Telematics and Informatics Reports 10 (2023) 100059

Contents lists available at ScienceDirect

Telematics and Informatics Reports

journal homepage: www.elsevier.com/locate/teler

Migrants vs. stayers in the pandemic – A sentiment analysis of Twitter content

Olga Czeranowska[a,*], Karol Chlasta[b], Piotr Miłkowski[c], Izabela Grabowska[b], Jan Kocoń[c], Krzysztof Hwaszcz[d], Jan Wieczorek[c], Agata Jastrzębowska[b]

[a] *SWPS University of Social Sciences and Humanities, Warsaw, Poland*
[b] *Kozminski University, Warsaw, Poland*
[c] *Wrocław University of Science and Technology, Wrocław, Poland*
[d] *University of Wrocław, Wrocław, Poland*

ARTICLE INFO

Keywords:
Sentiment analysis
Text mining
Text analytics
Social media
Twitter
Migrants

ABSTRACT

In this paper, we propose a sentiment analysis of Twitter data focused on the attitudes and sentiments of Polish migrants and stayers during the pandemic. We collected 9 million tweets and retweets between January and August 2021, and analysed them using MultiEmo, the multilingual, multilevel, multi-domain sentiment analysis corpus. We discovered that the sentiment of tweets differs between migrants and stayers over time, and it relates to the country of migration. The general sentiment is similar for migrants and stayers, but a more detailed analysis reveals that hashtags related to staying safe and staying at home, as well as vaccinations are more polarised for migrants than for stayers, and they reflect the general development trend of the pandemic in Europe. In addition to comparing migrants with stayers, we also compared migrants staying in different countries. amongst the countries of migration, for which we collected at least 3000 tweets, the most positive sentiment of Polish migrants' tweets was observed in Belgium, with the most negative sentiment coming from Estonia. We also observed that the sentiment of tweets written in Polish by stayers in Poland is less negative when compared to Polish migrants in most of the countries with the highest number of tweets.

Introduction

This article creates intersections and interplays between big data and migration studies. It is worth noting that we bring this interplay to the table at a time when big data is recognised as a third revolution in modern social sciences (cf. [34,52,91]). Interconnecting these phenomena along with "the age of migration" [101] reveals a niche of studies to which this article further aims to contribute. We also consider the remarkable structural conditions connected to the COVID-19 pandemic which suddenly constrained geographical mobility and forced people to stay put and the effect it had on migrants' who we consider to be a 'naturally mobile' population.

Big data provides new opportunities and challenges to social sciences, especially in exploring the connections between individuals and the 'vast social structures that shape us' ([35]: 584). Big social data provides opportunities to observe how big structures evolve and eventually change, and how they are related to people's everyday lives and interpersonal relations. Halfard and Savage [36], in their critical analysis of big data opportunities and utilisations in social research, acknowledge that these data offer insights in real time and overtime into the daily lives of a large number of people – millions – that were previously unattainable, showing their interactions, emotions and behaviours to some extent. They quote arguments that the numbers speak for themselves. But they also ask – is that all? Does context matter at all? Is there still a place for conventional enquiry in this massive assemblage? They acknowledge that big data might offer opportune resources for sociological research as a tool, though not as a solution. It can help to make powerful arguments about society, its structures and members. Research shows that the digital footprints of humans left in many online environments can be successfully employed to study social and psychological outcomes [97], emotions [56], cultural fit [18], and social networking [86].

In this paper, we use digital footprints to analyse Covid-19 impact on migrants/movers. Experience of lockdown and restriction of human freedoms, according to Hobfoll's [42] resource conservation theory, may cause a loss of resources, which will be associated with reduced physical and emotional health. With lower resources, people will be less able to cope with difficult and crisis situations. An analysis of Twitter data from general populations showed that prolonged restrictions on mobility, symptoms associated with developing Covid-19, economic uncer-

* Corresponding author: SWPS University of Social Sciences and Humanities, Chodakowska 19/31, 03-815 Warsaw, Poland.
E-mail address: oczeranowska@swps.edu.pl (O. Czeranowska).

https://doi.org/10.1016/j.teler.2023.100059
Received 23 December 2022; Received in revised form 21 March 2023; Accepted 31 March 2023

tainty and a low level of quality of social interactions have a negative impact on well-being [12]. Introducing physical distancing, quarantine and staying at home to reduce the spread of the coronavirus has reduced the quality of life for families [27]. Moreover, the spread of Covid-19 has profoundly changed the daily lives of citizens by introducing new ways of working and accessing services based on smart technologies [85].

In the existing research, migrants and refugees were often analysed as a topic of the Twitter discourse (cf. [19,40,94]), but not as tweet authors. In this paper, we focus on migrants as actors in the internet discourse and their sentiments regarding the pandemic and its outcome and compare them to stayers as other distinguished actors and a reference population. By using big (social) data presented in the context of the pandemic, we are able to delve into migrants' subjectivities, relations with others, perceptions of reality, and the impacts of structural conditions on their lives. In this way we create a novel intersection between big data and migration studies. The key aim of this article is to analyse the similarities and differences of pandemic-specific sentiments between migrants who have become immobile for some time, and stayers who were less mobile even before the pandemic.

Our analysis was conducted in the context of Polish migration. Poland is one of the countries of origin of intra-EU movers, including highly-skilled movers [26]. There are also some migration flows from Poland aimed outside Europe, mainly to the USA [83]. Presently, the majority of research regarding migrants during the pandemic, focuses on the most vulnerable migrant groups from developing countries [3,6,9,24,50]. Polish migrants were (mostly) not affected by the pandemic in the same way as those groups but an unprecedented and unpredictable shift in the world of open-borders they had known influenced their lives in many spheres.

This paper consists of four parts. In the first part, we will present the theoretical background on the interactions and interdependencies between geographical and virtual mobilities. We also discuss how those have changed in the context of the pandemic. In the second part, we present our methodology with a review of studies using Twitter data, also in relation to migration. In the third section of this manuscript, we describe the process of data collection and analysis using an automated LaBSE+BiLSTM sentiment classification model. Finally, we present the analysis of our data and results with conclusions.

Theoretical background

Geographical and virtual mobility

Geographical mobility and virtual mobility can be connected in multiple ways. New communication technologies and social media have long played many roles in migration processes and migrants' lives at every stage of migration [41]. They provide migrants with fast and low-cost communications channels, through which different types of content (audio, video, pictures, text) can be sent [55]. In 2008, Diminescu [22] coined the term 'connected migrant' to describe the role of ICT in migrants' lives. He argues that migration can no longer be viewed as movement between two separated places and two independent systems of social relations. Individuals moving to another country no longer have to become uprooted, as they are able to maintain remote relations. The 21st-century migrant is defined both by mobility and connectivity: 'Yesterday the motto was: immigrate and cut your roots; today it would be: circulate and keep in touch' ([22]: 568). Moreover, with new technologies enabling people to maintain ties, transnational identities can be built [54].

Research proves that the concept of a 'connected migrant' is a good description of reality. Wilding [90] researched migrant families and showed that even in the 1990s, emails were the mode of communication that enabled them to keep in contact despite physical distance. The emergence of new forms of electronic communication enabled those families to have more frequent contacts. Similarly, Komito [54] interviewed Polish and Filipino non-nationals in Ireland about their social media practices. For most of their respondents, social network sites were a daily (or at least weekly) tool. Respondents were 'new media experts', and felt comfortable switching between them depending on the context.

Long distance communication technologies are intrinsically linked with migration processes [7]. Nowadays, social media is lowering the threshold for migration [20]. Hiller and Franz [41] identified different roles of computer use in the three stages of migration: pre-migrant, post-migrant and settled migrants. Even almost 20 years ago, when levels of internet usage and its importance in everyday lives were lower, it was considered a useful tool in planning and organising international mobility. Social media makes snowball migration easier – people living abroad can keep in touch with family and friends in their country of origin and possibly provide them with help to move as well [54]. Moreover, once a person has moved (and settled), then provided with easy access to information, they may be more likely to move again, for example due to new economic opportunities emerging in another city or country [55].

Apart from providing the information needed to become established in a new country (administrative procedures, labour market opportunities and other practical information), new communication technologies might also have an emotional and relational importance for migrants. Migrants can maintain ties with their community in the country of origin despite physical distance via the internet and social media. They can, therefore, still feel part of their national/ethnic community, even without physical contact [41,54]. This means that the psychological cost of being away from loved ones is at least lowered by the possibility of remaining present in each other's lives [53]. New technologies have changed what can be (and is) shared. Constant, real-time communication enables migrants to share both important news and – most importantly – their feelings as these emerge [22]. This constant communication enables migrants to continue to fulfil their roles in family life, for instance in the form of a transnational mothering [43,62] or fathering [74].

At the same time, new technologies provide a novel resource in building and organising the social life of the diaspora [41]. However, the easier contact with other people of the same nationality (both in the country of origin and diaspora) may mean less motivation to integrate with the receiving society [8]. For example in Komito and Bates [55], Polish migrants in Dublin were found to maintain contacts mainly with other Poles: 'In so far as loneliness would have previously increased their motivation to socialise with those who they live near, now they can maintain contact with friends and family electronically' ([55]: 13). Similarly, in van den Bos and Nell [89], national links were overrepresented in migrants' web-surfing experience. New media seems, therefore, to reinforce territoriality [89].

Migrants in times of the pandemic

Migration research in the first year of the pandemic was focused on vulnerable groups of migrants and refugees [3,6,9,24,50]. However, migrants moving between developed countries were also affected by the current situation. Being far from home in turbulent times such as the global pandemic would certainly increase an individual's level of distress. While the virus has been called 'a great equaliser', in reality, both the chances of contracting the illness and possibilities of receiving proper medical help are mediated by socially constructed impacts, such as ethnicity or immigrant status. Low quality housing and other conditions resulting from poor living standards, as well as hesitancy to vaccinate and to seek help (e.g. due to undocumented status) might result in more severe forms of illness [21]. Depending on their legal status, migrants may also have limited access to the host country's health system. Additionally, limited language skills may also create barriers in accessing important information on the current situation and on the precautions which should be taken into consideration.

