# Framing COVID-19:
# How we conceptualize and discuss the pandemic on Twitter

Philipp Wicke[1*], Marianna M. Bolognesi[2]

[1]Department of Computer Science, University College Dublin, Dublin, Ireland

[2]Department of Department of Modern Languages, Literatures, and Cultures, University Bologna, Bologna, Italy

*Corresponding author

E-mail: philipp.wicke@ucdconnect.ie

# Abstract

Doctors and nurses in these weeks are busy in the trenches, fighting against a new invisible enemy: Covid-19. Cities are locked down and civilians are besieged in their own homes, to prevent the spreading of the virus. War-related terminology is commonly used to frame the discourse around epidemics and diseases. Arguably the discourse around the current epidemic will make use of war-related metaphors too,not only in public discourse and the media, but also in the tweets written by non-experts of mass communication. We hereby present an analysis of the discourse around #Covid-19, based on a corpus of 200k tweets posted on Twitter during March and April 2020. Using topic modelling we first analyze the topics around which the discourse can be classified. Then, we show that the WAR framing is used to talk about specific topics, such as the virus treatment, but not others, such as the effects of social distancing on the population. We then measure and compare the popularity of the WAR frame to three alternative figurative frames (MONSTER, STORM and TSUNAMI) and a literal frame used as control (FAMILY). The results show that while the FAMILY literal frame covers a wider portion of the corpus, among the figurative framings WAR is the most frequently used, and thus arguably the most conventional one. However, we conclude, this frame is not apt to elaborate the discourse around many aspects involved in the current situation. Therefore, we conclude, in line with previous suggestions, a plethora of framing options, or a metaphor menu, may facilitate the communication of various aspects involved in the Covid-19-related discourse on the social media, and thus support civilians in the expression of their feelings, opinions and ideas during the current pandemic.

# Introduction

On December 31, 2019, Chinese authorities alerted the World Health Organization of pneumonia cases in Wuhan City, within the Hubei province in China. The cause, they initially said, was unknown, and the disease was first referred to as 2019-nCoV and then named COVID-19. The next day, the Huanan seafood market was closed, because it was suspected to be the source of the unknown disease, as some of the patients presenting with the pneumonia-like illness were dealers or vendors at that market. Since then, the disease has spread quickly throughout China, and from there to the rest of the world. SARS-CoV-2 is the name of the virus responsible for this coronavirus pandemic that we are experiencing while the present article was in writing. The virus has so far spread throughout all the inhabited continents and affected millions people, killing thousands of individuals. Schools have been shut down, kids are at home, many workers are now working from remote, locked down in their houses, and leaving only for reasons of primary necessity, such as shopping for groceries and going to medical appointments.

With many countries implementing lock downs and promoting quarantines, suggesting or forcing citizens to stay inside their homes in order to avoid spreading the virus, millions of people are experiencing a global pandemic for the first time in their lives. The social distancing enforced by various governments stimulated many internet users to use social media to communicate and express their own concerns, opinions, ideas and feelings in relation to this new situation. On Twitter, for example, around 16K Tweets are posted by Twitter users every hour, containing a hashtag such as #coronavirus, #Covid-19 or #COVID. A variety of issues are debated on a daily basis on Twitter, in relation to the pandemic. These include, but are not limited to, the political and social consequences of various governmental decisions, the situations in the hospitals getting increasingly more crowded every day, the interpretation of the numbers associated with the spreading of the pandemic, the problems that families face with homeschooling their children while working from home, and so forth. Among these issues, the discussion around the treatment and containment of the virus is surely a central topic.

The present article aims at describing how the discourse around Covid-19 is framed on Twitter. In particular, we want to understand what are on Twitter the main topics related to the discourse around Covid-19 and to what extent the treatment of the disease is framed figuratively. Because previous research has shown that various social and political issues addressed in public discourse are framed in terms of wars [1], we assume that this tendency may emerge also on Twitter, in relation to the discourse around Covid-19. Unlike the articles on magazines and journals typically used for corpus analyses of this kind, Twitter does contain messages written by journalists and other experts in mass communication, as most tweets are provided by non-expert communicators. We hereby want to investigate to what extent Twitter users, and therefore non-expert communicators, frame Covid-19 in terms of a war, and whether other figurative framings arise from automated analyses of our corpus of Tweets containing virus-related hashtags.

