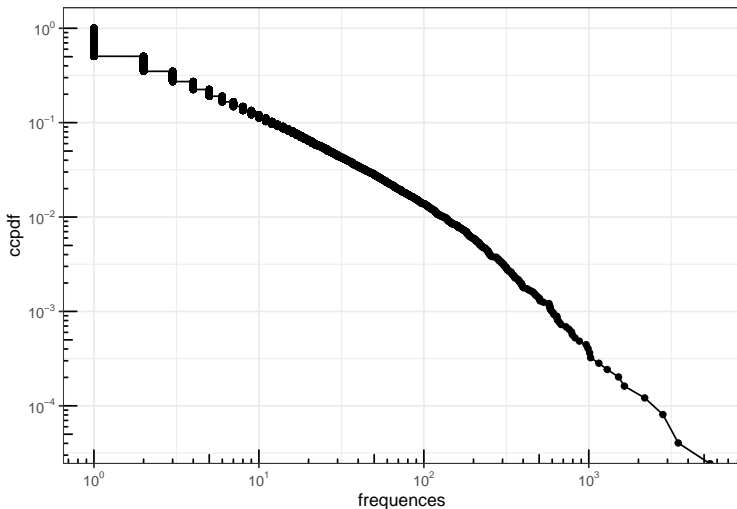


Figure 6: CCPDF for the frequency of tweets for a given user.

Retweet networks

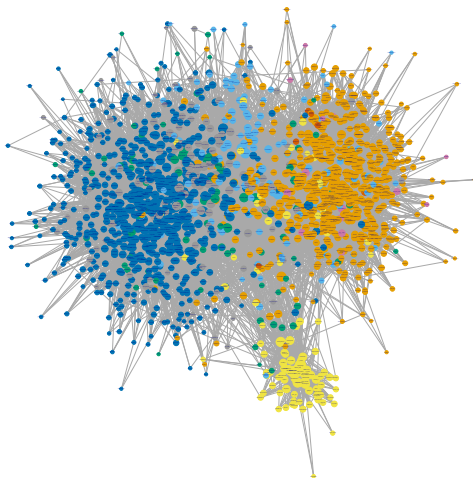
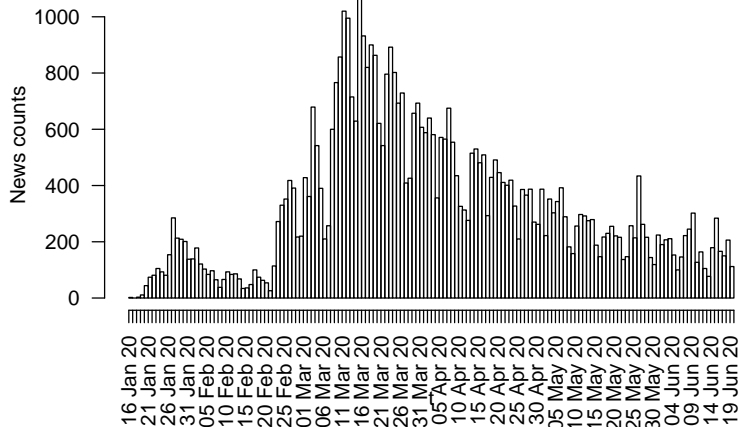


Figure 7: A network of retweets with the Koronawirus hashtag in Polish language during 15.01-30.06.2020 period (1500 most central nodes by weighted degree centrality, color by Louvain communities). Blue – the ruling coalition, Orange – the mainstream opposition, Yellow – a Protestant minority

- ▶ We collected 51962 representative 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 news data we will deploy temporal network analysis of news media sources based on the context similarity (*doc2vec*)