Moreover, the pandemic, similarly as other kinds of crises, resulted in growth of social disparities and heightened risk of discrimination of groups considered as 'outgroups' [33,49]. Migrants are therefore in general more affected by the pandemic than non-migrants, both in terms of health risk and other consequences of the pandemic e.g. economic or social ones [95].

In March/April 2020, when the Covid-19 pandemic caused 'a global mobility deadlock' ([102]: 1) many people became "stranded" far from home and their loved ones [87]. Nearly all the international borders were closed for non-essential travel. Additionally, various measures were issued for exceptional travel. Despite the mobility restrictions and measures implemented by governments, nationals returning to their home countries were frequently excluded from the travel bans [102]. Some countries even helped their citizens in organising their return migration. In the Polish context, this took the form of the 'Lot do domu' (eng.: 'Flight back home') action which was aimed at bringing Polish citizens, together with their spouses and children, back to Poland. Special charter flights were organised from selected oversea destinations (UK, Ireland, Cyprus, Malta, USA) as well as less popular destinations. Fees for these flights were meant to be subsidised by the government (https://www.pot.gov.pl/). Still, 'Lot do domu' program sparked some controversies regarding costs of tickets and overbooking.

For those who could not go back to the country of origin (or did not want to, due to higher level of integration into the receiving society or other personal circumstances), virtual mobility become the only option of contacting their family and friends back home [10,11,87]. During lockdowns, even contacts and mobility within the same city became limited and virtual mobility became the only option of seeing people outside one's household, working or shopping [11,95].

Apart from the physical health risks, pandemic also had wide and long lasting consequences for mental well-being and psychological health [79,81,16,73]. Because of limited social support in the country of residence and distance separating them from families and friends in the country of origin, migrants (especially vulnerable migrants) may also be more at risk of experiencing psychological and social consequences of the pandemic, such as loneliness [25,60,72]. The WHO ApartTogherther survey showed that migrants reported worsening mental health during the pandemic. Respondents were feeling more depressed, worried, anxious, lonely, angry, stressed, irritated, hopeless, had more sleep-related problems and used more drugs and alcohol [92].

Twitter as tool and topic in social sciences

Twitter data have been studied increasingly more extensively within the field of social sciences as it proves valuable in analysing users' opinions, ideas, emotions and views in a broad range of subjects. This is especially valid when investigating large-scale attitudes towards particular events or topics [1,5,30,38,61,66,88]. Important advantage of Twitter data is their suitability for sentiment analysis which enables research focused on public emotions [5]. Such analyses may provide a key source of information for reaching certain decisions, for instance, at government level (e.g. framing actions aimed at shaping public procedures and policies related to social distancing in the current Covid-19 crisis - [80]). Importantly, in the case of migration studies, Twitter can provide us with data sent from every place on the globe (there is no issue regarding different providers, etc. - [37]). Moreover, unlike the majority of official statistics, Twitter data can be available in real time, without a time lag [99].

Twitter data are especially useful for research during various crises when social media are used as a source of information, support systems and a place for sharing emotions and opinions [5,70,77]. Consequently, Twitter has also been used as a data source for research concerning Covid-19, especially as the pandemic itself enhanced the use of social media [68]. For example, at the beginning of the pandemic, Arpaci et al. [4] used evolutionary clustering to analyse public reactions to the new coronavirus. Further, in the spring of 2020, Xue et al. [93] used machine learning in the infodemiology study on twitter data. Geolocalisation was used by Gharavi, Nazemi and Dagostari [29] for mapping the warning signs of pandemic outbreaks in the US, and by Huang et al. [45] for testing the efficacy of mobility restrictions. Twitter has been important as a public discourse space in studies focused on conspiracy theories [2,32,84]. With the introduction of the Covid-19 vaccines, a new kind of conspiracy theories emerged, and a new debate started, which was also analysed with the use of Twitter data [103,108]. There have also been studies explicitly focusing on the attitude towards vaccines, e.g. Mir and Sevukan [68] analysed sentiments expressed by Indian people towards Covid vaccines, Kwok, Vadde, and Wang [57] studied Australian Twitter while Mahanti et al. [63] and Mir, Rathinam and Gul [69] did a comparative analysis of various countries. Moreover, since 21 January 2020, tweets connected with the pandemic have been collected in the COVID-19-TweetIDs GitHub repository (2020). Although the dataset is multilingual, authors admit that it is biassed in favour of English tweets. COVID-19 stimulated researchers to go deeper into Twitter data and explore also issues related to immigration (cf. [75] and [76,104]). Rowe and collaborators [75,76] studied the impact of pandemic on sentiment and attitudes towards immigrants and they found evidence on a growing social polarisation concerning migration, showing high concentrations of strongly positive and strongly negative sentiments.

Web 2.0 has brought about a "renaissance of geographic information" [47,84]. New forms of data provide various opportunities for researchers interested in the geographical mobility of humans. Hawelka at al. [37] demonstrated that the number of visitors estimated for different countries based on Twitter data converges with the official statistics on international tourism. They also confirmed that Twitter data shows similar statistical properties as other mobility datasets. However, only a small proportion (estimated at 1% in 2013) of Twitter data is geotagged, as users can disable the GPS function on their mobile devices. In 2009 Twitter introduced per-tweet geotagging (apart from per-user geotagging). This functionality can associate each tweet with latitude and longitude, but was used even less than per-used geotagging [14] and consequently the function was disabled in 2019. On the other hand, the percentage of content with locations listed by users is higher, but some of those are very general, unclear and/or sometimes false/imaginary [39].

Due to the low percentage of geotagged data, researchers have tried other techniques in order to "locate" social media data. Some of these strategies use place-based language – topics which are supposed to be more popular in some localities (e.g. local news), as well as the use of lexical variations typical for certain places [13,14,23,46,48]. However, this approach rooted in sociolinguistics and dialectology would be less useful in locating data created by migrants. Compton, Jurgens and Allen [17] were able to establish the localization of a significant percentage of Twitter users based on their friends' localizations. Again, this social network-based approach to geotagging would not be an ideal solution for migrants' tweets, as friends remaining in the country of origin may constitute a considerable part of their social media networks (cf. [89]). In situations where exact locations are needed (not for the academic purposes, but for 'life or death' matters such as crisis management in cases of earthquakes, or other natural disasters), special software has been developed, such as Stanford Named Entity Recognition. Still, even these programmes are not one hundred percent accurate [28]. Location can be also determined by IP addresses, but these are unreliable due to the dynamic allocation of IP addresses by Internet Service Providers. Moreover, virtual private networks may disguise users' real locations [51].

Another important methodological issue in Twitter-based research is selectivity connected with social media data. The global representation of Twitter data is affected by the demographic profile of Twitter users [58]. Internet coverage and usage in itself is not universal, with some groups being more frequently excluded. Additionally, profiles of the 'typical uses' of each social media are different, which may result in disparate or varying research outcomes [98]. Migrants' access and use of social media depends on their cultural, socio-economic and linguistic backgrounds, as well as general computer literacy, with more vulnerable groups facing greater challenges. Still, [67] argue that 'coverage bias' (bias connected with unequal access to the internet as well as unequal digital literacy) is generally lower for migrants than for the general population because of migrants' high online presence described in the previous section.

Methodology

About the data/project

The data presented in this article were collected within the project "(IT)Mobility. Immobility of the mobile, mobility of the immobile - migrants in the times of pandemics and new information/communication technologies" (Institute of Social Sciences, SWPS University of Social Sciences and Humanities). The aim of the project was to analyse the situation of Polish migrants during the Covid-19 pandemic, with special focus on interconnections between virtual and geographical mobility. Migrants were chosen as the researched group because they can be defined as the 'most mobile' individuals within the society. We were interested in how those who are 'the most mobile' would deal with the forced immobility resulting from the pandemic. Moreover, we assumed that migrants, apart from geographical mobility, may also be more accustomed to various forms of virtual mobilities, which they could have been using even before the pandemic, for instance, to keep in touch with the loved ones in their country of origin (according to the 'connected migrants' hypothesis - [22]).

'Migrants *vs* stayers'

As we position this analysis on the junction of social sciences and informatics we would like to describe and justify who our 'respondents' are. Due to the very low percentage of the geolocalized content in tweets returned by Twitter API (1.44% of values not empty, with only 0.24% of tweets with coordinates for the tweets gathered in Q1 2021), we decided to use an alternative strategy for migrants' identification.

We propose a two-stage process to identify the immigrants on Twitter and distinguish them from stayers. First, we validate Twitter profiles using Botometer API[1] to exclude the bot profiles (with bot index greater than 3.5). Second, within the genuine human profiles, three annotators identified migrants, based on their self-declared profile location and the twitter language. We believe that although there are obvious challenges with the quality of the self-declared information on Twitter profiles (e.g. users stating imaginary locations like 'Milky Way' or 'San Escobar', cf. [39]), we were able to identify migratory profiles in the dataset based on the direct suggestions from the Twitter users (e.g. location of 'Kraków-Poland, now Edinburgh-Scotland'), as well as the country flags added to the profile. This is in line with the prior research of Graham et al. [31] stating that the profile locations tell us much about how users perceive, present, and place themselves, rather than an exact location.