In particular, we address the following research questions:

1. What type of topics are discussed on Twitter, in relation to Covid-19?

2. To what extent is the WAR figurative frame and the conventional metaphor DISEASE TREATMENT IS WAR used to talk about Covid-19 on Twitter? Which lexical units are used within this metaphorical frame and which lexical units are not?
3. Are there alternative figurative frames used to talk about Covid-19 on Twitter? And how does their use compare to the use of the WAR frame?

Each of the three questions is addressed in a specific study, which we describe in the Methods Section 3.

# Theoretical background

Mining the information encoded by private internet users in the short texts posted on Twitter (the tweets) is becoming an increasingly fruitful field of research. In relation to health discourse, tweets have been used by epidemiologists to access supplementary data about epidemics. For example, tweets about particular diseases have been compared to gold-standard incidence data, showing that there are positive correlations between the number of tweets discussing flu-symptoms and official statistics about the virus spread such as those published by Centers for Disease Control and Prevention and the Health Protection Agency [2]. Already a decade ago, in Brazil, tweets have been used to track the spreading of the dengue fever, a mosquito-transmitted virus [3]. More recently, Pruss and colleagues [4] used a topic model applied to a large corpus of tweets to automatically identify and extract the key topics of discussion about the Zika disease, a virus that spread mainly in the Americas in early 2015. The authors found also that rises in tweeting activity tended to follow major events related to the disease, and Zika-related discussions were moderately correlated with the virus incidence. Moreover, it has been demonstrated that the combination of data collected from hospitals about specific diseases, and data collected from social media, can improve surveillance and forecasting about the disease more effectively [5, 6].

Besides providing a valuable tool for tracking the spread of epidemics, and thus helping experts to take more effective decisions, social media have been used to investigate public awareness, attitudes and reactions about specific diseases [7, 8]. As Pruss and colleagues report in their review [4], the 2013 measles outbreak in the Netherlands, for example, has been analysed in this perspective by Mollema and colleagues [9], who compared the number of tweets (and other messages posted on social media) with the number of online news articles as well as with the number of reported measles cases and found a strong correlation between social media messages and news articles and a mild correlation between number of tweets and number of reported measles cases. Moreover, through a topic analysis and a sentiment analysis of the tweets, they found that the most common opinion expressed in the tweets was frustration regarding people who do not vaccinate because of religious reasons (the measles outbreak in the Netherlands began among Orthodox Protestants who often refuse vaccination for religious reasons).

The 2014 Ebola outbreak in Africa was also used as a case study to mine the attitudes, concerns and opinions of the public, expressed on Twitter. For example, Lazard and colleagues [10] analysed user-generated tweets to understand what were the main topics that concerned the American public, when the panic widespread on US soil after one case of Ebola was detected. The authors found that the main topics of concern for the American public were the symptoms and lifespan of the virus, the disease transfer and contraction, whether it was safe to travel, and how they could protect their body from the disease. In relation to the same outbreak, Tran and Lee [11] built Ebola-related information propagation models to mine the Ebola related tweets and the information encoded therein, focusing on the distribution over six topics, broadly defined as: 1. Ebola cases in the US, 2. Ebola outbreak in the world, 3. fear and pray, 4. Ebola spread and warning, 5. jokes, swear and disapproval of jokes and 6. impact of Ebola to daily life. The authors found that the second topic had the lowest focus, while the fifth and sixth had the highest.

More recently, tweets have been mined to understand the discussion around the Zika epidemics. Miller and colleagues [12] used a combination of natural language processing and machine learning techniques to determine the distribution of topics in relation to four characteristics of Zika: symptoms, transmission, prevention, and treatment. The authors managed to outline the most persistent concerns or misconceptions regarding the Zika virus, and provided a complex map of topics emerged from the tweets posted within each of the four categories. For example, in relation to the issue of prevention they observed the emergence of the following topics: need for control and prevent spread, need for money, ways to prevent spread, bill to get funds, and research. Vijaykumar and colleagues [13] analysed how content related to Zika disease spreads on Twitter, thanks to tweets amplifiers and retweets. The authors found that, of the 12 themes taken into account, Zika transmission was the most frequently talked about, on Twitter. Finally, Pruss and colleagues [4] mined a corpus of tweets in three different languages (Spanish, Portuguese and English) with a multilingual topic model and identified key topics of discussion across the languages. The authors reported that the Zika outbreak was discussed differently around the world, and the topics identified were distributed in different ways across the three languages.

