Perception of emergent epidemic of COVID-2019 / SARS CoV-2 on the Polish Internet

Andrzej Jarynowski*

Interdisciplinary Research Institute, Wroclaw, Poland

Monika Wójta-Kempa†

Department of Public Health, Faculty of Health Science, Wroclaw Medical University, Poland

Vitaly Belik‡

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

Problem: Due to the spread of SARS CoV-2 virus infection and COVID-2019 disease, there is an urgent need to analyze COVID-2019 epidemic perception in Poland. This would enable authorities for preparation of specific actions minimizing public health and economic risks.

Methods: We study the perception of COVID-2019 epidemic in Polish society using quantitative analysis of its digital footprints on the Internet (on *Twitter*, *Google*, *YouTube*, *Wikipedia* and electronic media represented by *Event Registry*) from January 2020 to 12.03.2020 (before and after official introduction to Poland on 04.03.2020). To this end we utilize data mining, social network analysis, natural language processing techniques. Each examined internet platform was analyzed for representativeness and composition of the target group.

Results: We identified three temporal major cluster of the interest before disease introduction on the topic COVID-2019: China- and Italy-related peaks on all platforms, as well as a peak on social media related to the recent special law on combating COVID-2019. Besides, there was a peak in interest on the day of officially confirmed introduction as well as an exponential increase of interest when the Polish government "declared war against disease" with a massive mitigation program. From sociolingistic perspective, we found that concepts and issues of threat, fear and prevention prevailed before introduction. After introduction, practical concepts about disease and epidemic dominate. We have found out that Twitter reflected the structural division of the Polish political sphere. We were able to identify clear communities of governing party, mainstream oppostition and a protestant group and potential sources of disinformation. We have also detected bluring boundaries between comminities after disease introduction.

Conclusions: Traditional and social media do not only reflect reality, but also create it. Due to filter "bubbles" observed on Twitter, public information campaigns might have less impact on society than expected. For greater penetration, it might be necessary to diversify information channels to reach as many people as possible which might already be happening. Moreover, it might be necessary to prevent the spread of disinformation, which is now possible in Poland due to the special law on combating COVID-2019.

* ajarynowski@interdisciplinary-research.eu
† monika.wojta-kempa@umed.wroc.pl
‡ vitaly.belik@fu-berlin.de

INTRODUCTION

Although a large part of the Polish population has heard about coronaviruses for the first time a few weeks ago, in reality they face less dangerous coronaviruses causing simple cold all the time. Only the emergence of a novel strain from Wuhan gave the word "Coronavirus" a new meaning. Before disease introduction, less than a half of surveyed Poles believed that corona virus is the most important topic in the second half of February 2020 (IBRIS [1]). The disease was not detected in Poland until 03.03.2020, but this topic was relatively important already before introduction and started to drive media and social life after the disease introduction. During opinion poll performed on 09.03.2020 and 10.03.2020, 63% of respondents reported, that this is a serious threat for Poland, 40% that this is a serious threat for their family and 73% that this is a serious threat for the Polish Economy ([2]). We observe several small peaks of interest on the Internet for Coronavirus in Poland during investigated period (since January 2020 till 11.03.2020), although the largest surge in interest occurs after the official confirmation of the virus introduction to Poland on 04.03.2020 . The last day of analysis 11.03.2020 was a day of pandemic declaration by WHO.

Recently social media activities are being analyzed worldwide to better understand perception and spread of diseases. This helps in some cases to track the spread of diseases (Ginsberg et al. [3], Lu et al. [4], Joshi et al. [5]) with a higher precision than other methods. The Internet is a good ground for propagation of views often contradicting the current state of medical knowledge [6]. Social media can serve as a valuable source of information as well as disinformation about the virus globally, fueling panic and creating so-called infodemic ([7]), at unprecedented speed and massively desrupting entire countries such as Italy (Guardian [8]). Propaganda and persuasion techniques are widely used on the Internet easily reaching certain target groups susceptible to conspiracy theories and effectively polarizing societies, possibly due to the interference by foreign intelligence [9]. Panic-related behaviors accompanying a virus outbreak [10] are influencing the epidemic spread (Perra et al. [11], Fenichel et al. [12], Wang et al. [13]), with the Internet being the main mediating mechanism. This could lead to destructive collective behavior, such as e.g. xenophobia against people from affected countries.

Methodology

In this study, by quantitative analysis of digital traces on the Internet (e.g. social media), we concetrate on the number and nature of social media events such as information queries and deploy social network analysis, as well simple natural language processing topic modeling techniques.

Up to our knowledge there were no previous studies quantitatively linking the Internet activities and risk

perception of infectious diseases in Poland ([14, 15]), expect our own study ([16]). Thus the present study is a first exploratory attempt filling this gap to continuing prelimiary research on Coronovirus perception in Poland before disease introduction ([16]).

We primarily analyze quantitative digital footprint data on the Internet from January 2020 to 11.03.2020, including their representativeness. The number of internet users in Poland in January 2020 was 28.1 million (PBI [17]) and 28.6 milion in 2019 (PBI [18]). Internet covers 85% of the total literate population in Poland. Thus, the passive representativeness of the Internet is relatively high, but active (own content creation) is biased towards younger age groups and women. The former group has a high activity on the Internet, e.g. an average Polish teenager spends about 5 hours a day on the Internet (Tanaś et al. [19])) and the later group is responsible for generation of up to 85% of health-related content in social media((PBI [17]). Over 99% of young Polish women use the Internet (Jarynowski and Belik [20])).

As a keyword in queries of web services ([15]) we chose a colloqial term "Koronawirus"/"Coronavius" due to its penetration in the society. Other related keywords in use are much less popular, except for Wikipedia, where medical term SARS-CoV-2 was choosen.

In our study each considered internet platform is described separately and has its own specific bias. Data can be biased e.g. due to content presenting algorithms on the media platforms. For instance, technological giants Google, Twitter and Facebook are supposed to implement fact verification algorithms to filter out false informations. Being aware of this, computational techniques of social sciences (Jarynowski et al. [21]), despite some disadvantages and their exploratory nature, provide us with an opportunity to analyze a huge amount of digital footprint data at low cost and in short time.

GOOGLE TRENDS

Google has 95% share among Polish Internet users with over 8 billion entries per month and is the undisputed leader on the Polish Internet market (PBI [17]). Interest in novel Coronavirus on Google can be measured by the number of queries [Fig. 1]. According to Google there were $2 \cdot 10^5$ searches of term "Koronawirus" monthly. However, it is calculated based on historical data. Thus there were dozens of thousand daily searches in the late February 2020.