Therefore our proxy of 'the migrant status' was a self-declared (single or multiple) location outside Poland in the user profile and the Twitter language set to Polish. Such users were treated as Poles who had emigrated. We assumed that the change of location for a foreign one would not happen in case of a short stay abroad (for example, holidays or business trips) but rather in case of a more permanent or long-term move that would imply some degree of immigrant integration. Consequently, users whose set location was Poland and who had Polish as their Twitter account language were identified as Polish stayers. Finally, the dataset was annotated with unified countries, based on the unified location of a user's profile (e.g. the self-declared location of 'Gdansk' was considered to be 'Poland'). This unification was necessary to allow for data analysis using a country dimension.

We are aware that this strategy for identifying migrant vs stayers profiles is not perfect, but with a very low level of geotagged content we decided that it can be used as a sufficient proxy for the location. Levels of geolocated content being extremely low were mentioned in earlier studies, e.g. [37] and [71]). Consequently, strategies alternative to geotagging were implemented in numerous studies for the localization of the tweets (e.g. [5,68,96,100]).

Data collection

We collected our dataset using a server provisioned by Microsoft Azure cloud. The instance type was B2s, hosted in Stockholm. The operating system of our server was Windows (version 10.0.17763). The machine had the following key software packages installed: (1) Microsoft SQL Server 2019 GDR build version 15.0.2080.9, (2) R version 4.0.3 (2020–1010) (3) Rtweet package version 0.7.0.

The data release contains two flat files extracted from the Microsoft SQL Server 2019 tables. The physical model of our database is driven by the output standard available from querying Twitter's API [64] with the Rtweet R package. Each tweet was recorded with 89 data elements (fields) that were pre-selected as useful for the potential analysis.

All the tweets were pre-processed in R. The pre-processing did not change the content of the data. It was only done to adjust the data types in order to make them compatible with the requirements of the SQL Server (e.g. the database fields representing the tweet publication time were shortened to the first 50 characters, per each tweet).

The data also underwent a post processing stage, when the following additional dimensions were added to the database:

- the sentiment scores of each tweet, using MultiEmo sentiment analysis tool;
- the origin country of Twitter accounts, added manually on the basis of the 'Location' field provided by the Rtweet R package - migrants were defined as users with location outside of Poland and user's language set to Polish;
- supplementary fields, added to facilitate ad-hoc data analysis, calculated from other fields in the dataset (e.g. the field called WeekOfTheYear, based on the tweet's publication date).

The post-processed results were analysed in Microsoft Power BI Desktop.

Data description

Our Twitter analysis presented in this paper encompasses tweets gathered between 1 January and 30 August 2021. The data was gathered based on 34 hashtags relating to the COVID-19 pandemic and migration through Twitter API (cf. [64]). Queries run on the Twitter API were executed independently for each day, using a filter on the current date. After the exploratory analysis, only the first set of hashtags (pandemic-related) was included in the analysis. We identified hashtags through a brainstorming in our team consisting of both sociologists, psychologists, and information technologists - we made a list based on our expertise. We also checked manually for the popularity and co-occurrence of hashtags. The full list of the pandemic hashtag names and their translations from Polish can be found in the Table 2 below.

We collected a total of 9 058 194 tweets in two tables (Emigrants and Tweets) refreshed using custom R scripts connecting to Twitter API. The first table (Emigrants) focused on Migrants' tweets (so tweets generated by non-bot, humans, whose language was set to Polish and the location was set to a place outside of Poland. The second table (Tweets) focused on tweets related to 34 hashtags. Apart from the genuine user tweets, our tweet database included:

- 5 327 155 retweets – instances where somebody else's tweet was posted into a user's feed to allow more people to see and respond to it;
- 316 136 quotes – cases where a tweet is shared with an additional comment.

[1] Botometer at Indiana University https://botometer.osome.iu.edu/

Table 1
Data description in SQL Server tables and Power BI.

	Table: Tweets	Table :Emigrants
Content Description	All tweets returned by Twitter API queried by selected hashtags until Aug 30th, 20,201	All tweets published by selected users identified as migrants until Aug 30th, 2021
Number of Columns	92 (includes the originating hashtag)	91
Additional Information	Field indicating a hashtag provided to Twitter API to get a tweet record (text + metadata)	Table populated through a Twitter API query based on a vector of user profiles, manually identified as migrants (Twitter language = Polish, user location outside of Poland.
Size (uncompressed)	7.625 GB	6.479 GB
Number of Tweets	4,629,501	4,428,693
Including Retweets	2,858,418	2,468,737
Including Quotes	126,122	190,014
Frequency of Refresh	Daily	Daily
Analysed Period	Jan - Aug 15th, 2021	Feb - Aug 15th, 2021
Tweets in Analysed Period	4,405,128	4,134,710
Tweets by Language (1% or more)	English (67.29%), Russian (7.84%), Polish (6.93%), German (6.56%), Japanese (3.59%), Italian (2.68%), Dutch (1.91%), Indonesian (1.31%)	Polish (55.49%), English (30.53%), Spanish (2.27%), Japanese (2.55%), Russian (2.32%), German (1.58%)

Table 2
Full list of Twitter hashtags.

Hashtag	English Translation
#emigracja	Emigration
#lockdown	Lockdown
#lotdodomu	fly home
#maseczki	Masks
#narodowakwarantanna	national quarantine
#NarodowyProgramSzczepien	National Vaccination Programme
#otwieraMY	WE are opening [businesses]
#otwieramy	we are opening [businesses]
#pracazdalna	remote work
#remotework	Remotework
#stayathome	Stayathome
#staysafe	Staysafe
#szczepienia	Vaccination
#SzczepimySie	Let's Vaccinate
#szczepimysie	let's vaccinate
#szczepionka	Vaccine
#wfh	Wfh [work from home]
#zostanwdomu	stay at home

As with all Twitter data, certain well-known caveats always apply when using such data, including the influence of bots and disinformation [82]. We identified 146 bots using the Botometer service (formerly BotOrNot; [78]) provided by the Observatory on Social Media (OSoMe) at Indiana University . This tool allowed us to check the prior activity of a given Twitter account, and index it with a score between 0 and 5, based on the likelihood it was run by bots. Higher scores indicate that a profile is more bot-like. We assume that a genuine human migrant has a score of less than, or equal to, 3.5.

Sentiment analysis

We used sentiment analysis, a natural language processing (NLP) approach, to classify the main sentiments of a given Twitter message, such as in Martinez-Camara et al. [65]. The approach used to calculate the sentiment scores changed during the course of the study from Chlasta [105] to MultiEmo [106] — a tool for multilingual sentiment analysis which supports more than 100 languages. As per the initial tests, this new approach turned out to be more suitable for our sentiment analysis task than the method originally anticipated. The key advantage was that it did not require the semi-manual update of dictionaries (of positive and negative words). The methodological advantage of MultiEmo was that the model has already been validated. Cross-language experiments carried out on the model in 11 languages proved that LaBSE embeddings with an additional attention layer implemented in the BiLSTM architecture outperformed other methods [109].

Table 3
Sentiment of tweets (stayers vs migrants) by hashtag from January 2021 to July 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	64,547	567,538	221,572	527,478	1,381,135	61	488	160	465	1174	1,382,309
lotdodomu	1	26	8	403	438				10	10	448
maseczki	45	887	87	342	1361	1	32	4	12	49	1410
NarodowyProgramSzczepien	79	1393	518	6049	8039	3	37	2	79	121	8160
otwieraMY	1770	9580	5533	52,889	69,772	130	560	274	2076	3040	72,812
pracazdalna	28	44	104	586	762		1	5	23	29	791
remotework	4630	6041	31,147	233,291	275,109	6	3	20	164	193	275,302
stayathome	9784	75,336	59,881	96,208	241,209	3	47	41	52	143	241,352
staysafe	31,141	250,149	412,064	346,945	1,040,299	124	1613	2700	490	4927	1,045,226
szczepienia	137	2904	992	5523	9556	3	113	47	113	276	9832
SzczepimySie	9942	38,548	21,343	82,674	152,507	247	1369	607	1740	3963	156,470
szczepionka	117	1918	327	2954	5316	3	69	13	84	169	5485
wfh	11,265	43,579	58,744	135,407	248,995	16	34	64	277	391	249,386
workfromhome	33,678	31,695	121,348	150,668	337,389	21	15	120	255	411	337,800
zostanwdomu	1111	1603	644	3373	6731	8	39	13	41	101	6832
Total	168,275	1,031,241	934,312	1,644,790	3,778,618	626	4420	4070	5881	14,997	3,793,615

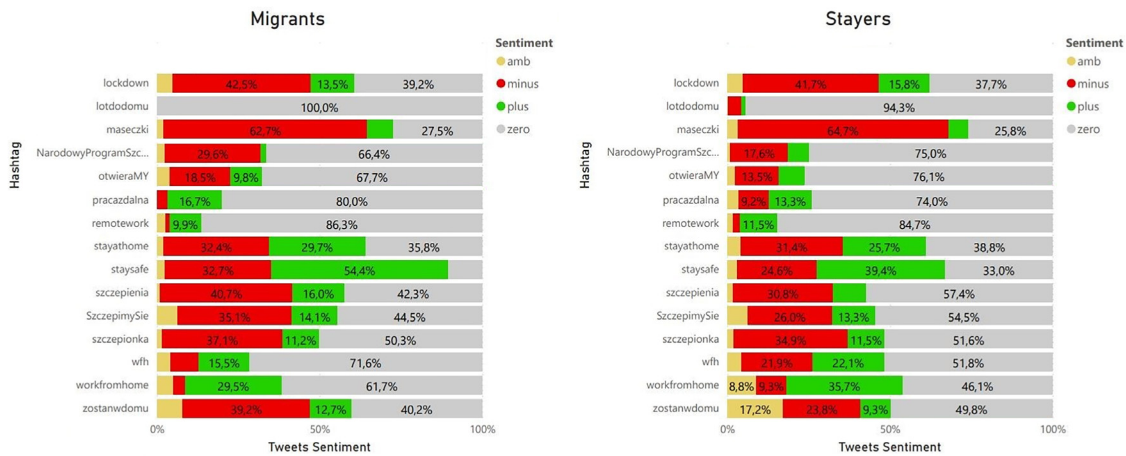


Fig. 1. Migrants vs. stayers - Sentiment by Hashtag.