In cognitive linguistics, and in particular in metaphor studies, it has been shown that the discourse around diseases and epidemics is often framed by means of figurative constructions. In particular, metaphor is often used to talk about different aspects of diseases, such as their treatment, their outbreak and their symptoms. The framing power of metaphor is particularly relevant in health-related discourse, because it has been shown that it can impact patients' general well-being. For example, in a seminal study Sontag [14] criticized the popular use of war metaphors to talk about cancer, a topic of research recently investigated also by Semino and colleagues [15]. As these authors explain, the military metaphor that we tend to use to talk about the development, spreading and cure of cancer inside the human body has been repeatedly rejected by cancer patients as well as by many relatives and doctors, who indicate that such framing provokes anxiety and a sense of helplessness that can have negative implications for cancer patients. In a series of experiments, for example, Hendricks and colleagues [16] found that framing a person's cancer situation within the war metaphor, and therefore as a battle, has the consequence of making people believe that the patient may feel guilty in the case that the treatment does not succeed. Conversely, framing the cancer situation as a journey encourages the inference that the patient will experience less anxiety about her health condition.

The military metaphor thanks to which we frame diseases such as cancer is a very common one to be found in public discourse [1]. According to Karlberg and Buell [17] 17% of all articles in the Time Magazine published between 1981 and 2000, contained at least one war metaphor. The war metaphor is not used solely to frame the discourse around diseases, but also the discussion around political campaigns, crime, drugs and poverty. As explained in [1], war metaphors are pervasive in public discourse and span a wide range of topics because they provide a very effective structural framework for communicating and thinking about abstract and complex topics, notably because of the emotional valence that these metaphors can convey. In the special case of the diseases, the war metaphor is typically used to frame the situation relatively to the treatment of the disease. As indicated in MetaNet, the Berkeley-based structured repository of conceptual metaphors and frames [18], the metaphor can be formalized as DISEASE TREATMENT IS WAR, or TREATING DISEASE IS WAGING WAR (https://metaphor.icsi.berkeley.edu/pub/en/index.php/Metaphor:DISEASE_TREATMENT_IS_WAR). Within this metaphor, a variety of mappings can be identified, including: the diseased cells are enemy combatants, medical professionals are the army of allies, the body is the battlefield, medical tools are weapons, and applying a treatment is fighting.

The figurative frame of WAR, used in discourses around diseases, is certainly a conventional one, frequently used, often unconsciously. As argued by [1], such a frame is handy and frequently used because it draws on basic knowledge that everyone has, even though for most people this is not knowledge coming from first hand experience. Moreover, this frame expressed in an exemplary way the urgency associated with a very negative situation, and the necessity for actions to be taken, in order to achieve a final outcome quickly. The outcome can be either positive or negative, in a rather categorical way. The inner structure of the frame is also relatively simple, with opposing forces clearly labelled as in-groups and out-groups, or allied and enemies. Each force has a strategy to achieve a goal, which involves risks and can potentially be lethal. For these reasons, this frame is arguably very well suited to appear in the discourse around

Covid-19, as previously observed in relation to other diseases. The adversarial relationship between doctors and the virus, the different goals afforded by the two forces and the human body as the battlefield for this operation, are possible mappings that we seek to trace down, with our analysis on Covid-19 related tweets.

Despite the undebatable frequency by which public discourse around diseases uses war metaphors, this frame is often not well received, as mentioned above, and war-related metaphors can be resisted. As we will further elaborate in the Discussion section, in some cases the press opposes deliberately the war frame, advancing alternative figurative frames. Tracking down alternative frames to the war one in a qualitative manner has been a recent endeavor initiated by scholars in cognitive linguistics and corpus analysis, on Twitter. The hashtag #ReframeCovid (first proposed, to the best of our knowledge, by Inés Olza and Paula Sobrino) has been recently used to harvest texts such as articles, advertisements and notes showing how the virus has been opposed and framed in alternative ways by a few journalists and writers. Notably, the discourse has been reframed using lexical units related to the domain of FOOTBALL, of GAMES, of STORMS and so forth. In Study 3, hereby reported, we explored the structure and functioning of alternative frames too, in our corpus-based analysis of tweets about Covid-19, and compared them to the WAR frame as well as a literal frame, that is the frame of family and friends.