The subject of the corona virus occurred recently after the outbreak in Wuhan. It is important to note, that until disease introduction to Poland there were no Coronavirus related searches in top 25 Google queries at all. Although "coronavirus" related queries were observed in top trends before the introduction of the disease to Poland, they start to dominate top queries only after the massive mitigation measures were taken such as school/university and boarder closures around 09-12.03.20 [Fig. 1]. Prior to the disease introduction

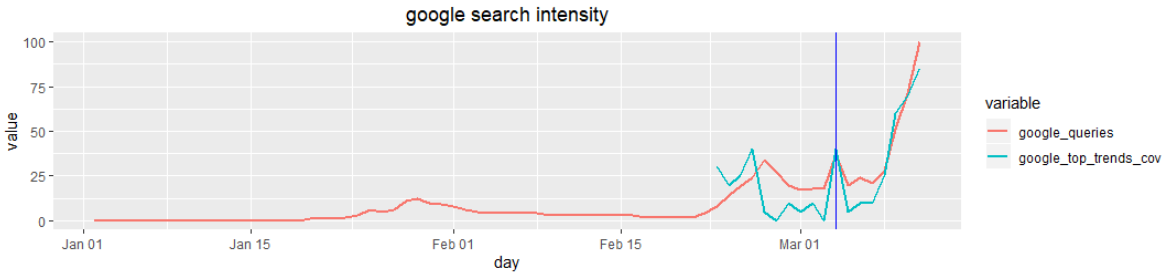

Figure 1. The intensity of searched queries with the word "corona virus"/“Koronawirus” in Polish Google (01.01-11.03.20) and percentage of “coronavius” related queries in top trend topics (23.01-11.03.20) both generated using Google Trend tool. Disease introduction marked with the blue vertical line.

two phases of interest can be distinguished [Fig. 1]:

1. from the end of January and beginning of February when the epidemic was announced and confirmed in China. We see a small peak around 25.01.20 (e.g. death of Liang Wudong) and around 29.01.20 (e.g. first case in Germany);

2. from the end of February till beginning of March (when the number of infections rapidly increased in Italy). We see a clear peak around 27.02.20 (e.g. fake news about possible introduction of the disease to Poland [22]).

After the first confirmed case in the country we can see a peak on that day (04.03.20) and substantial growth after important measures were implemented by Polish authorities (09-11.03).

People are looking for information on further epidemiological topics related to infection and epidemics as well [Fig. 2]. It should be noted that professional vocabulary such as "hand hygiene" practically does not appear in queries (below the “noise threshold” compared to other epidemiological terms [Fig. 2]). Queries related to hand hygiene had a peak in late February and beginning of March. Queries related to masks had their peak of popularity at the end of February and the lack of increase in popularity in March (compared to other epidemiological queries [Fig. 2]) may be due to the success of information campaigns on their alleged low effectiveness or simply due to the lack of the masks on the market.

Search activity for information on the infection in Poland is still much lower than in countries with high global mobility (Lai et al. [23]) and already confirmed cases of infection. Such a peripheral location of Poland (among others) lead to the first confirmed case introduced by land rather than by air (Interdisciplinary [24]).

In the corresponding semantic networks we recognize the most common co-occurring phrases in the search engine together with the word "Coronavirus" [Fig. 3]. Such a network contains information on how

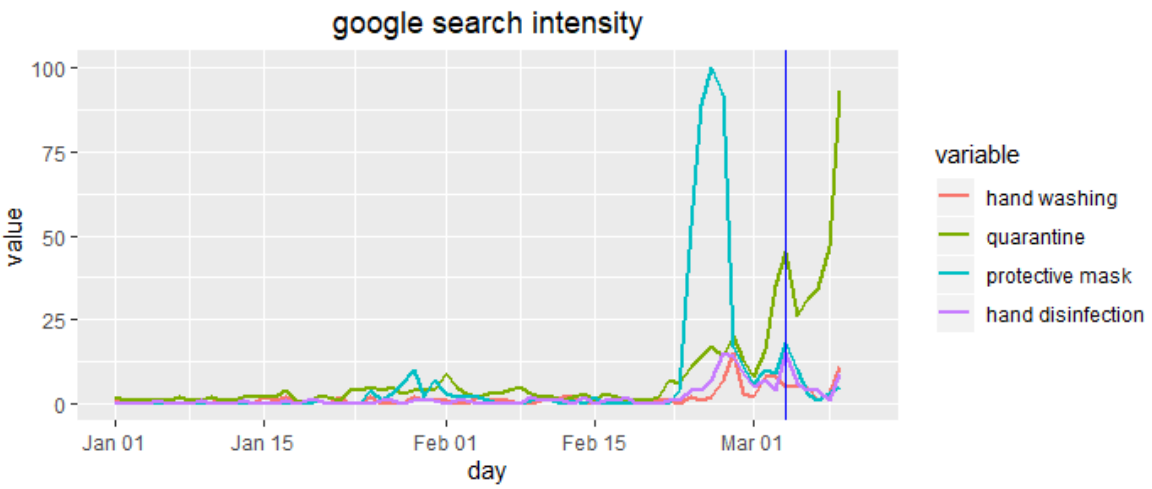

Figure 2. The intensity of queries with the phrases “quarantine”, "protective mask", "hand washing", "hand disinfection" (“kwarantanna” “maseczka ochronna”, “mycie rąk”, “dezynfekcja rąk”) in Polish Google (01.01-10.03/2020) generated using the Google Trend tool. Disease introduction marked with the blue vertical line.

Figure 3. Semantic web (prepositions) of the noun "corona virus" in Google search engine before introduction (snapshot on 29.02.20) [left] and after introduction (snapshot on 12.03.20) [right] generated using the Answer the Public tool for Poland and Polish language (AnswearPublic [26]).

the predictor - the corona virus noun - relates to its arguments in a query phrase ([25]).

Before virus introduction, we observed [Fig. 3], that most often the questions are associated with a threat, e.g. if *it is / will reach / to Poland / close to Poland; can one die / how it kills* (czy jest / dotrze w / do Polski(ce) / blisko Polski; czy można umrzeć / jak zabija). Second level searches concerns prevention, e.g. *how to prevent / guard / protect yourself* (jak zapobiec / chronić się / zabezpieczyć się). In addition, there are

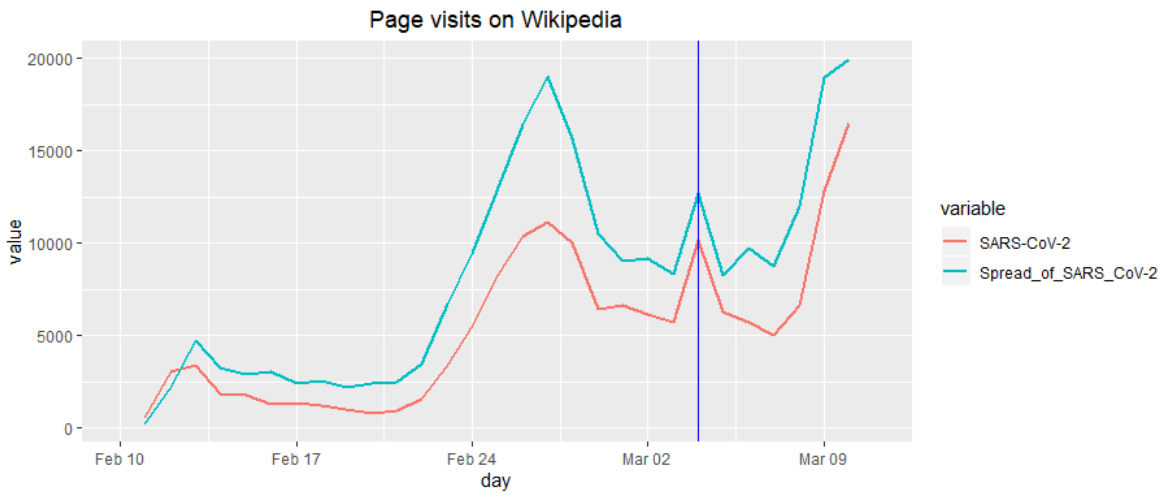

Figure 4. The number of views on the article "Spread of SARS-CoV-2 virus infection"/ “Szerzenie się zakażeń wirusem SARS-CoV-2” and “SARS-Cov-2” (10.02-10.03.2020) on Wikipedia. Disease introduction marked with the blue vertical line.

third level threads such as symptoms, history, or restrictions. The aspect of geographical proximity is also very important, thus the terms *nearby*, *near* next to dominant semantic field around the word "Coronavirus".