The input document can be analysed using MultiEmo at the level of the whole text, individual paragraphs or sentences. The training corpus for MultiEmo was built upon 8216 review texts in Polish related to four domains: hotels, medicine, universities, and products [107]. The documents used in creating the tool contained 57,466 sentences. Sentiment annotations were performed manually by three independent annotators per element: at the level of a sentence and at the level of a full-text document. A Positive Specific Agreement value [44] was ultimately achieved at the level of 0.88 for sentences and 0.91 for texts. After that, the collection of documents was translated into 11 languages (Chinese, Dutch, English, French, German, Italian, Japanese, Polish, Portuguese, Russian and Spanish) with the use of a DeepL neural machine translation service.

The current classification model utilises LaBSE – a language-agnostic cross-lingual sentence embedding for 109 languages (Feng et al. 2020). The result of the model is a sentiment distribution at the selected text level of analysis. MultiEmo is publicly available under the Creative Commons Attribution 4.0 International Licence [106].

In total, after the step of prepossessing the raw data in MS SQL Server tables, our final data set in Power BI included 4 405 128 tweets. These included both tweets and retweets, as well as quotes (retweets with a comment). The flat files annotated with sentiment were loaded to Microsoft Power BI for the final analysis and visualisation of the dataset.

Results

During the research period, the most popular hashtag in our dataset was #lockdown. Its popularity, however, kept decreasing with time, and at the end of the research period it was almost as popular as #staysafe. The hashtag #otwieramy (eng. 'we are opening up [the businesses]'), connected with the protest against lockdown restrictions, was most popular at the beginning of 2021, but its popularity systematically decreased over the research period. Figure showing the popularity of each of the hashtags over time can be seen in the Appendix 1.

Particularly negative sentiment was identified in the tweets related to protective masks (#maseczki). Nearly 65% of all the tweets found for the hashtag #maseczki were evaluated as negative by our sentiment analysis algorithm. The second hashtag which stands out is #lockdown, for which around 42% of tweets were evaluated as having a negative sentiment. MultiEmo identified a particularly positive sentiment in 40% of tweets related to #staysafe. Another hashtag which showed particularly positive sentiment was #workfromhome. Nearly 36% tweets were evaluated as positive.

In terms of the migrants vs. stayers analysis, the general sentiments were similar, however migrants showed more of both positive and negative sentiments and stayers more of neutral ones. Moreover, it is apparent that the sentiment for some hashtags differs more strongly between migrants and stayers than for others (Fig. 1). A few examples are:

- #szczepienia (eng. 'vaccinations'), where just over 40% of tweets published by migrants were evaluated as having a negative sentiment. For stayers, the percentage of negative tweets was only 30.8%.
- #zostanwdomu (eng. 'stay at home'), where nearly 40% of tweets published by migrants were evaluated as having a negative sentiment, vs just under 24% negative for stayers.
- #staysafe, where 33% of tweets published by migrants were evaluated as having a negative sentiment, vs nearly 25% negative for stayers. Migrants were also more positive about this hashtag (54.4%) vs stayers (39.4%).

Additionally, Fig. 2 presents the general sentiment of tweets per week. Migrants tend to become less positive over time, whereas stayers seem to become more negative. (Fig. 3)

We further analysed the top 20 receiving countries, and looked at their sentiment structure of tweets (written in Polish, as recognized by Twitter). We discovered that the structure of tweet sentiment coming from stayers in Poland is less negative than Polish migrants in most countries (Fig. 3), with the exception of Belgium, where it is similar to Poland (22% vs 23%, respectively). Meanwhile, 23% of the tweets attributable to stayers in Poland were evaluated as having a positive sentiment, which is a higher score than for Poles in Germany (22%), France (22%), China (22%), Italy (22%), Netherlands (21%), Canada (19%), Ireland (17%), Island (17%), Czech Republic (17%), Greece (16%), Wales (14%) and Estonia (13%).

Discussion

Our findings show the changing popularity of particular pandemic-related hashtags, mirroring the changing emphasis on those issues in social life. This result confirms the high topicality of Twitter data (and social media data in general) and their links to current news. Moreover, we can see that the focus of Twitter discussions (at least regarding listed hashtags) was mostly on protective measures (masks, lockdowns and working from home). Amongst those, masks and limitations of mobility were seen negatively and working from home and generally 'staying safe' as less limiting had more positive connotations.

Secondly, in general, migrants' and stayers' tweets with pandemic-related hashtags showed similar sentiments, according to MultiEmo. However, migrants' tweets had higher positive and negative scores, while stayers' were more neutral. Consequently, some of the hashtags #zostanwdomu (eng. 'stay at home'), #szczepimysie (eng. 'we are vaccinating ourselves'), #szczepienia (eng. 'vaccinations'), #staysafe,

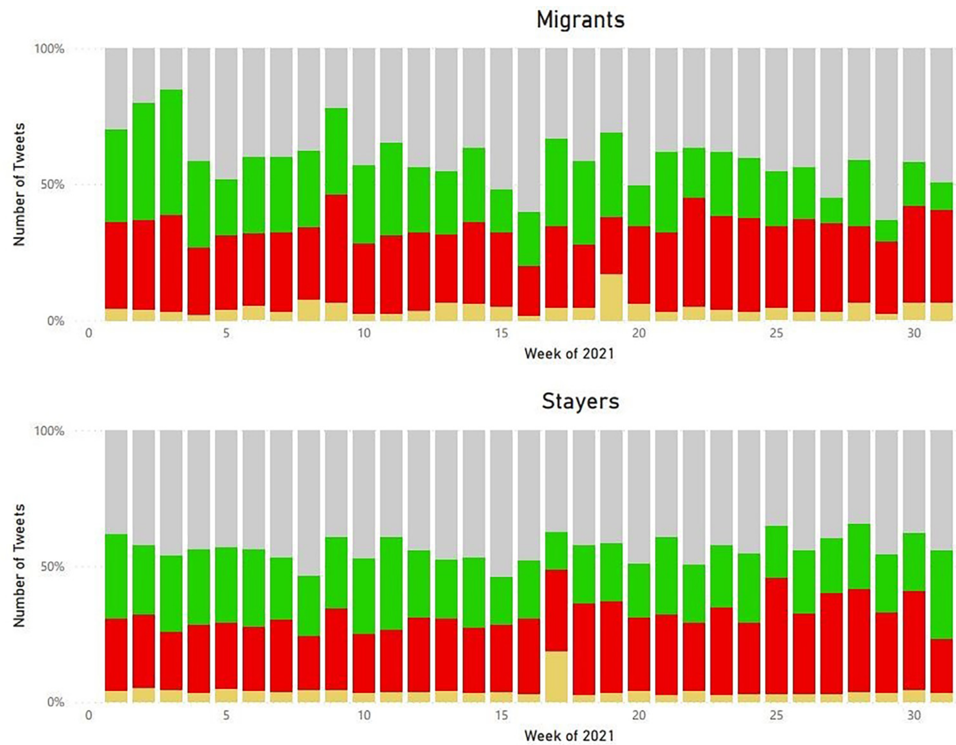


Fig. 2. Sentiment (migrants vs. stayers) by Week.

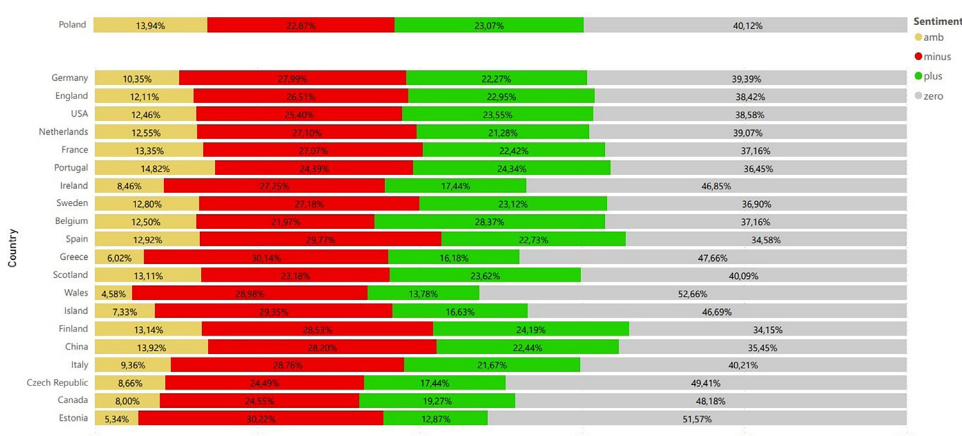


Fig. 3. Migrants vs. stayers by Country.

#otwieram y(eng. 'we are opening up [the businesses]'), #narodowyprogramszczepien (eng. 'national vaccination program') were more polarised for migrants than for stayers. The more polarised character of migrants' tweets may be connected with the fact that they were at risk of being more affected by the pandemic than the stayers [95]. Even not counting the health risk (heightened by limited knowledge of the receiving country's healthcare and sometimes also restricted access to it, cf. [59]), migrants were more likely to be affected by the economic consequences of the pandemic as well as different forms of discrimination [33,49,95]. Also, they had additional stressors in the distance and limited contact with their family and friends in the country of origin due to limits in international mobility [10,11,87]. Consequently, migrants' emotional well-being and mental health may be more at risk during the pandemic [25,60,72,92], resulting in more polarised sentiments in relation to various aspects of the Covid-19 pandemic and the limitations or protective measures resulting from it. Additionally, for migrants who stay in contact with their country of origin, social media such as Twitter are an important platform for sharing their experiences and opinions on actual topics. Therefore, more emotional investment in those online discussions may also result in more polarised sentiments.