# Methods

## General design of the 3 studies

In Study 1 we explored the range of topics addressed in the discourse on Covid-19 on Twitter. To do so, we used a topic modelling technique. In Study 2 we explored the actual usage of the WAR frame, and in which topics related to Covid-19 is the WAR frame more frequently used. To do so, we compiled a list of war-related lexical units and ran it against our corpus of Covid-19 tweets, observing and discussing which lexical units of each frame were used in the tweets. In Study 3 we explored alternative frames that could be used to frame the discourse around Covid-19 on Twitter. To do so, we compiled lists of lexical units for selected alternative frames (three figurative frames and one literal one) and compared the percentages by which they appear to be used in the corpus of tweets, against the percentages by which the WAR frame is used.

## Constructing the corpus of Covid-19 tweets

In order to identify tweets that relate to the Covid-19 epidemic, we defined a set of relevant hashtags used to talk about the virus: #covid19, #coronavirus, #ncov2019, #2019ncov, #nCoV, #nCoV2019, #2019nCoV, #COVID19. Using Twitter's official API in combination with the *Tweepy* python library (tweepy.org) for 12 days we have collected 25.000 tweets per day that contain at least one of the hashtags. We did not collect retweets or mentions without the hashtag. The tweets were collected in accordance with the Twitter terms of service. Two main restrictions of those terms and service motivated our decision to limit the extent of our corpus: Firstly, the free streaming API only allows access up to one week of Twitter’s history. Secondly, there is a limit of 180 requests per 15-minutes. The algorithm to obtain the tweets will automatically increase our corpus in the upcoming weeks, such that new data will be freely available for future research. As previously mentioned, we do not store users' ID and release the data in compliance with ethical standards.

To balance our corpus, we needed to consider how a single tweet or a single user weights on the overall corpus. For example, a scientific analysis of fake news spread during the 2016 US presidential election showed that about 1% of users accounted for 80% of fake news and report that other research suggests that 80% of all tweets can be linked to the top 10% of most tweeting users [19]. Therefore, we decided to filter the corpus and retain only one tweet per user. Table 1 provides an overview of the dates, collected and filtered tweets.

**Table 1**: **Dates of collection for tweets containing hashtags related to the Covid-19 epidemic.**

| **Date** | 20.03.2020 | 21.03.2020 | 22.03.2020 | 23.03.2020 | 24.03.2020 | 25.03.2020 | 26.03.2020 |
|---|---|---|---|---|---|---|---|
| **Filtered / Collected** | 20316 /25000 | 39284 /50000 | 57073 /75000 | 73346 /100000 | 89785 /125000 | 103614 /150000 | 118866 /175000 |
| | | | | | | | |
| **Date** | 27.03.2020 | 28.03.2020 | 29.03.2020 | 30.03.2020 | 31.03.2020 | 01.04.2020 | 02.04.2020 |
| **Filtered / Collected** | 132995 /200000 | 146654 /225000 | 156775 /250000 | 167847 /275000 | 180234 /300000 | 191278 /325000 | 203756 /350000 |

The filtered tweets are single tweets per user, the total number of collected tweets is 25k per day.

As the streaming API starts collecting tweets from 23:59 CET of each day and has been limited to English, our corpus arguably encompasses mainly tweets produced by users residing in the

USA, where the time of data collection corresponds to awake hours, and the targeted language corresponds to the first language of many (if not most) US residents. The total number of collected tweets from individual users over 14 days is 203,756. This results in 41,78% of the collected tweets being tweets of a user tweeting more than once, thus being filtered out. We can analyse the most common words in our corpus and use topic modeling to understand what topics are connected when people tweet about the epidemic.

The corpus therefore contains over 300.000 tweets collected between 20.03.2020 and 01.04.2020 mentioning the Covid-19 epidemic. Each tweet is stored with the public username, a timestamp and location. As we are only interested in a linguistic analysis and considering the privacy rights of Twitter users for research [20], we have only stored the tweet along with a timestamp. This corpus is publicly available in the online repository on OSF, retrievable at the following link: https://osf.io/bj5a6/?view_only=1644595a66dd4adebeeb6b2bb0449c89.