After virus introduction, there is domination of practical questions e.g. *how to do the test / when to call a doctor / how to treat / asymptomatic course / can you go for a walk / is it in air/ do you do shopping / leave from the employer / map* (jak zrobić test/ kiedy dzwonić po lekarza/ jak się leczyć/ przebieg bezobjawowy/ czy można iść na spacer/ czy jest w powietrzu/ czy robić zasapy/ urlop od pracodawcy/ mapy).

WIKIPEDIA

Wikipedia traffic is another indicator of the social activity. Wikipedia has an Internet coverage of 57% with over 350 million entries per month among Polish Internet users (PBI [17]). There is a significant overepresentation of users with tertiary education inhabiting big cities (affinity index>120 ([27])). We looked at the history of page views on discussions around the articles “SARS-CoV-2” ([28]) and "Spread of virus infection SARS-CoV-2"/“Szerzenie się zakażeń wirusem SARS-CoV-2” (on 11.03.2020 this page was taken down and new page about the pandemic appeared in its place ([29])) in the Polish Wikipedia [Fig. 4].

Before the disease introduction, we see a growing trend in the number of queries with a small peak around 13.02.20 (which does not appear in other media and is related probably to giving a new name to the virus and the disease) and a clear peak around 27.02.20 (e.g. a fake news about disease in Poland ([22])). The first days of March are characterized by a slight decline in interest, perhaps due to the saturation of

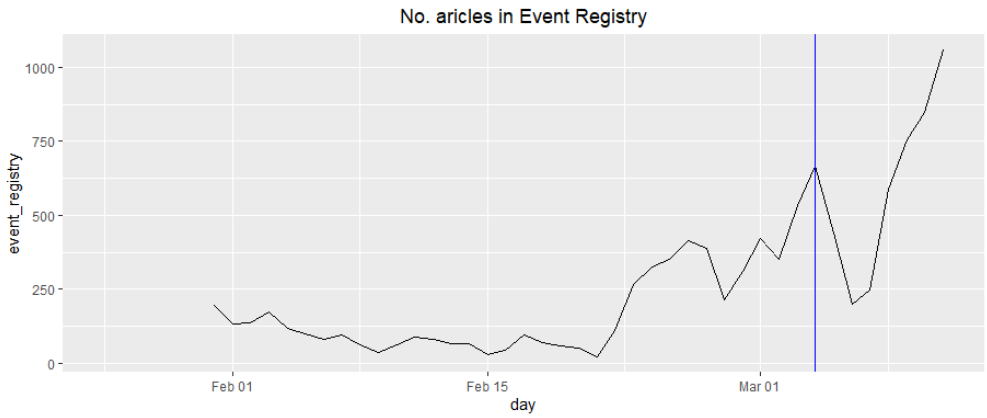

Figure 5. Number of articles in time 31.01-11.03 (generated using the Event Registry tool). Disease introduction marked with the blue vertical line.

knowledge of basic definitions about the disease in the society. After the first confirmed case, there is a small peak around this day and a slight growth during actual epidemic and massive mitigation strategies. On 04.03.20, a new page for epidemic spread in Poland was launched.

The intensive discussions between editors in Wikipedia concern, among other, the effectiveness of protective masks or the reliability of data from the PRC (People's Republic of China). No data is available before 10.02.2020, due to changes in the title of the articles together with the name of the virus and disease by WHO.

EVENT REGISTRY

We choose EventRegistry (EventRegistry [30]) as a traditional media search engine because it has a large range of online magazines representing various political sites. In addition, it gives priority to the digital versions of other broadcasting channels, including television, radio or newspapers. Between January 31 and March 11, 10755 (before 4603 and 6152 after introduction) representative articles were selected (the non-systematic sampling method was applied).

In the traditional media the weekly seasonality of the articles and 3 peaks of interest: at the end of January, the second half of February and the beginning of March can be cleary seen [Fig. 5]. Counts of news article seem to coindence with all peaks observed at other platforms and expect the peak on 26.02, ahead all of other platforms.

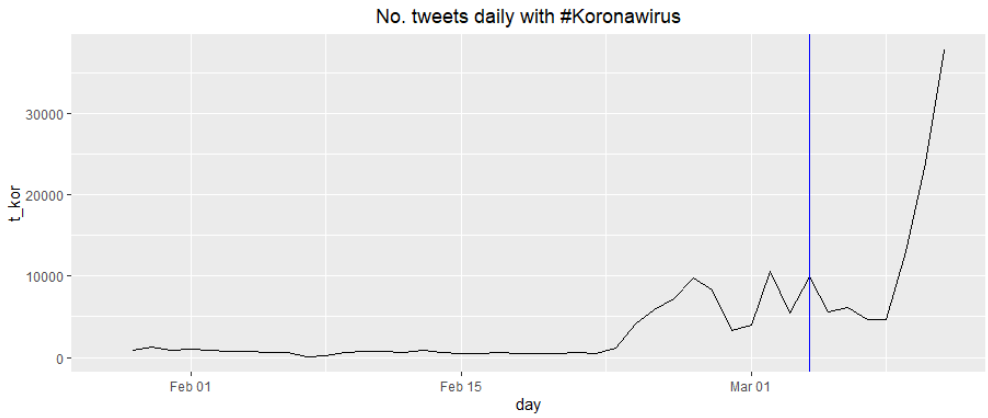

Figure 6. Number of tweets per day with the Koronawirus hashtag in Polish language (28.01-11.03.2020). Disease introduction marked with the blue vertical line

TWITTER

Twitter in Poland has relatively low popularity (~3 million registered users or less than 8% of the population) and is mainly used by expats, journalists and politicians (Sotrender [31]). However, Twitter provides an API for data acquisition available to general public almost for free. This makes possible to analyze not only content of tweets, but also their context (location, retweeting, commenting, etc.). A lot of interest in infection can be seen on Twitter in Polish (210182 tweets with #Koronawirus during 28.01-11.03.2020).

There is not so much attention on Twitter until late February 2020. There are peaks on 27.02.20 (fake news about possible case in Poland), 02.03.20 (discussion about special anti-COVID-2010 act ([32])), 04.03.20 (disease introduction) and huge increase in interest on 9-11.03.20 (massive mitigation strategies implemented).