Moreover, the sentiment was related to the country of migration, with the top three most positive countries of migration for Polish-speaking migrants being Belgium (28%), Portugal (24%), and Finland (24%), and the top three most negative counties being Greece (30%), Estonia (30%), and Spain (30%). This may depend on the severity of the pandemic situation of each of the countries but also on how the local policies and measures were seen by migrants (also in reference to the ways in which the Polish government dealt with the pandemic).

Our study is not without limitations, the main one being connected with the identification of migrants vs stayers. The challenges connected with the availability of Twitter's geo-tagged data motivated us also to use the alternative approach, exploiting the profile locations provided by tweet's authors themselves. These may be less specific, but are sufficient to establish the country a person is tweeting from which offers a novel approach as compared to the available scholarship (e.g. [15,37,58]). In connection with the language used, we were able to identify at least those Polish migrants who were still using Polish for their Twitter presence. We propose therefore a new operational definition of migrants on Twitter, based on a combination of their self-declared Twitter location and language used. Still, we are aware that this methodology of migrants vs stayers identification is not perfect, as there are many users who do not provide information on localization or give information that may not be true (cf. [39]).

Another limitation is related to the timing of the project. The project was started in late 2020, therefore, we do not have the data covering the initial months of the pandemic. Our data covers therefore mainly the period when pandemic has already become a prolonged 'new reality'.

Conclusion

Despite described limitations, our study provides a new approach to quantitative migration studies using automated sentiment analysis. While in the existing Twitter studies regarding migration, migrants are treated as a topic of the Twitter discourse [19,40,94], we were able to focus on their role as actors in the conversation. This is, to our knowledge, the first study on the Twitter dataset aiming to analyse migrants' tweets in particular. Additionally, we also focused on migration experiences of a group, which is currently under-researched in the context of Covid-19 pandemic, that is migrants from developed, EU countries.

The outcomes of our study can be used to understand migrants' sentiments and experiences connected to the Covid-19 pandemic. Social media data can be useful to policymakers and health authorities both in the migrants' countries of origin and the receiving countries. Taking into consideration the importance of social media for migrants' lives and the general role of communication via social media during the crises such as the pandemic, Twitter can be seen both as a source of information on the migrants' population and as a platform for communication of health policies and combating misinformation. Twitter represents a global virtual repository of specific social interactions within and between various social groups - in our case between migrants and stayers. We wanted to predominantly define migrants or movers by attributing to them various proxies but as a value added to it we also managed to define stayers as a reference group. We also exercised established and new measures of sentiment scores and their interpretations.

Further research is needed to refine the methodology behind identifying migrants' accounts on Twitter. As the rise in the percentage of geotagged content in the future is not probable, other methods, using self-declared Twitter location, language and other available data, should be tested. Moreover, it would be interesting to repeat our sentiment analysis on migrants from other countries to compare their reactions to the Covid-19 pandemic. Another possible, promising direction of future research is including in the analysis also other social media, preferably combining multiple of them in order to take into account different characteristics of users (demographics but also different popularity of each social media in different countries).

Despite limitations mentioned above, we believe that using Twitter data gives a more comprehensive understanding of the digital footprints and traces of migrants on Twitter stamped by time and space. We also believe that our approach and the analyses presented in this article are extra arguments for using the potential of Twitter data, which is an objective and freely accessible source of data for migration studies.

Table 1, Table 3, Table 4, Table 5, Table 6, Table 7, Table 8, Table 9, Table 10

Funding

This paper was prepared within the project "(IT)Mobility. Immobility of the mobile, mobility of the immobile - migrants in the times of pandemics and new information/communication technologies" (Institute of Social Sciences, SWPS University of Social Sciences and Humanities). The project was financed by the Ministry of Science and Higher Education in Poland under the 2019–2022 program „Regional Initiative of Excellence", project number 012 / RID / 2018/19 of the day 18.12.2018, amount of funding 11 985 347,00 pln.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

CRediT authorship contribution statement

Olga Czeranowska: Conceptualization, Funding acquisition, Project administration, Writing – original draft, Writing – review & editing. **Karol Chlasta:** Conceptualization, Funding acquisition, Data curation, Formal analysis, Software, Visualization, Writing – original draft, Writing – review & editing. **Piotr Miłkowski:** Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Izabela Grabowska:** Conceptualization, Writing – original draft, Writing – review & editing. **Jan Kocoń:** Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Krzysztof Hwaszcz:** Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Jan Wieczorek:** Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing. **Agata Jastrzębowska:** Writing – original draft, Writing – review & editing.

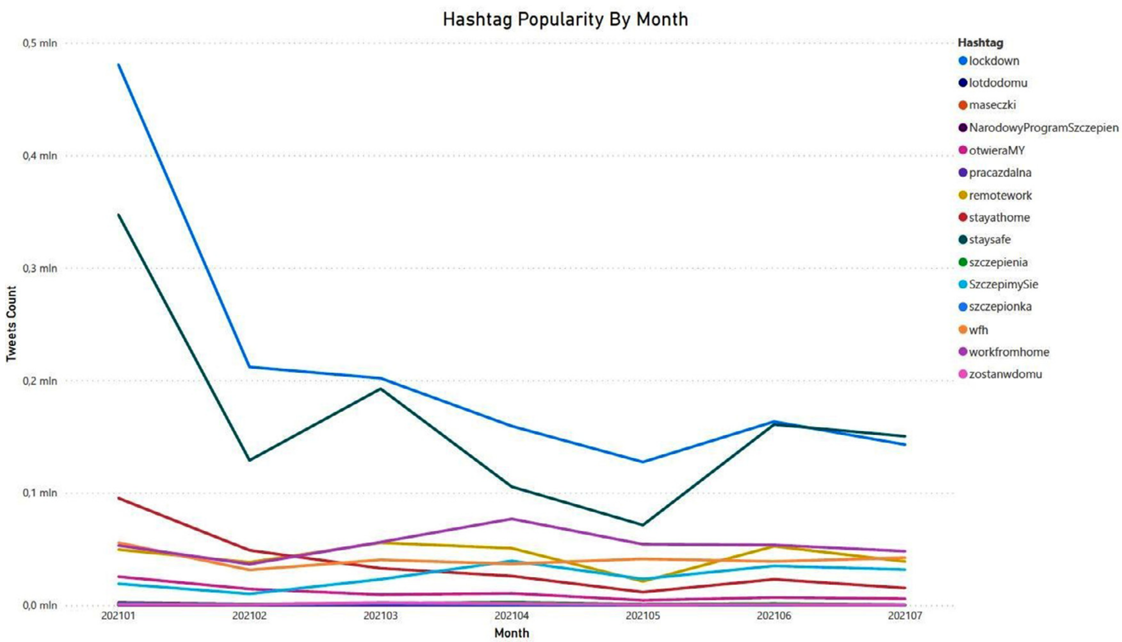
Data availability

The data that has been used is confidential.

Acknowledgements

We thank Piotr Adamowicz for his assistance with data collection and analysis.

Appendix figure



Appendix 1

Table 4
Sentiment of tweets (stayers vs migrants) by hashtag in January 2021.

Hashtag	Stayers Amb.	Neg.	Pos.	Neutr.	Sum	Migrants Amb.	Neg.	Pos.	Neutr.	Sum	Total
Lockdown	25,897	156,453	94,473	203,403	480,226	28	173	70	179	450	480,676
Lotdodomu				2	2						2
Maseczki		53	10	4	67		3	1	2	6	73
NarodowyProgramSzczepien	37	695	195	1401	2328	2	30	2	29	63	2391
otwieraMY	848	3111	2700	17,888	24,547	34	172	98	658	962	25,509
Pracazdalna	11		31	85	127			2	8	10	137
Remotework	967	1029	5507	42,012	49,515	2		4	27	33	49,548
stayathome	4805	24,359	27,405	38,665	95,234	2	17	30	30	79	95,313
staysafe	14,326	64,260	138,619	127,933	345,138	39	768	1111	180	2098	347,236
szczepienia	14	592	62	888	1556		18	2	25	45	1601
SzczepimySie	1052	5276	2289	10,040	18,657	32	152	38	310	532	19,189
szczepionka	37	594	62	678	1371	1	21	3	13	38	1409
wfh	2154	15,316	15,510	22,655	55,635	3	4	9	54	70	55,705
workfromhome	2560	4427	23,325	22,820	53,132	3	4	15	49	71	53,203
zostanwdomu	173	122	70	436	801		2	1	4	7	808
Total	52,881	276,287	310,258	488,910	1,128,336	146	1364	1386	1568	4464	1,132,800

Table 5
Sentiment of tweets (stayers vs migrants) by hashtag in February 2021.