## Study 1: Identifying topics in Covid-19 discourse on Twitter through Topic Modeling

In order to extract and identify topics from the corpus, we used a Latent Dirichlet Allocation algorithm (henceforth, LDA) [21]. LDA is an unsupervised machine learning algorithm that aims to describe samples of data in terms of heterogeneous categories. It is mostly used to identify categories in documents of text and thus appropriate to identify topics within the Covid-19 corpus of tweets. The study reported by Pruss and colleagues on the corpus of tweets related to the Zika epidemics [4], for example, used the same algorithm to identify topics within the corpus. For the purpose of our study we used the *Gensim* LDA-Multicore algorithm, which allows us to parallelize the training of our data on multiple CPUs. As an unsupervised learner, LDA needs to be given the number of topics that it will try to divide the data into. Our exploratory approach includes the search space for several different amounts of topics, thus varying in the level of granularity represented within each topic. We hereby report the results obtained from the division of the data into a relatively small number of topics (N=4) and a relatively large number of topics (N= 16), to show and compare a less granular and a more granular division of the data. We expect to find broader and more generic concepts listed in the first analysis (N=4) and more specific concepts in the fine-grained topic analysis (N=16). We have used 6 passes of the algorithm to go over the entire corpus (higher numbers cost more training time, but can improve the topic quality). To train the LDA model, we preprocessed our corpus with the following operations:

- converting each tweet into a list of tokens (using *Gensim*'s simple_preprocess function)
- removing tokens with less than 3 characters (e.g. “aa”, “fo”, “#o”)
- removing stopwords from the list of tokens (stopwords from Stone et al. [22])
- removing Covid-19 words from the list of tokens (e.g. “covid”, “nCov”, “coronavirus” etc)
- turning the tokens into a bag-of-words, i.e. a list of tuples with the token and its number of occurrences in the corpus

We excluded terms like “coronavirus”, “covid”, “corona”, “virus” or “nCov19” from the topic modeling, because these do not add information about the topics themselves. The preprocessing resulted in a list of 415,329 tokens, that is, lexemes. We did not lemmatize the corpus, nor pos-tagged it for the purpose of our study. Therefore, for example, gerundive forms of verbs, as well as plural forms of nouns are present in the corpus, and the list of frame-related words is also composed by lexemes and not simple lemmas. Additionally, we trained another LDA model with a tf-idf (term frequency-inverse document frequency) version of the tokens. The tf-idf assigns a statistical relevance to each token based on how many times the token occurs and the inverse document frequency (a measure of whether the token is rare or common in the corpus) of that token. As its results did not add any further insight to our research, we provide it in the online repository but do not discuss this model further.

## Study 2: Determining lexical units associated with the WAR frame

To investigate to what extent users use the WAR frame, to talk about Covid-19, we needed to assess the amount of tweets that use war language in our corpus. To explore the lexical units associated with the WAR frame we took a double approach, using two tools. The first tool was the web-service *relatedwords.org* This web-service provides a list of words (lexemes, not lemmas) related to a target word. This list is ranked through competing scores by several algorithms. One of which finds similar words in a word embedding [23] another one queries *ConceptNet* [24] to find words with meaningful relationships to the target word. Choi and Lee [25] used the same web-service to expand the list of categories used to model conceptual representations for crisis-related tweets [25]. The list of words retrieved on *relatedwords* was adapted to our purpose. As a matter of fact, the list featured words such as “franco-prussian war” or “aggression”. The former is a specific type of war and it includes the term “war” itself, we dropped any kind of specific war or terms that include a compound of war, e.g. “state of war”. The latter term “aggression” is too broad if yet closely related to the target word to be kept in the list. Additionally, in case of doubt, we checked the term in an online dictionary to verify its relation to the war framing. The term list includes the following 79 terms. The second tool used to prepare the list of lexical units related to the WAR frame was the MetaNet repository of conceptual metaphors and frames housed at the International Computer Science Institute in Berkeley, California [18]. Here, from the WAR frame (https://metaphor.icsi.berkeley.edu/pub/en/index.php/Frame:War) we selected the 12 words that were not yet included in the selection of lexical units based on *relatedwords.* Moreover, we dropped compound units that included words that we had already included in the list (e.g., “combat zone”, because we featured already the word “combat”) and two mis-spelled units (“seige”instead of “siege” and “beseige” instead of “besiege”). The total number of lexical units for the WAR framing was 91:

> **WAR (91)**: allied, allies, armed, armies, army, attack, attacks, battle, battlefield, battleground, battles,belligerent, bloodshed, bomb, captured, casualties, combat, combatant, combative, conflict, conquer, conquering, conquest, crusade, defeat, defend, defenses, destruction, disarmament, enemies, enemy, escalation, fight, fighter, fighting, foe, fortify, fought, grenade, guerrilla, gunfight, holocaust, homeland, hostilities,

hostility, insurgency, invaded, invader, invaders, invasion, liberation, military, peace, peacetime, raider, rebellion, resist, resistance, riot, siege, soldier, soldiers, struggle, tank, threat, treaty, trench, trenches, troops, uprising, victory, violence, war, warfare, warrior, wars, wartime, warzone, weapon, alliance, ally, arsenal, blitzkrieg, bombard, front, line, minefield, troop, vanquish, vanquishment.

In order to understand where in relation to our predicted LDA topics the WAR frame is located, we have collected all tweets that mention at least one term of the WAR frame and asked the LDA model to predict its topic. This way, we could identify the topics with the most or the least terms related to WAR.

## Study 3: Search method for alternative framings and relevant lexical units therein

In order to identify whether or not the war framing is particularly relevant, we explored alternative framings used in discourses on viruses. For this purpose we used the metaphor exploration web services by [26], called *MetaphorMagnet* http://bonnat.ucd.ie/metaphor-magnet-acl). Using the keywords “virus” and “epidemic” we selected the following alternative frames, which could be in principle used to frame the discourse around Covid-19: STORM, MONSTER and TSUNAMI. We did not include alternative frames that have been observed in the press and harvested with the #ReframeCovid keyword, because these featured words that arguably are used also in a very literal way on Twitter, in relation to

Covid-19. Notably these frames were GAME and the sub-frame SOCCER GAME, used in the Spanish press according to the community of Spanish cognitive linguists. However, lexical units such as “game”, “football”, “soccer”, “game season”, and so on, are likely to be used literally in the tweets, to refer to the fact that all sport events and thus all games have been suspended, due to the epidemic. Another frame that has been observed in the press and tagged as #ReframeCovid is the FLOOD frame. However, through a quick search on Metaphor Magnet, we realized that this frame has too many shared lexical units with STORM and TSUNAMI, and was therefore discarded.

In order to select the lexical units within each of the alternative frames, we used the tool *relatedwords*, already used for the WAR frame, for consistency. However, because these alternative frames are arguably less conventionalized, none of them is included in the list of frames on MetaNet. Thus, *relatedwords* was the only tool we used to harvest lexical units for the alternative frames. We created three list of lexical units:

> **STORM (57)**: thunderstorm, rain, lightning, snowstorm, blizzard, wind, hurricane, weather, rainstorm, typhoon, tempest, precipitation, beaufort, snow, cyclone, meteorology, hail, hailstorm, windstorm, flooding, thunder, tornado, monsoon, rainfall, rage, force, disaster, ice, storm, atmospheric, disturbance, wildfire, clouds, firestorm, ramp, tornadoes, fog, winds, rains, waves, landfall, thunderhead, duststorm, tides, gusts, floodwaters, wave, cloud, swells, cloudburst, anticyclone, downpour, sandstorm, stormy, whirlwinds, storms, oceanographic.

**MONSTER (51)**: freak, demon, devil, giant, ogre, fiend, zombie, frankenstein, bogeyman, werewolf, horror, mutant, creature, dragon, superhero, goliath, behemoth, monstrosity, colossus, legend, evil, lusus, naturae, mouse, beast, boogeyman, leviathan, dracula, monstrous, teratology, villain, killer, ghost, gigantic, siren, superman, vampire, undead, psycho, monster, chimera, godzilla, fiction, mythology, mutation, demoniac, manatee, mermaid, monsters, spider, bug.