To apply Social Network Analysis (Wasserman et al. [33]) methods to the Twitter data, we build a network with vertices representing Twitter accounts and edges representing retweets (Jarynowski et al. [34]). The network revealed various connections (social impact, trust, friendship, etc.) between accounts being social actors and the characteristics of the actors (political affiliation, views, etc.). An unsupervised weighted Louvain algorithm (Blondel et al. [35]) for community analysis was used and the vertex color denotes the community it belongs to. Retweet network [Fig. 7] shows how the discourse is divided into the ruling party (gray), opposition (orange) and the protestant religious and political group / „Idz pod prad" (light blue).

In addition, further discourse participants were identified [Fig. 7] as far right (dark blue). The subject of "Coronavirus" in Poland has a conflict-generating potential. As a consequence a dispute has emerged between the ruling party promoting information content and confirming the belief that the Polish state is

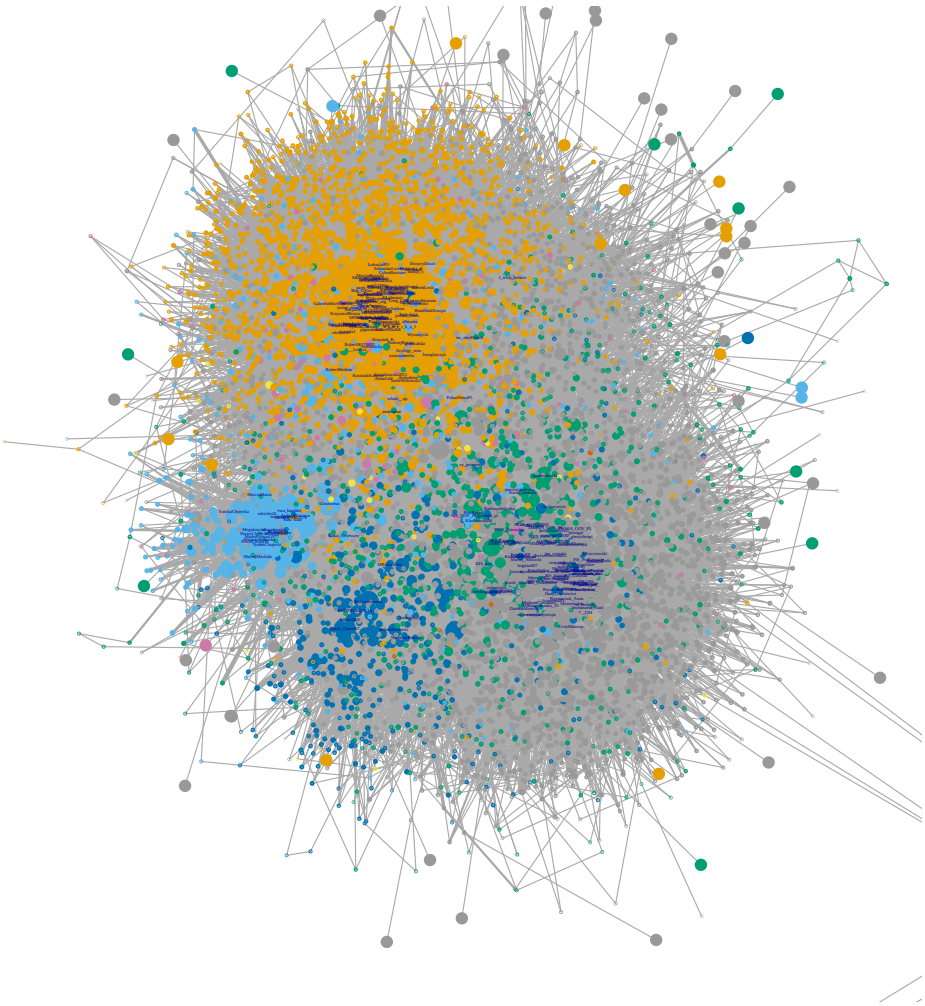

Figure 7. A network of Twitter accounts (vertices) connected by retweets (edges) with the Koronawirus hashtag. Colorcode: gray - the ruling camp, orange - the opposition, light blue - the Protestant group, dark blue - far right (28.01-11.03.2020). This network only shows accounts that have generated at least 3 tweets and connections that represent at least 2 retweets, labels are provided fo 100 most central (weighted degree centrality) .

well prepared to fight the virus (gray cluster), and the opposition negating its ability to fight the virus [Fig. 7], especially before disease introduction (Jarynowski et al. [16]). However after introduction, the most distinguished discourse is represented in the cores of the network [Fig. 7] and boundaries between communities are more blured than it was before confirmation of the first case in Poland ([16]). It could mean, that (at least in this weekly resolution) for an average Twitter user, mechanisms of solidarity and community building (for example to help neigbors and support healthworkers) are less politically driven than before.

For example, Twitter accounts already classified as potentially belonging to the so-called trolls (which in other studies were attributed to the extreme right in the context of elections to the European Parlament

Rank	item	count	item English	Rank	item	count	item English
1	polska	32266	Poland	16	informacja	12248	information
2	rzad	22500	Government	17	szpital	12217	hospital
3	przypadek	21868	case	18	szkola	12168	school
4	osoba	18723	person	19	chiny	12153	China
5	polski	18482	Polish	20	zarazic	12127	to infect
6	wszystek	17924	all	21	epidemia	12117	epidemic
7	móc	17908	can	22	wirus	11349	virus
8	zdrowie	17476	health	23	walka	10904	fight
9	wlochy	14113	Italy	24	zakazic	10888	to infect
10	czlowiek	13681	men	25	kwarantanna	10552	quarantaine
11	wszyscy	13586	all	26	potwierdzic	10423	to confirm
12	minister	13079	Minister	27	dzien	10393	day
13	test	13079	test	28	polak	10341	Pole
14	pis	12810	Law and Justice	29	panstwo	9769	country
15	ludzie	12602	people	30	zamknac	9713	to close

Table I. Counts of 30 most frequent words on Tweeter related to COVID-10 without stop words during 28.01-11.03.2020.

(OKO [36]), or to the far-left side in the context of the African Swine Fever epidemic (Jarynowski et al. [37]), promoted content in the buffer area (attacking both the ruling party and the mainstream opposition).

We also looked on the most frequent words after stemming and excluding stop words [Tab. I]. There are mainly words related to topics around health, but with politicical context as well, due to the political and journalistic bias of Twitter. In comparition to situation before the first confrimed case ([16]), there is less politics and less fear related concepts.