Stayers						Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	10,959	74,408	39,470	86,764	211,601	18	90	30	69	207	211,808
lotdodomu		2		2	4						4
maseczki	17	86	4	70	177		6		1	7	184
NarodowyProgramSzczepien	7	108	53	918	1086		2		19	21	1107
otwieraMY	482	3058	1625	8580	13,745	58	208	106	416	788	14,533
pracazdalna	4	8	9	105	126			1	2	3	129
remotework	598	815	4385	32,441	38,239	1		2	56	59	38,298
stayathome	1481	9626	7303	30,527	48,937		4	2	6	12	48,949
staysafe	4670	26,430	53,226	43,684	128,010	36	259	512	80	887	128,897
szczepienia	22	172	56	879	1129		2	2	7	11	1140
SzczepimySie	702	1466	1024	6870	10,062	12	36	18	114	180	10,242
szczepionka	14	189	36	422	661		6	1	11	18	679
wfh	1354	3764	7659	18,650	31,427	2	3	7	55	67	31,494
workfromhome	1141	2746	14,940	17,756	36,583	3	1	6	41	51	36,634
zostanwdomu	103	129	81	358	671	1	4	4	2	11	682
Total	21,554	123,007	129,871	248,026	522,458	131	621	691	879	2322	524,780

Table 6
Sentiment of tweets (stayers vs migrants) by hashtag in March 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	10,579	75,143	30,633	85,242	201,597	5	94	31	93	223	201,820
lotdodomu	1		7	5	13						13
maseczki	21	358	23	77	479	1	12	1	3	17	496
NarodowyProgramSzczepien	4	141	94	1319	1558	1	1		7	9	1567
otwieraMY	182	1686	624	6765	9257	16	54	16	196	282	9539
pracazdalna	6	11	25	205	247			1	7	8	255
remotework	723	1117	6251	47,488	55,579	2	1	6	60	69	55,648
stayathome	1292	16,176	6737	8690	32,895	1	6	2	11	20	32,915
staysafe	4416	41,546	93,810	51,823	191,595	21	276	554	118	969	192,564
szczepienia	17	453	116	998	1584	3	21	5	23	52	1636
SzczepimySie	1172	4754	2640	14,042	22,608	32	190	104	218	544	23,152
szczepionka	17	537	99	981	1634		14	3	32	49	1683
wfh	2799	3735	10,601	23,134	40,269	5	9	20	104	138	40,407
workfromhome	1792	3788	22,305	28,208	56,093	5	2	23	90	120	56,213
zostanwdomu	262	872	212	963	2309	6	23	3	23	55	2364
Total	23,283	150,317	174,177	269,940	617,717	98	703	769	985	2555	620,272

Table 7
Sentiment of tweets (stayers vs migrants) by hashtag in April 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	5810	77,765	18,652	56,980	159,207	3	38	9	61	111	159,318
lotdodomu		1		4	5						5
maseczki		211	30	121	362		7		2	9	371
NarodowyProgramSzczepien	26	140	112	1907	2185				20	20	2205
otwieraMY	168	716	312	9057	10,253	10	60	26	374	470	10,723
pracazdalna	6	17	20	70	113		1		3	4	117
remotework	746	1002	5086	43,901	50,735	1	2	4	9	16	50,751
stayathome	834	9904	7193	8093	26,024		11	4	2	17	26,041
staysafe	1640	35,353	34,063	33,908	104,964	23	117	294	31	465	105,429
szczepienia	46	764	376	1398	2584		28	16	27	71	2655
SzczepimySie	2162	9486	5114	21,758	38,520	56	345	186	394	981	39,501
szczepionka	29	239	57	376	701	1	10	2	12	25	726
wfh	1782	5007	8319	21,693	36,801	4	11	14	52	81	36,882
workfromhome	24,991	8020	17,692	26,096	76,799	4	4	18	60	86	76,885
zostanwdomu	221	245	133	957	1556	1	7	4	12	24	1580
Total	38,461	148,870	97,159	226,319	510,809	103	641	577	1059	2380	513,189

Table 8
Sentiment of tweets (stayers vs migrants) by hashtag in May 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	4479	61,703	18,148	43,013	127,343	4	34	4	22	64	127,407
lotdodomu		20	1	3	24						24
maseczki	4	118	17	35	174		4	2	2	8	182
NarodowyProgramSzczepien	3	137	48	157	345		4			4	349
otwieraMY	84	677	234	3195	4190	10	44	26	198	278	4468
pracazdalna	1	2	6	46	55			1		1	56
remotework	435	617	2657	17,596	21,305			1	6	7	21,312
stayathome	400	4545	3667	3162	11,774		2	1	1	4	11,778
staysafe	1238	18,894	27,057	23,911	71,100	3	99	118	37	257	71,357
szczepienia	16	390	247	376	1029		18	14	11	43	1072
SzczepimySie	2050	6171	4219	10,262	22,702	54	270	118	292	734	23,436
szczepionka	10	212	31	236	489	1	9	2	10	22	511
wfh	1493	6915	7921	24,984	41,313	2	3	5	6	16	41,329
workfromhome	1805	5176	19,899	27,382	54,262	4	2	28	11	45	54,307
zostanwdomu	145	127	61	295	628		1	1		2	630
	12,163	105,704	84,213	154,653	356,733	78	490	321	596	1485	358,218

Table 9
Sentiment of tweets (stayers vs migrants) by hashtag in June 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	5403	97,194	16,438	44,210	163,245	2	46	15	31	94	163,339
lotdodomu		3		383	386				10	10	396
maseczki	2	57	3	33	95				2	2	97
NarodowyProgramSzczepien	2	143	16	298	459				2	2	461
otwieraMY	6	324	34	6420	6784	2	22	2	212	238	7022
pracazdalna		6	11	62	79				3	3	82
remotework	1035	1281	6244	43,965	52,525			3	5	8	52,533
stayathome	767	9580	6550	6232	23,129		6	2	2	10	23,139
staysafe	4107	47,110	54,209	55,122	160,548	1	93	93	38	225	160,773
szczepienia	18	505	121	933	1577		24	8	20	52	1629
SzczepimySie	2228	9783	5425	16,796	34,232	51	326	135	354	866	35,098
szczepionka	10	134	39	239	422		8	2	6	16	438
wfh	1502	8219	7481	21,767	38,969		4	8	6	18	38,987
workfromhome	1223	7001	20,395	24,990	53,609	2	2	30	4	38	53,647
zostanwdomu	182	92	84	306	664		1			1	665
Total	16,485	181,432	117,050	221,756	536,723	58	532	298	695	1583	538,306

Table 10
Sentiment of tweets (stayers vs migrants) by hashtag in July 2021.

	Stayers					Migrants					
Hashtag	Amb.	Neg.	Pos.	Neutr.	Sum	Amb.	Neg.	Pos.	Neutr.	Sum	Total
lockdown	1420	24,872	3758	7866	37,916	1	13	1	10	25	37,941
lotdodomu				4	4						4
maseczki	1	4		2	7						7
NarodowyProgramSzczepien		29		49	78				2	2	80
otwieraMY		8	4	984	996				22	22	1018
pracazdalna			2	13	15						15
remotework	126	180	1017	5888	7211				1	1	7212
stayathome	205	1146	1026	839	3216		1			1	3217
staysafe	744	16,556	11,080	10,564	38,944	1	1	18	6	26	38,970
szczepienia	4	28	14	51	97		2			2	99
SzczepimySie	576	1612	632	2906	5726	10	50	8	58	126	5852
szczepionka		13	3	22	38		1			1	39
wfh	181	623	1253	2524	4581			1		1	4582
workfromhome	166	537	2792	3416	6911						6911
zostanwdomu	25	16	3	58	102		1			1	103
Total	3448	45,624	21,584	35,186	105,842	12	69	28	99	208	106,050

References

[1] O.Y. Adwan, M. Al-Tawil, A.M. Huneiti, R.H. Al-Dibsi, R.A. Shahin, A.A. Abu Zayed, Twitter sentiment analysis approaches: a survey, Int. J. Emerg. Technol. Learn. (iJET) (2020) 15.

[2] W. Ahmed, J. Vidal-Alaball, J. Downing, F.L. Seguí, COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data, J. Med. Internet Res. 22 (5) (2020) e19458.

[3] Q. Alemi, C. Stempel, H. Siddiq, E. Kim, Refugees and COVID-19: achieving a comprehensive public health response, Bull. World Health Organ. 98 (8) (2020) 510.

[4] I. Arpaci, S. Alshehabi, M. Al-Emran, M. Khasawneh, I. Mahariq, T. Abdeljawad, A.E. Hassanien, Analysis of twitter data using evolutionary clustering during the COVID-19 pandemic, Comput. Mater. Continua 65 (1) (2020) 193–204.

[5] S. Bashir, S. Bano, S. Shueb, S. Gul, A.A. Mir, R. Ashraf, N. Noor, Twitter chirps for Syrian people: sentiment analysis of tweets related to Syria Chemical Attack, Int. J. Disaster Risk Reduct. 62 (2021) 102397.

[6] D. Bhandari, Y. Kotera, A. Ozaki, S. Abeysinghe, M. Kosaka, T. Tanimoto, COVID-19: challenges faced by Nepalese migrants living in Japan, BMC Public Health 21 (1) (2021) 1–14.

[7] M. Borkert, P. Cingolani, V. Premazzi, The State of the Art of Research in the EU On the Uptake and Use of ICT By Immigrants and Ethnic Minorities (IEM), European Commission, Joint Research Centre, Institute for Prospective Technological Studies, Siviglia, 2009 available on http://ftp://s-jrcsvqpx101p.jrc.es/pub/EURdoc/JRC53398.pdf (retrived on 10/02/2018). DOI, 10, 28811.

[8] M. Brekke, Young refugees in a network society, Mobility and place: enacting Northern European peripheries (2008) 103–114 https://www.taylorfrancis.com/chapters/edit/10.4324/9781315595757-18/young-refugees-network-society-marianne-brekke.

[9] M. Buheji, B. Mavrić, G. Beka, T.C. Yein, Alleviation of refugees COVID-19 pandemic risks-a framework for uncertainty mitigation, Int. Bus. Res. 13 (7) (2020) 69–79.

[10] E. Cabalquinto, 'Without technology we'd be very stuck': ageing migrants' differential (im) mobile practices during a lockdown, Media Int. Australia (2022) 1329878×221095582.

[11] E. Cabalquinto, "Come on, put viber, we can drink coffee together": performing (Im) mobile intimacy in turbulent times among aging migrants, Commun. Culture Critique 15 (2) (2022) 244–260.

[12] T. Carpi, A. Hino, I. Stefano, G Porro, The Impact of COVID-19 on subjective well-being: evidence from twitter data, J. Data Sci. (2022) 1–20 10.6339/22-JDS1066.