News articles II



Frequency distribution of news sources I

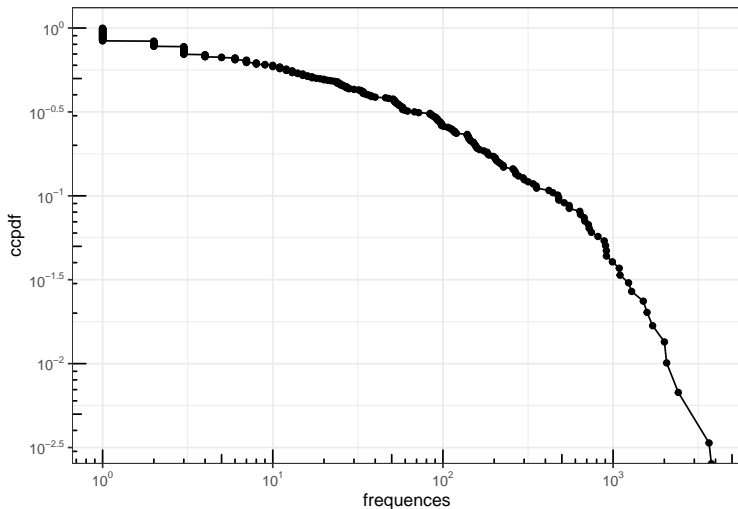


Figure 8: CCPDF for the frequency of news sources.

Frequency distribution of news sources II

Table 1: The most frequent news sources.

Source	Frequency
sportowefakty.wp.pl	3780
Super Express	3664
tvn24	2417
Fakty RMF FM	2064
Wirtualna Polska	2003
wGospodarce.pl	1707
Onet Wiadomooci	1583
Radio ZET	1505
gazetaprawna.pl	1285
PolskieRadio24.pl	1233
Niezalezna.PL	1099
Interia Fakty	1087
dziennikpolski24.pl	991
PolsatNews.pl	914
forsal.pl	913
Rzeczpospolita	903
Onet.pl	887
FAKT24.pl	815
onet.pl	746
dziennik.pl	724

Networks of news media sources

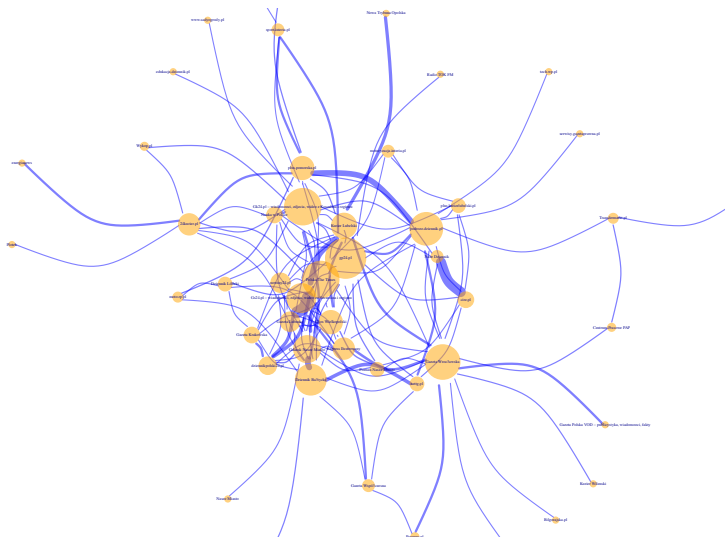


Figure 9: We constructed a network of 297 media sources based on cosine similarity measures of doc2vec of news excerpts.

Temporal networks of news media sources



Figure 10: Temporal network of the 50 largest media sources based on cosine similarity measures of doc2vec of news excerpts [<https://tinyurl.com/infodemics>].

Dynamics of similarity in news media

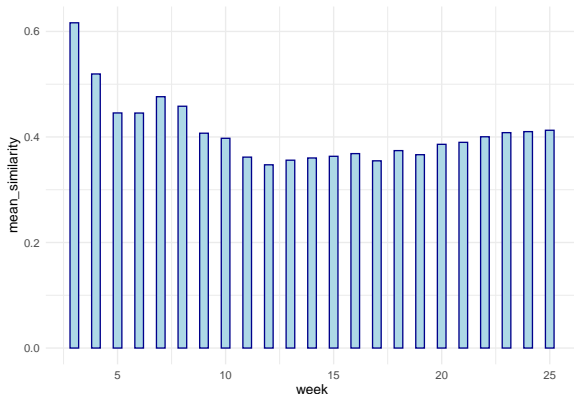


Figure 11: The average topic similarity (cosine similarity based on doc2vec) between news excerpts.