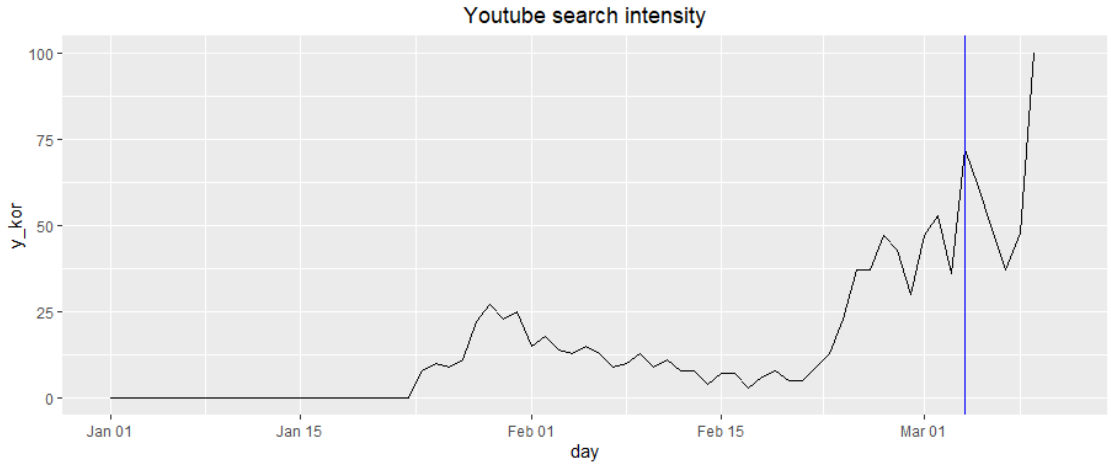

Figure 8. The intensity of queries for the word "Koronawirus" on Youtube (01.01-11.03.20) generated using the Google Trend tool. Confirmed disease introduction is marked by the blue vertical line.

YOUTUBE

Youtube has a 68% share among Polish Internet users with about 700 million entries per month (PBI [17]). In addition, streams from the mobile app should be taken into account, as it is most popular app on Poles' smartphones (PBI [17]). For our analysis, we selected videos on the main subject of "corona virus" ([38]) using keyword search. One observes multiple peaks (25.01, 31.01, 27.02, 2.03, 04.03, and 9-11.03.20) on Youtube similar to all other social media [Fig. 8].

OTHER MEDIA

The fastest growing profile in January and February 2020 in the entire Polish Facebook was "Conflicts and global disasters" („Konflikty i katastrofy światowe") which gained over 120 thousand followers in one month correlating with recent increased activity related to corona virus information. The most popular post in the Vlog category in January 2020 was a video material titled "Wuhan market" on SA Wardęga's profile (@sawardega), which was eventually marked as containing false information (Sotrender [39]).

There are other social platforms not covered by this study such as Instagram (a popular platform among teenagers with affinity index>120 in this age category ([40])). The topic of Coronavius was not so popular at Polish Instagram until the disease introduction. The most popular hashtag after introduction of the virus "odwolajcieszkolyxd" and is related to school closures.

For a better topic coverage blogs (e.g. blog Media@jesion gathers hundreds thousands entries by day ([41])) and user comments in Internet media presents additional reach information source. E.g. articles

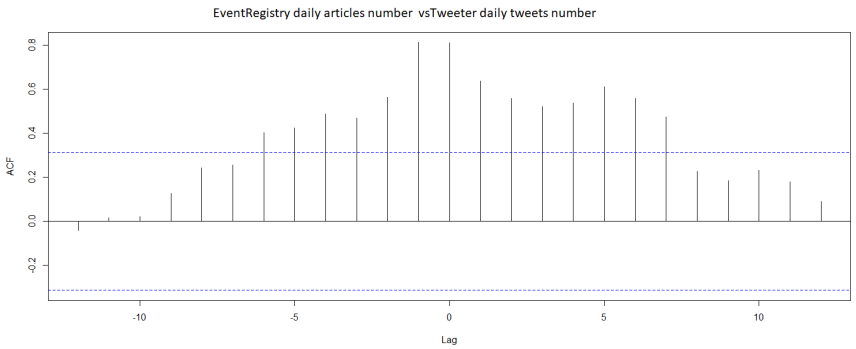

Figure 9. Lagged (in days) correlation between daily series of article counts from Event Registry and Tweets numbers.

from media such as wp.pl, onet.pl, interia.pl have billions entries monthly and their articles on Coronavirus reaches on average dozens of thousands entries with hundreds of comments ([17]).

COMPARISON OF DIFFERENT PLATFORMS

To compare the interest on COVID-19 on different Internet platforms we visualized the available queries and interest measurements as time series together and marked events important to the Polish public [Fig. 10]. From the time series as well as Google queries semantics [Fig. 3] and tags/topics in the news, we observe that geographal proximity of the disease drives a lot of interest (e.g. outbreak in Germany). We see that traditional news agencies (well represented in EventRegistry) as well as Google search could anticipate and form more distinct interest peaks than social media platforms. Moreover, discussions on social-content media (YouTube and Wikipedia) cause smoother interest trajectories [Fig. 10] than information providers (news from EventRegistry, Google). We detected time lags between different platforms for given topics (e.g news from event registry are ahead of commentatory media in fake news on possible disease introduction). It could mean that the social media dissiminate infomation via word-of-mouth speading mechanisms. On the other hand, traditional media have journalists hired to search and select most interesting topics quickly ([42]). It leads to faster respond to events by news agencies (repersented by EventRegistry) than social media. A stronger lag (1 day) is observed between EventRegitry (the most ahead of) and Twitter (the most delayed) [Fig. 9].

On Fig 10 we can distinguish 4 phases of interest in Poland ([16]):

1) Chinese phase (COVID-19 emerges in Asia ([43]));

2) Itialian phase (second wave of COVID-19 ([44]));

3) Waiting phase (awearness building and waiting for a first case in Poland);

4) Epidemic and Mitigation phase ([45]).

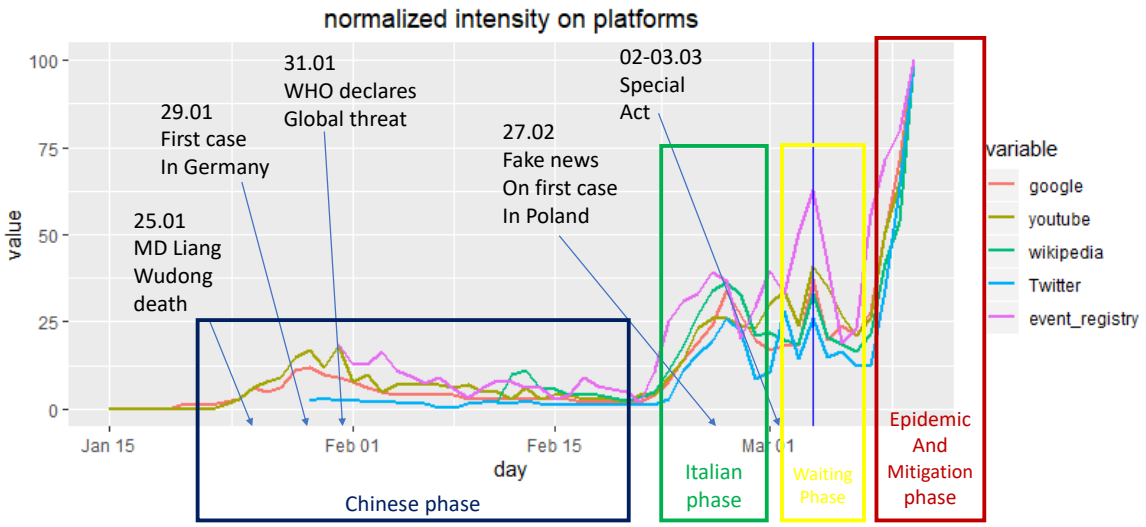

Figure 10. The intensity of topic Koronawirus on various media platforms during 15.01-11.03.2020. Time series were normalized to 100 by maximal value for a given serie. Disease introduction is marked by the blue vertical line.