[13] Chandra, S., Khan, L., & Muhaya, F.B. (2011, October). Estimating twitter user location using social interactions–a content based approach. In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing (pp. 838–843). IEEE.

[14] Z. Cheng, J. Caverlee, K. Lee, You are where you tweet: a content-based approach to geo-locating twitter users, in: Proceedings of the 19th ACM international conference on Information and knowledge management, 2010, pp. 759–768.

[15] Z. Cheng, J. Caverlee, K. Lee, D. Sui, Exploring millions of footprints in location sharing services, in: Proceedings of the International AAAI Conference on Web and Social Media, 2011, pp. 81–88. Vol. 5, No. 1.

[16] K.R. Choi, M.V. Heilemann, A. Fauer, M. Mead, A second pandemic: mental health spillover from the novel coronavirus (COVID-19), J. Am. Psychiatr. Nurses Assoc 26 (4) (2020) 340–343.

[17] R. Compton, D. Jurgens, D. Allen, Geotagging one hundred million twitter accounts with total variation minimization, in: 2014 IEEE international conference on big data (big data), IEEE, 2014, pp. 393–401.

[18] C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, C. Potts, No country for old members: user lifecycle and linguistic change in online communities, in: Proceedings of the 22nd international conference on World Wide Web, 2013, pp. 307–318.

[19] A.S. de Rosa, E. Bocci, M. Bonito, M. Salvati, Twitter as social media arena for polarised social representations about the (im) migration: the controversial discourse in the Italian and international political frame, Actinide Speciation High Ionic Strength Media: Exp. Model. Approaches Predict. Actinide Speciation Migr. Subsurf., [Proc. Am. Chem. Soc. Symp. Exp. Model. Stud. Actinide Speciation Non-Ideal Syst.] 9 (3) (2021) 1167–1194.

[20] R. Dekker, G. Engbersen, How social media transform migrant networks and facilitate migration, Global Netw. 14 (4) (2014) 401–418.

[21] D. Devakumar, S.S. Bhopal, G. Shannon, COVID-19: the great unequaliser, J. R. Soc. Med. 113 (6) (2020) 234–235.

[22] D. Diminescu, The connected migrant: an epistemological manifesto, Social Sci. Inf. 47 (4) (2008) 565–579.

[23] J. Eisenstein, B. O'Connor, N.A. Smith, E. Xing, A latent variable model for geographic lexical variation, in: Proceedings of the 2010 conference on empirical methods in natural language processing, 2010, pp. 1277–1287.

[24] M. Falkenhain, U. Flick, A. Hirseland, S. Naji, K. Seidelsohn, T. Verlage, in: Setback in Labour Market Integration Due to the Covid-19 crisis? An explorative Insight On Forced migrants' Vulnerability in Germany, European Societies, 2020, pp. 1–16.

[25] T. Fokkema, R.O. Ciobanu, Older migrants and loneliness: scanning the field and looking forward, Eur. J. Ageing 18 (3) (2021) 291–297.

[26] E. Fries-Tersch, M. Jones, L. Siöland, Annual Report On intra-EU Mobility, Publications Office of the European Union, Luxembourg, 2021.

[27] M. Gayatri, M. Puspitasari, The impact of COVID-19 pandemic on family well-being: a literature review, Family J. (2022) 106648072211310, doi:10.1177/10664807221131006.

[28] J. Gelernter, N. Mushegian, Geo-parsing messages from microtext, Trans. GIS 15 (6) (2011) 753–773.

[29] Gharavi, E., Nazemi, N., & Dadgostari, F. (2020). Early outbreak detection for proactive crisis management using twitter data: covid-19 a case study in the us. *arXiv preprint arXiv:2005.00475*.

[30] S.A. Golder, M.W. Macy, Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures, Science 333 (2011) 1878–1881.

[31] M. Graham, S.A. Hale, D. Gaffney, Where in the world are you? Geolocation and language identification in Twitter, Prof. Geogr. 66 (4) (2014) 568–578.

[32] A. Gruzd, P. Mai, Going viral: how a single tweet spawned a COVID-19 conspiracy theory on Twitter, Big Data Soc. 7 (2) (2020) 2053951720938405.

[33] L. Guadagno, Migrants and the COVID-19 pandemic: An initial Analysis, International Organization for Migration, Geneva, Switzerland, 2020.

[34] J. Guan, S. Yao, C. Xu, H. Zhang, Design and implementation of network user behaviors analysis based on hadoop for big data, in: International Conference on Applications and Techniques in Information Security, Springer, Berlin, Heidelberg, 2014, pp. 44–55.

[35] A. Halavais, Bigger sociological imaginations: framing big social data theory and methods, in: Inf. Commun. Soc., 18, 2015, pp. 583–594.

[36] S. Halford, M. Savage, Speaking sociologically with big data: symphonic social science and the future for big data research, Sociology 51 (6) (2017) 1132–1148.

[37] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, C. Ratti, Geo-located Twitter as a proxy for global mobility patterns, Cartogr. Geogr. Inf. Sci 41 (3) (2014) 260–271.

[38] N. Heaivilin, B. Gerbert, J.E. Page, J.L. Gibbs, Public health surveillance of dental pain via Twitter, J. Dent. Res. 90 (2011) 1047–1051.

[39] B. Hecht, L. Hong, B. Suh, E.H. Chi, Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2011, pp. 237–246.

[40] T. Heidenreich, J.M. Eberl, F. Lind, H. Boomgaarden, Political migration discourses on social media: a comparative perspective on visibility and sentiment across political Facebook accounts in Europe, J. Ethn. Migr. Stud. 46 (7) (2020) 1261–1280.

[41] H.H. Hiller, T.M. Franz, New ties, old ties and lost ties: the use of the internet in diaspora, New Media Soc. 6 (6) (2004) 731–752.

[42] S. Hobfoll, Conservation of resources theory: its implication for stress, health, and resilience, Oxford Handb. Stress Health Coping (2012), doi:10.1093/oxfordhb/9780195375343.013.0007.

[43] P. Hondagneu-Sotelo, E. Avila, I'm here, but I'm there" the meanings of Latina transnational motherhood, Gender Soc. 11 (5) (1997) 548–571.

[44] G. Hripcsak, A.S. Rothschild, Agreement, the f-measure, and reliability in information retrieval, J. Am. Med. Inform. Assoc. 12 (3) (2005) 296–298.

[45] X. Huang, Z. Li, Y. Jiang, X. Li, D. Porter, Twitter reveals human mobility dynamics during the COVID-19 pandemic, PLoS One 15 (11) (2020) e0241957.

[46] Y. Huang, D. Guo, A. Kasakoff, J. Grieve, Understanding US regional linguistic variation with Twitter data analysis, Comput. Environ. Urban Syst. 59 (2016) 244–255.

[47] A. Hudson-Smith, M. Batty, A. Crooks, R. Milton, Mapping for the masses: accessing Web 2.0 through crowdsourcing, Soc. Sci. Comput. Rev. 27 (4) (2009) 524–538.

[48] Y. Ikawa, M. Enoki, M. Tatsubori, Location inference using microblog messages, in: Proceedings of the 21st international conference on the world wide web, 2012, pp. 687–690.

[49] ILO (2020). Covid-19 and the World of Work: impacts and Policy Responses 2020.

[50] Kassem, I.I. (2020). Refugees besieged: the lurking threat of COVID-19 in Syrian war refugee camps. Travel medicine and infectious disease.

[51] S. Kinsella, V. Murdock, N O'Hare, I'm eating a sandwich in Glasgow" modelling locations with tweets, in: Proceedings of the 3rd international workshop on Search and mining user-generated contents, 2011, pp. 61–68.

[52] R. Kitchin, Big Data, new epistemologies and paradigm shifts, Big Data Soc. 1 (1) (2014) 2053951714528481.

[53] L. Komito, The Experience of Virtual Communities: cosmopolitan or Voyeur?, in: P. O'Donovan, L. Rascaroli (Eds.) The Cause of Cosmopolitanism: Dispositions, Models, Transformations, Peter Lang, 2010, pp. 135–149.

[54] L. Komito, Social media and migration: virtual community 2.0, J. Am. Soc. Inf. Sci. Technol. 62 (6) (2011) 1075–1086.

[55] L. Komito, J. Bates, Virtually local: social media and community among Polish nationals in Dublin, in: Aslib Proceedings, Emerald Group Publishing Limited, 2009.

[56] A.D. Kramer, J.E. Guillory, J.T. Hancock, Experimental evidence of massive-scale emotional contagion through social networks, Proc. Natl. Acad. Sci. 111 (24) (2014) 8788–8790.

[57] S.W.H. Kwok, S.K. Vadde, G. Wang, Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: machine learning analysis, J. Med. Internet Res. 23 (5) (2021) e26953.

[58] K. Leetaru, S. Wang, G. Cao, A. Padmanabhan, E. Shook, Mapping the global Twitter heartbeat: the geography of Twitter, First Monday (2013).

[59] A. Liem, C. Wang, Y. Wariyanti, C.A. Latkin, B.J. Hall, The neglected health of international migrant workers in the COVID-19 epidemic, Lancet Psychiatry 7 (4) (2020) e20.

[60] M. Lim, A. Van Hulst, S. Pisanu, L. Merry, Social isolation, loneliness and health: a descriptive study of the experiences of migrant mothers with young children (0–5 Years Old) at La Maison Bleue, Front. Global Women's Health 3 (2022) 823632.

[61] J.B. Lowe, M. Barnes, C. Teo, S. Sutherns, Investigating the use of social media to help women from going back to smoking post-partum, Aust. N. Z. J. Public Health 36 (2012) 30–32.

[62] M. Madianou, Migration and the accentuated ambivalence of motherhood: the role of ICTs in Filipino transnational families, Global Netw. 12 (3) (2012) 277–295.