Topic Analysis for News

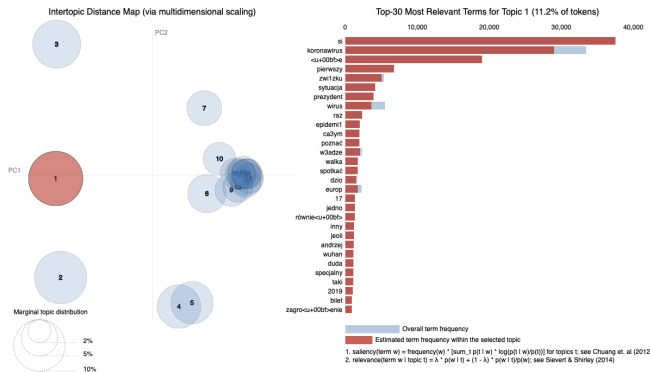


Figure 12: Major topics of news corpus with pyLDAvis [https://tinyurl.com/infodemics]. Topics: football (8,10,11); election (9,15); politics (12,14); schools (13), churches (19), economics (3), medical (4,16).

Topic Analysis for Twitter

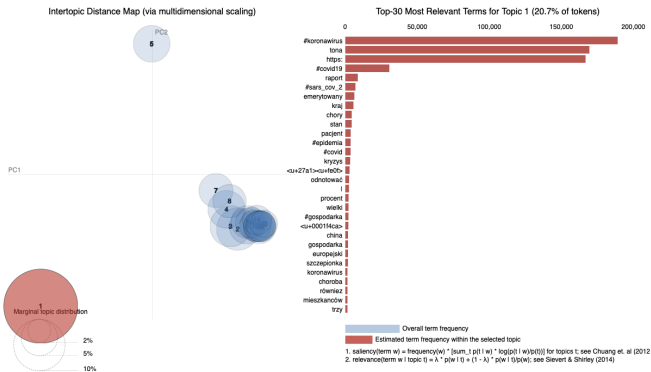


Figure 13: Major topics of news corpus with pyLDAvis [https://tinyurl.com/infodemics]. Topics: Government (1,4,12,18,19); restrictions (6,17); Warsaw (3,8); Silesia (13,16); Wroclaw (14); economy (5); surveillance (2,7).

Sentiment analysis of news

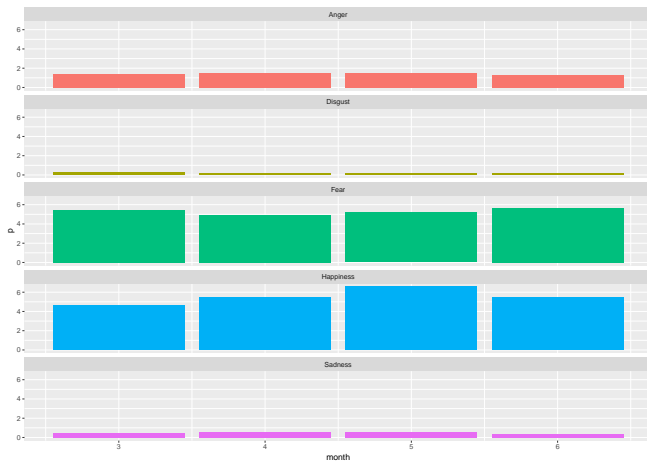


Figure 14: Monthly sentiment analysis of news media.

Conclusions

- ▶ Announcement of measurements and not their actual enacting spikes the interest on the Internet
- ▶ Retweet networks revealed a decrease of a strong political polarization in Poland during the peak of the interest to COVID-19
- ▶ *Diversity* of both news and tweets seem to show a pronounced minimum during the major Interest spike
- ▶ Topic similarity networks in the news allowed to identify the relevant topics such as *sports, travel*, regional clusters of news sources
- ▶ LDA topic modeling for both Twitter and news revealed identifiable topics such as *governmental, elections, economical, football, schools, regional*
- ▶ Sentiment analysis of the news revealed a counter-intuitive peak of *happinnes* in May and minimum of *fear* in April