Moreover, Twitter (15% penetration rate among Polish Internet users ([40])) has a significantly different time pattern than other analyzed platforms. In particular, it does not have China-related phase at all and it has at least 3-fold faster growth rate than other media in epidemic and mitigation phase.

To quantify the similarity in interest on different platforms we calculated Pearson correlations coefficients of the time series [Fig. 11]. All measured intensities are positively correlated. *Protective mask* as a signal of fear/perception of risk is less correlated with other variables more related to information needs. *Protective mask* and *washing hands* is much less correlated than *protective mask* and *hand disinfection*, which could mean, that people search for professianal solutions rather thans simple and practical ones. Moreover, high amounts of searching terms as *antiviral mask* (there is no such a medical term) implies that people are searching for colloquial terms ([16]).

CONCLUSIONS

In the face of the COVID-19 pandemic, there is an unprecedented flood of information (information 'noise') and our task was to extract important events and features from widest range of Internet media in Poland. The knowledge of quantitative characteristics of the "Coronavirus" perception in Poland is an important prerequisite for a proper crisis management (Trzos et al. [46]) such a design of protection

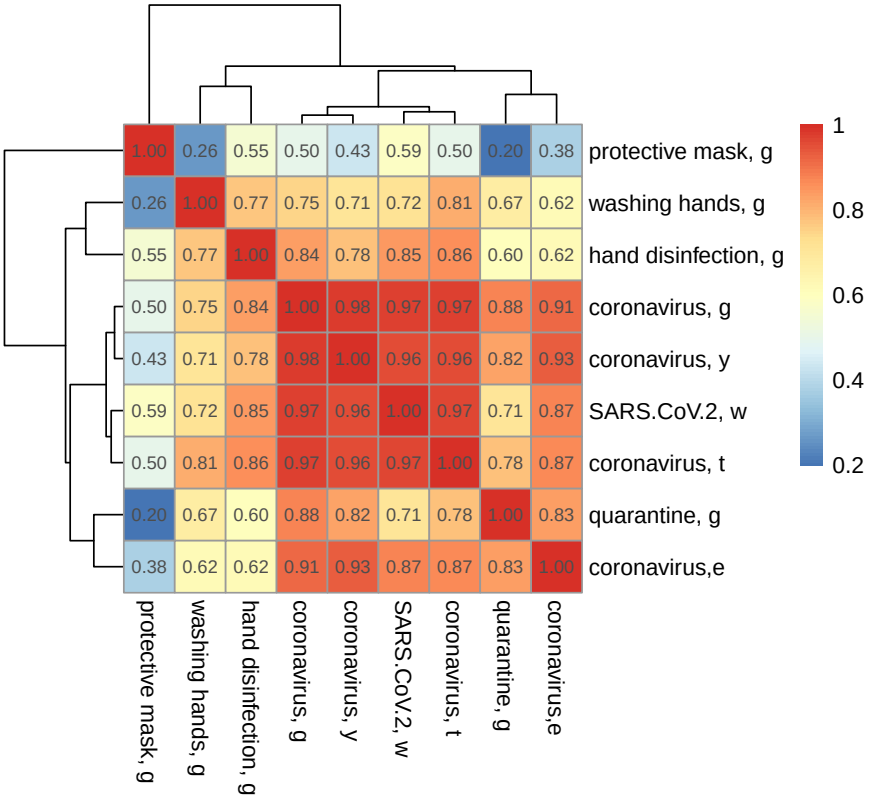

Figure 11. The Pearson's correlation matrix and the corresponding hierarhical clustering for the terms "Koronawirus"/"Covronavius" and related epidemiological queries on various media platforms (g – Google, w – Wikipedia, y – Youtube, t – Twitter, e – EventRegistry). With the significance level of 5% all correlations were significant except the pair "antiviral mask, g"/"quarantine, g". Colorcode corresponds to correlation strenght.

policies for risk management and adequate education of citizens by the stakeholders identified in Poland. For example, at present, government mitigation programs or hygiene education principles published on Twitter fall (to the great extend) into the information bubble or echo chambers ([47]) of supporters of the ruling party [Fig. 7]. Lack of interest in general public ([43, 48]) could be associated with a low epidemiological awareness of average Poles, especially when COVID-19 pandemic is massively discussed in traditional and social media by a relatively small but loud group of people.

Before official disease introduction we observed in Poland two information phases related to outbreaks in China (the end of January) and in Italy (the second half of February) and one commentary and update phase in social media in the beginning of March related to, among others, the special act on COVID-19 mitigation [Fig. 10]. Information media (Wikipedia and Google) do not display the third phase, because probably the awareness about the virus and the knowledge on the disease has already saturated, and people are interested mainly in the current update (new infomation on COVID-19) on Twitter, Youtube, or other electronic media.

After the disease introduction to Poland, there is a peak on the introduction day and fast growth of interest in epidemic and mitigation phase [Fig. 10]. Under conditions of a market-consumer society, private goals of individuals come into conflict with responsibility of the whole society or community. For example, economic consequence of supply and demand lead to price increases of medical devices ([49]).

Due to the data availability, only by analyzing Twitter we have the full control over the methodology and research techniques (Alshaabi et al. [50]). The most difficult is to access the Facebook data, despite the highest population penetration and the largest reach in Poland ([51]). Facebook does not allow automated analysis (Facebook [52]) for the public and we can only rely on manual work-intensive research of commercial companies (Sotrender [31, 39]), whose research methodology may differ from scientific standards. However, there are attempts to use available Facebook Ads campaigns in context to “Coroanavirus” [53].

In order to prepare and manage the crisis in an optimal way, a deeper perception analysis in the form of reliable quantitative and qualitative analysis is required. Especially, according to the results of empirical research, society expects institutional activities and in the event of an epidemic, it is the "state (...) that is responsible for the poor health of the population" (Taranowicz [54]). Perhaps one of the reasons the Chinese have been so successful in controlling the spread of the infection is that social media like WeChat (Wang et al. [55], Zhang et al. [56]), or Internet forums (Liu and Lu [57]) were analyzed by algorithms (Lu et al. [58], Paul and Dredze [59], Salathe et al. [60]) with a goal to mitigate the spread of COVID-19. Combining the behavioral changes detected via social media analysis with the detailed information on human mobility via e.g. mobile phone tracking (Schneider et al. [61], Brockmann et al. [62], Gonzalez et al. [63]), sophisticated computational models of infectious disease spread could be implemented (Belik et al. [64], Vespignani [65], Maier and Brockmann [66], Hufnagel et al. [67], Ferguson et al. [68], Prasse et al. [69]) allowing to simulate various scenarios and assess possible human and economic losses and properly distribute the available resources. Furthermore, impact of public information campaigns could be measured by internal surveys of public opinion. In addition, Internet media analysis could fill gaps in socio-medical research on collective actions during an important public health disruption event such as infectious diseases (Jarynowski and Belik [70]).