[63] Mahanti, A., Chakraborty, S.D., Goel, S., Veerelli, D., & Basu, S. Are you positive about Covid vaccination? A comparative study of Twitter Sentiment.

[64] Makice, K. (2009). *Twitter API: Up and running: Learn how to Build Applications With the Twitter API*. " O'Reilly Media, Inc.".
[65] E. Martínez-Cámara, M.T. Martín-Valdivia, L.A. Urena-López, A.R. Montejo-Ráez, Sentiment analysis in Twitter, Nat. Lang. Eng. 20 (1) (2014) 1–28.
[66] T.H. McCormick, L. Hedwig, N. Cesare, S. Ali, E.S. Spiro, Using Twitter for demographic and social science research: tools for data collection and processing, Sociol. Methods Res. (2015) 1–32 October.
[67] D. McGhee, C. Moreh, A. Vlachantoni, An 'undeliberate determinacy'? The changing migration strategies of Polish migrants in the UK in times of Brexit, J. Ethn. Migr. Stud. 43 (13) (2017) 2109–2130.
[68] A.A. Mir, R. Sevukan, Sentiment analysis of Indian Tweets about Covid-19 vaccines, J. Inf. Sci. (2022) 01655515221118049.
[69] A.A. Mir, S. Rathinam, S. Gul, Public perception of COVID-19 vaccines from the digital footprints left on Twitter: analyzing positive, neutral and negative sentiments of Twitterati, Trends Biotechnol. 40 (2) (2022) 340–356.
[70] M. Miyabe, A. Miura, E. Aramaki, Use trend analysis of twitter after the great east japan earthquake, in: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion, 2012, pp. 175–178.
[71] F. Morstatter, J. Pfeffer, H. Liu, K. Carley, Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose, in: Proceedings of the international AAAI conference on web and social media, 2013, pp. 400–408. Vol. 7, No. 1.
[72] H. Pan, T. Fokkema, L. Switsers, S. Dury, S. Hoens, L. De Donder, Older Chinese migrants in coronavirus pandemic: exploring risk and protective factors to increased loneliness, Eur. J. Ageing 18 (2) (2021) 207–215.
[73] U. Panchal, G. Salazar de Pablo, M. Franco, C. Moreno, M. Parellada, C. Arango, P. Fusar-Poli, The impact of COVID-19 lockdown on child and adolescent mental health: systematic review, Eur. Child Adolesc. Psychiatry (2021) 1–27.
[74] R.D. Parke, J.T. Cookston, Transnational fathers: new theoretical and conceptual challenges, J. Fam. Theory Rev. (2020).
[75] F. Rowe, M. Mahony, E. Graells-Garrido, M. Rango, N. Sievers, Using Twitter to track immigration sentiment during early stages of the COVID-19 pandemic, Data Policy 3 (2021) e36.
[76] F. Rowe, M. Mahony, E. Graells-Garrido, M. Rango, N. Sievers, Using Twitter to track immigration sentiment during early stages of the COVID-19 pandemic—ADDENDUM, Data Policy 4 (2022) e11.
[77] A. Saravanou, G. Valkanas, D. Gunopulos, G. Andrienko, Twitter floods when it rains: a case study of the UK floods in early 2014, in: Proceedings of the 24th International Conference on World Wide Web, 2015, pp. 1233–1238.
[78] M. Sayyadiharikandeh, O. Varol, K.C. Yang, A. Flammini, F. Menczer, Detection of novel social bots by ensembles of specialized classifiers, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 2725–2732.
[79] L. Shanahan, A. Steinhoff, L. Bechtiger, A.L. Murray, A. Nivette, U. Hepp, D. Ribeaud, M. Eisner, Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study, Psychol. Med. 52 (5) (2022) 824–833.
[80] C. Shofiya, S. Abidi, Sentiment analysis on COVID-19-related social distancing in Canada using Twitter data, Int. J. Environ. Res. Public Health 18 (2021) 5993 2021.
[81] M.M. Sloan, M. Haner, A. Graham, F.T. Cullen, J.T. Pickett, C.L. Jonson, Pandemic emotions: the extent, correlates, and mental health consequences of fear of COVID-19, Sociol. Spectrum 41 (5) (2021) 369–386.
[82] K. Starbird, Disinformation's spread: bots, trolls and all of us, Nature 571 (7766) (2019) 449–450.
[83] Statistics Poland (2021). Main directions of emigration and immigration in the years 1966-2018 (migration for permanent residence)
[84] M. Stephens, A geospatial infodemic: mapping Twitter conspiracy theories of COVID-19, Dialogues Hum. Geogr. 10 (2) (2020) 276–281.
[85] O. Troisi, G. Fenza, M. Grimaldi, F. Loia, Covid-19 sentiments in smart cities: the role of technology anxiety before and during the pandemic, Comput. Human Behav. 126 (2021) 106986, doi:10.1016/j.chb.2021.106986.
[86] Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*.
[87] A.A. Ullah, F. Nawaz, D. Chattoraj, Locked up under lockdown: the COVID-19 pandemic and the migrant population, Social Sci. Humanities Open 3 (1) (2021) 100126.
[88] P.M. Valkenburg, J. Peter, A.P. Schouten, Friend networking sites and their relationship to adolescents' well-being and social self-esteem, Cyberpsychol. Behav. 9 (2006) 584–590.
[89] M. Van den Bos, L. Nell, Territorial bounds to virtual space: transnational online and offline networks of Iranian and Turkish–Kurdish immigrants in the Netherlands, Global Netw. 6 (2) (2006) 201–220.
[90] R. Wilding, 'Virtual' intimacies? Families communicating across transnational contexts, Global Netw. 6 (2) (2006) 125–142.
[91] T. Wolbring, The digital revolution in the social sciences: five theses about big data and other recent methodological innovations from an analytical sociologist, Soziologie des Digitalen–Digitale Soziologie. Soziale Welt–Sonderband 23 (2020) 60–72.
[92] World Health Organization. (2020). ApartTogether survey: preliminary overview of refugees and migrants self-reported impact of Covid-19.
[93] J. Xue, J. Chen, R. Hu, C. Chen, C. Zheng, Y. Su, T. Zhu, Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach, J. Med. Internet Res. 22 (11) (2020) e20550.
[94] V. Yantseva, Migration discourse in Sweden: frames and sentiments in mainstream and social media, Social Media+ Society 6 (4) (2020) 2056305120981059.
[95] D.A.W. Yen, B. Cappellini, H.P. Yang, S. Gupta, Coping with coping: international migrants' experiences of the Covid-19 lockdown in the UK, British J. Manage. 32 (4) (2021) 1219–1241.
[96] S. Yousefinaghani, R. Dara, S. Mubareka, A. Papadopoulos, S. Sharif, An analysis of COVID-19 vaccine sentiments and opinions on Twitter, Int. J. Infect. Dis. 108 (2021) 256–262.
[97] W. Youyou, M. Kosinski, D. Stillwell, Computer-based personality judgments are more accurate than those made by humans, Proc. Natl. Acad. Sci. 112 (4) (2015) 1036–1040.
[98] E. Zagheni, I. Weber, Demographic research with non-representative internet data, Int. J. Manpow. (2015).
[99] E. Zagheni, V.R.K. Garimella, I. Weber, B. State, Inferring international and internal migration patterns from twitter data, in: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 439–444.
[100] P. Zola, P. Cortez, M. Carpita, Twitter user geolocation using web country noun searches, Decis. Support Syst. 120 (2019) 50–59.
[101] H. De Haas, S. Castles, M.J. Miller, The age of migration: International population movements in the modern world, Bloomsbury Publishing, 2019.
[102] DTM (2020) Global mobility restriction overview 9th April 2020: https://reliefweb.int/sites/reliefweb.int/files/resources/DTM-Covid19%20Global%20Overview%20Output%2009.04.2020.pdf
[103] E. Bonnevie, S.D. Rosenberg, C. Kummeth, J. Goldbarg, E. Wartella, J. Smyser, Using social media influencers to increase knowledge and positive attitudes toward the flu vaccine, Plos one 15 (10) (2020) e0240828.
[104] Y. Freire-Vidal, E. Graells-Garrido, F. Rowe, A framework to understand attitudes towards immigration through Twitter, Applied Sciences 11 (20) (2021) 9689.
[105] Chlasta, K. (2019). Sentiment analysis model for Twitter data in Polish language. arXiv preprint arXiv:1911.00985.
[106] J. Kocoń, P. Miłkowski, K. Kanclerz, MultiEmo: Multilingual, Multilevel, Multidomain Sentiment Analysis Corpus of Consumer Reviews, in: M. Paszynski, D. Kranzlmüller, V.V. Krzhizhanovskaya, J.J. Dongarra, P.M.A. Sloot (Eds.), Computational Science – ICCS 2021. ICCS 2021. Lecture Notes in Computer Science, Springer, Cham, 2021 vol 12743.
[107] J. Kocoń, P. Miłkowski, M. Zaśko-Zielińska, Multi-level sentiment analysis of PolEmo 2.0: Extended corpus of multi-domain consumer reviews, in: Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), 2019, pp. 980–991.
[108] A.M. Jamison, D.A. Broniatowski, M. Dredze, A. Sangraula, M.C. Smith, S.C. Quinn, Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic'on Twitter, Harvard Kennedy School Misinformation Review (2020) 1.
[109] P. Miłkowski, M. Gruza, P. Kazienko, J. Szołomicka, S. Woźniak, J. Kocoń, Multi–model Analysis of Language-Agnostic Sentiment Classification on MultiEmo Data, in: Computational Collective Intelligence: 14th International Conference, ICCCI 2022, Hammamet, Tunisia, September 28–30, 2022, Proceedings, Cham, Springer International Publishing, 2022, pp. 163–175.