Acknowledgments: we thanks PNFN (2019-21), NCN (2016/22/E/HS2/00034), and FU Berlin (FU AvH: 08166500) for partial finanacial support and Łukasz Krzowski, Daniel Płatek, Ireneusz Skawina, Andrzej Buda, and Marcus Doherr for fruitful discussions.

REFERENCES

[1] IBRIS. Koronawirus ponad polityka. https://wiadomosci.onet.pl/tylko-w-onecie/wybory-prezydenckie-2020-sondaz-andrzej-duda-prowadzi-potrzebna-ii-tura/cmk8ssw, 2020. Accessed: 2020-03-04.

[2] OKO. Sondaż ipsos o koronawirusie. dwie trzecie widzi duże zagrożenie, ale nie panikujemy. https://oko.press/sondaz-ipsos-o-koronawirusie/, 2020. Accessed: 2020-03-12.

[3] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.

[4] Fred Sun Lu, Suqin Hou, Kristin Baltrusaitis, Manan Shah, Jure Leskovec, Jared Hawkins, John Brownstein, Giuseppe Conidi, Julia Gunn, Josh Gray, et al. Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the boston metropolis. *JMIR public health and surveillance*, 4(1):e4, 2018.

[5] Aditya Joshi, Ross Sparks, James McHugh, Sarvnaz Karimi, Cecile Paris, and C Raina MacIntyre. Harnessing tweets for early detection of an acute disease event. *Epidemiology*, 31(1):90–97, 2020.

[6] Anna Kata. Anti-vaccine activists, web 2.0, and the postmodern paradigm–an overview of tactics and tropes used online by the anti-vaccination movement. *Vaccine*, 30(25):3778–3789, 2012.

[7] Natasha Kassan. Disinformation and coronavirus. https://www.lowyinstitute.org/the-interpreter/disinformation-and-coronavirus, 2020.

[8] Guardian. Italian minister tries to calm coronavirus panic and attacks profiteers. https://www.theguardian.com/world/2020/feb/27/italian-minister-tries-to-calm-coronavirus-panic-and-attacks-profiteers, 2020. Accessed: 2020-03-04.

[9] Sean Jacobs. Political propaganda spread through social bots. *Media, Culture, & Global Politics*, pages 1–22, 2017.

[10] Gregg Gonsalves and Peter Staley. Panic, paranoia, and public health—the aids epidemic's lessons for ebola. *New England Journal of Medicine*, 371(25):2348–2349, 2014.

[11] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PloS one*, 6(8), 2011.

[12] Eli P Fenichel, Carlos Castillo-Chavez, M Graziano Ceddia, Gerardo Chowell, Paula A Gonzalez Parra, Graham J Hickling, Garth Holloway, Richard Horan, Benjamin Morin, Charles Perrings, et al. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311, 2011.

[13] Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch. Coupled disease–behavior dynamics on complex networks: A review. *Physics of life reviews*, 15:1–29, 2015.

[14] Loukas Samaras, Elena García-Barriocanal, and Miguel-Angel Sicilia. Syndromic surveillance using web data:

a systematic review. In *Innovation in Health Informatics*, pages 39–77. Elsevier, 2020.

[15] Sudhakar V Nuti, Brian Wayda, Isuru Ranasinghe, Sisi Wang, Rachel P Dreyer, Serene I Chen, and Karthik Murugiah. The use of google trends in health care research: a systematic review. *PloS one*, 9(10), 2014.

[16] Andrzej Jarynowski, Monika Wojta-Kempa, and Vitaly Belik. Percepcja „koronawirusa" w polskim internecie do czasu potwierdzenia pierwszego przypadku zakażenia sars-cov-2 w polsce. *accpeted: Pielegniarstwo i Zdrowie Publiczne*, 2020.

[17] PBI. Polski internet w styczniu 2020. `http://pbi.org.pl/badanie-gemius-pbi/polski-internet-w-styczniu-2020/`, 2020. Accessed: 2020-03-04.

[18] PBI. Polski internet w styczniu 2019. `http://pbi.org.pl/badanie-gemius-pbi/polski-internet-w-styczniu-2019/`, 2019. Accessed: 2020-03-04.

[19] M Tanaś, W Kamieniecki, M Bochenek, and R Lange. Raport z badania nastolatki 3.0. *Warszawa, NASK*, 2017.

[20] Andrzej Jarynowski and Vitaly Belik. Choroby przenoszone drogą płciową w dobie internetu i e-zdrowia: kalkulatory ryzyka. *Kraków: Biblioteka Jagiellońska*, 2018.

[21] Andrzej Jarynowski, Andrzej Buda, and Piotr Nyczka. *Obliczeniowe nauki spoleczne w Praktycze*. WN:Glogow, 2014.

[22] Gazeta Wyborcza. Chybicka: "w Łodzi potwierdzono koronawirusa". `https://wiadomosci.gazeta.pl/wiadomosci/7,114884,25741826,chybicka-w-lodzi-potwierdzono-koronawirusa-minister-zdrowia.html`, 2020. Accessed: 2020-03-12.

[23] Shengjie Lai, Isaac Bogoch, Nick Ruktanonchai, Alexander Watts, Yu Li, Jianzing Yu, Xin Lv, Weizhong Yang, Hongjie Yu, Kamran Khan, et al. Assessing spread risk of wuhan novel coronavirus within and beyond china, january-april 2020: a travel network-based modelling study. *medRxiv*, 2020.

[24] Interdisciplinary. Kiedy 2019-ncov trafi do polski? `http://interdisciplinary-research.eu/kiedy-2019n-cov-trafi-do-polski`, 2020. Accessed: 2020-03-04.

[25] Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. plwordnet 3.0–a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268, 2016.

[26] AnswearPublic. Answear the public. `https://answerthepublic.com`, 2020. Accessed: 2020-03-04.

[27] PBI. Wikipedia i jej użytkownicy. `http://pbi.org.pl/wp-content/uploads/2017/09/2017-09-26-Wikipedia_analiza.pdf`, 2018. Accessed: 2020-03-12.

[28] Wikipedia. Sars-cov-2. `https://pl.wikipedia.org/wiki/SARS-CoV-2`, 2020. Accessed: 2020-03-04.

[29] Wikipedia. Szerzenie sie zakazen wirusem sars-cov-2. `https://pl.wikipedia.org/wiki/Szerzenie_się_zakazen_wirusem_SARS-CoV-2`, 2020. Accessed: 2020-03-04.

[30] EventRegistry. Event registry. `https://www.eventregistry.org`, 2020. Accessed: 2020-03-04.

[31] Sotrender. Twitter w polsce – podsumowanie 2018 roku. `https://www.sotrender.com/blog/pl/2019/02/twitter-w-polsce-podsumowanie-2018-roku-infografika/`, 2019. Accessed:

2020-03-04.

[32] Sejm. Rzadowy projekt ustawy o szczegolnych rozwiazaniach zwiazanych z zapobieganiem, przeciwdzialaniem i zwalczaniem covid-19. https://www.sejm.gov.pl/Sejm9.nsf/PrzebiegProc.xsp?id=016EAA75EDD551EBC125851E0077C1C2, 2020. Accessed: 2020-03-04.

[33] Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[34] Andrzej Jarynowski, Michal B Paradowski, and Andrzej Buda. Modelling communities and populations: an introduction to computational social science. *Studia metodologiczne*, 39:117–139, 2019.

[35] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[36] OKO. Proba wplyniecia na wyniki wyborow? dwie siatki patriotycznych trolli wspieraly konfederacje. https://oko.press/proba-wplyniecia-na-wyniki-wyborow-dwie-siatki-patriotycznych-trolli-wspieraly-konfede 2019. Accessed: 2020-03-04.

[37] Andrzej Jarynowski, Daniel Platek, Łukasz Krzowski, Anton Gerylovich, and Vitaly Belik. African swine fever-potential biological warfare threat. Technical report, EasyChair, 2019.

[38] Youtube. Koronawirus. https://www.youtube.com/, 2020. Accessed: 2020-03-04.

[39] Sotrender. Facebook trends styczen 2020 – burzliwych 31 dni nowego roku. https://www.sotrender.com/blog/pl/2020/02/facebook-trends-styczen-2020-burzliwych-31-dni-nowego-roku/, 2020. Accessed: 2020-03-04.

[40] PBI. Instagram i jego polscy użytkownicy. http://pbi.org.pl/wp-content/uploads/2018/07/2018-07-Instagram.pdf, 2018. Accessed: 2020-03-12.

[41] Media@jesion. Pytania i odpowiedzi dotyczące covid-19. https://medium.com/@jesion/dlaczego-powinniśmy-się-przejmować-często-zadawane-pytania-i-odpowiedzi-dotyczące-covi 2020. Accessed: 2020-03-12.

[42] Friederike Schultz, Sonja Utz, and Anja Göritz. Is the medium the message? perceptions of and reactions to crisis communication via twitter, blogs and traditional media. *Public relations review*, 37(1):20–27, 2011.

[43] Artur Strzelecki. Infodemiological study using google trends on coronavirus epidemic in wuhan, china. *arXiv preprint arXiv:2001.11021*, 2020.

[44] Artur Strzelecki. The second worldwide wave of interest in coronavirus since the covid-19 outbreaks in south korea, italy and iran: A google trends study. *arXiv preprint arXiv:2003.10998*, 2020.

[45] Interdisciplinary. Sars-cov-2 opracowanie pierwszych ognisk. http://interdisciplinary-research.eu/sars-cov-2-opracowanie-pierwszych-ognisk, 2020. Accessed: 2020-03-16.

[46] Arkadiusz Trzos, Łukasz Krzowski, and Katarzyna Długosz. Specyfika działań ratownictwa medycznego. *Na Ratunek*, 4:17, 2017.

[47] Fabian Baumann, Philipp Lorenz-Spreen, Igor M Sokolov, and Michele Starnini. Modeling echo chambers and

polarization dynamics in social networks. *Physical Review Letters*, 124(4):048301, 2020.

[48] Widzialni. Marketing wirusowy koronawirusowa. https://www.widzialni.pl/blog/marketing-wirusowy-koronawirusa/, 2020. Accessed: 2020-03-04.

[49] Ceneo. Koronawirus atakuje sklepy. boom na maski i... nie tylko. https://subiektywnieofinansach.pl/koronawirus-atakuje-a-w-sklepach-boom-na-maski-i-nie-tylko-w 2020. Accessed: 2020-03-04.

[50] T. Alshaabi, J. R. Minot, M. V. Arnold, J. L. Adams, D. R. Dewhurst, A. J. Reagan, R. Muhamad, C. M. Danforth, and P. S. Dodds. How the world's collective attention is being paid to a pandemic: Covid-19 related 1-gram time series for 24 languages on twitter, 2020.

[51] Mobirank. Social media users in poland. https://mobirank.pl/2018/12/04/liczba-uzytkownikow-facebooka-instagrama-i-messengera-w-polsce-11-2018/, 2018. Accessed: 2020-03-04.

[52] Facebook. Facebook for developers. documentation. https://developers.facebook.com/docs, 2020. Accessed: 2020-03-04.

[53] Yelena Mejova and Kyriaki Kalimeri. Advertisers jump on coronavirus bandwagon: Politics, news, and business. *arXiv preprint arXiv:2003.00923*, 2020.

[54] Iwona Taranowicz. Zdrowie i sposoby radzenia sobie z jego zagrożeniami. *Analiza socjologiczna. Wrocław: Oficyna Wydawnicza Arboretum*, 2010.

[55] Wenjun Wang, Yikai Wang, Xin Zhang, Yaping Li, Xiaoli Jia, and Shuangsuo Dang. Wechat, a chinese social media, may early detect the sars-cov-2 outbreak in 2019. *medRxiv*, 2020.

[56] Jun Zhang, Weili Wu, Xin Zhao, and Wei Zhang. Recommended psychological crisis intervention response to the 2019 novel coronavirus pneumonia outbreak in china: a model of west china hospital. *Precision Clinical Medicine*, 2020.

[57] Chuchu Liu and Xin Lu. Analyzing hidden populations online: topic, emotion, and social network of hiv-related users in the largest chinese online community. *BMC medical informatics and decision making*, 18(1):2, 2018.

[58] Xin Lu, Shuo Qin, Petter Holme, Fanhui Meng, Yanqing Hu, Fredrik Liljeros, and Gad Allon. Beyond the coverage of information spreading: Analytical and empirical evidence of re-exposure in large-scale online social networks. *arXiv preprint arXiv:1907.12389*, 2019.

[59] Michael J Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.

[60] Marcel Salathe, Linus Bengtsson, Todd J Bodnar, Devon D Brewer, John S Brownstein, Caroline Buckee, Ellsworth M Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L Mabry, et al. Digital epidemiology. *PLoS computational biology*, 8(7), 2012.

[61] Christian M Schneider, Vitaly Belik, Thomas Couronné, Zbigniew Smoreda, and Marta C González. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84):20130246, 2013.

[62] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075): 462–465, 2006.

[63] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.

[64] Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.

[65] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature physics*, 8 (1):32–39, 2012.

[66] Benjamin F Maier and Dirk Brockmann. Effective containment explains sub-exponential growth in confirmed cases of recent covid-19 outbreak in mainland china. *arXiv preprint arXiv:2002.07572*, 2020.

[67] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences*, 101(42):15124–15129, 2004.

[68] Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, et al. Impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand. *Imperial College, London doi.org/10.25561/77482*, 2020.

[69] Bastian Prasse, Massimo A Achterberg, Long Ma, and Piet Van Mieghem. Network-based prediction of the 2019-ncov epidemic outbreak in the chinese province hubei. *arXiv preprint arXiv:2002.04482*, 2020.

[70] Andrzej Jarynowski and Vitaly Belik. Analiza kosztow rozprzestrzeniania sie afrykanskiego pomoru swin w polsce. *Public Health Forum*, V(XIII):72, 2019.