**TSUNAMI (50)**: earthquake, disaster, tide, oceans, calamity, catastrophe, tragedy, wavelength, wind, period, cataclysm, flood, eruption, tidal, seiche, quake, thucydides, floods, floodwater, cyclone, devastation, ocean, surface, wave, coastlines, typhoon, waves, hurricane, magnitude, aftershock, mudslide, seafloor, richter, seawall, seismic, landslide, tsunamis, aftershocks, flooding, torrential, earthquakes, deepwater, triggering, tsunami, tremors, mudslides, riptide, rains, whirlpool, pacific.

As for the WAR frame, we ran these lists against our corpus and compared the frequency of occurrences within the corpus across the different framings.

## The literal frame of FAMILY used as control

The frames outlined above (WAR; TSUNAMI; STORM; MONSTER) are not expected to be found in a literal, but figurative sense in tweets relating to the Covid-19 pandemic. To evaluate their relevance across Covid-19 related tweets, we compared the occurrence of the lexical units listed therein with those listed within a frame that we can expect to occur in the literal sense. We have chosen the concept “family” and have compiled a comparable list using the same strategy as with the other term lists:

> **FAMILY (66)**: marriage, household, kin, house, kinfolk, home, lineage, kinship, parent, relative, clan, cousin, children, child, sister, mother, father, uncle, nephew, brother, grandson, son, grandfather, grandmother, kinsfolk, ancestor, consanguinity, tribe, sibling, subfamily, kindred, stepfamily, couple, family, sib, foster, parentage, menage, phratry, folk, daughter, kinsperson, aunt, grandma, granddaughter, grandaunt, stepbrother, niece, stepson, dad, stepdaughter, stepfather, wife, husband, daddy, parents, elder, daughters, mom, siblings, stepmother, grandpa, grandparents, relatives, widow, spouse

# Results and Discussions

## General Corpus Analytics

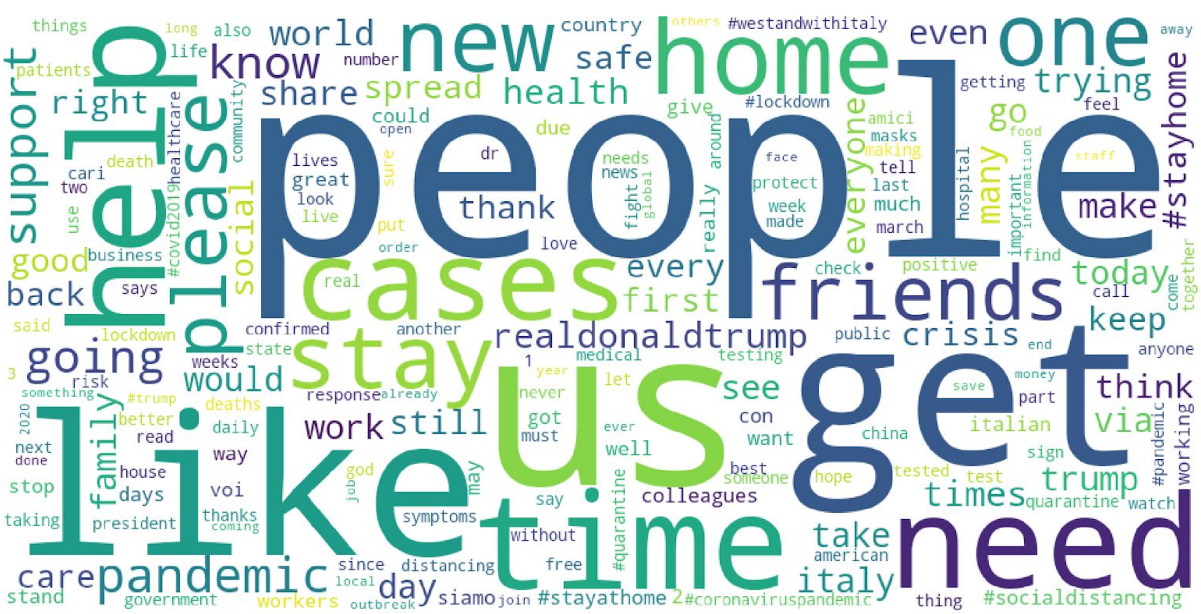
The corpus comprised 203,756 tweets, in which the 30 most common words, excluding stopwords and online tags (e.g. “&amp”, “https”) are:

> people (19153), us (13368), get (11270), like (10451), time (10263), help (10091), need (9993), cases (9205), home (9044), stay (8788), new (8752), one (8725), friends (8465), please (8232), pandemic (7614), support (7255), know (6931), going (6788),

realdonaldtrump (6659), times (6462), world (6451), health (6449), day (6153), family (6010), go (5986), trump (5967), work (5862), would (5705), today (5602), take (5532)

In this list, the number in brackets represents the frequency of occurrence, also visualized in Fig 1 (in which greater size of the word indicates greater occurrence in the corpus). In this word cloud, the larger the word print, the more frequent its occurrence in the corpus. The most frequent word is“people” with about 19k occurrences, followed by “us” with about 13k occurrences. Notably, we cannot distinguish whether “us” means “United States” or the pronoun “us”.

**Fig 1: Word cloud of the most common words in the corpus of over 200k collected tweets with at least one hashtag relating to the covid19 epidemic.**



# Study 1: Topic Model Analysis

## Analysis of 4 Topics

Dividing the corpus into four topics through the LDA, we obtain a list of words for each topic and the weightage (importance). Fig 2 shows the word clouds with greater words signalling greater significance. Except for topic #II, all of the other topics include the word “pandemic” among their most important words and show a strong overlap. The weights (importance) and words for each topic allocated by the LDA model with N=4 topics are the following:

LDA (N=4, no tf-idf, 6 passes):

**Topic #I**: 0.008 "pandemic", 0.005 "news", 0.004 "data", 0.004 "update", 0.004 "world", 0.004 "youtube", 0.003 "information", 0.003 "latest", 0.003 "today", 0.003 "april"

**Topic #II**: 0.029 "times", 0.028 "friends", 0.027 "family", 0.022 "share", 0.021 "italy", 0.021 "trying", 0.014 "support", 0.014 "sign", 0.013 "stand", 0.011 "colleagues"

**Topic #III**: 0.014 "people", 0.013 "cases", 0.011 "trump", 0.009 "realdonaldtrump", 0.008 "like", 0.006 "china", 0.006 "world", 0.005 "deaths", 0.005 "pandemic", 0.004 "going"

**Topic #IV**: 0.010 "home", 0.008 "help", 0.008 "stay", 0.007 "people", 0.007 "time", 0.007 "need", 0.007 "pandemic", 0.006 "health", 0.006 "work", 0.005 "safe"

**Fig 2: Word clouds form N=4 LDA topic modeling with greater words signalling greater significance.**


## Analysis of 16 Topics

The results for 16 topics show a much greater diversity among the classes and we present the 16 topics in word clouds in Fig 3. The results for the LDA with tf-idf are also included in the supplementary files.

**Fig 3: Depiction of the word clouds for each of the 16 topics clustered by the LDA.**

## Study 1: Topic Modelling Discussion

The LDA algorithm does not provide labels for the topics, therefore we need to identify labels for the topics. This is an interpretative task and aims to enrich the results with respect to our research questions. The topics identified by LDA analysed above and visualized in Fig 2 can be labelled as follows:

**Topic #I**: *Communications and Reporting*.

**Topic #II**: *Community and Social Compassion*.

**Topic #III**: *Politics*.

**Topic #IV**: *Reacting to the epidemic*.

The sixteen topic LDA model provided a more fine-grained view of topics that could be related to the 4 general topics. In the field of *Communication and Reporting,* we observed finer distinctions in topics #4, #11 and some in #15. Where #11 is more focused on "World", "Trump" and "China", we also find a topic in #4 closer to "News", "Lockdown", "Press", "Media" etc. In the domain of *Community and Social Compassion*, topic #3 is very close to topic #II. Whereas, topic #13, #16, #5 relate to topics around the quarantine, self-isolation and in general *Reacting to the Epidemic* (#IV).

There are also some novel topics around treatment and medical needs (#1, #6, #7), around testing (#10) or working/studying from home (#2, #9 and parts of #12). Rather unrelated to the whole epidemic, a conglomerate of words can be found in topic #8 and parts in #12.

## Study 2: WAR Framing Results

Analysing all tweets from the database, a total of 9,502 tweets contained at least one term from the WAR framing, which is 4.66% of all tweets. Of these, 1,029 tweets had more than one war-related term. The 20 most common war terms found in our database are hereby reported with relative percentage to all war terms and number of overall occurrence:

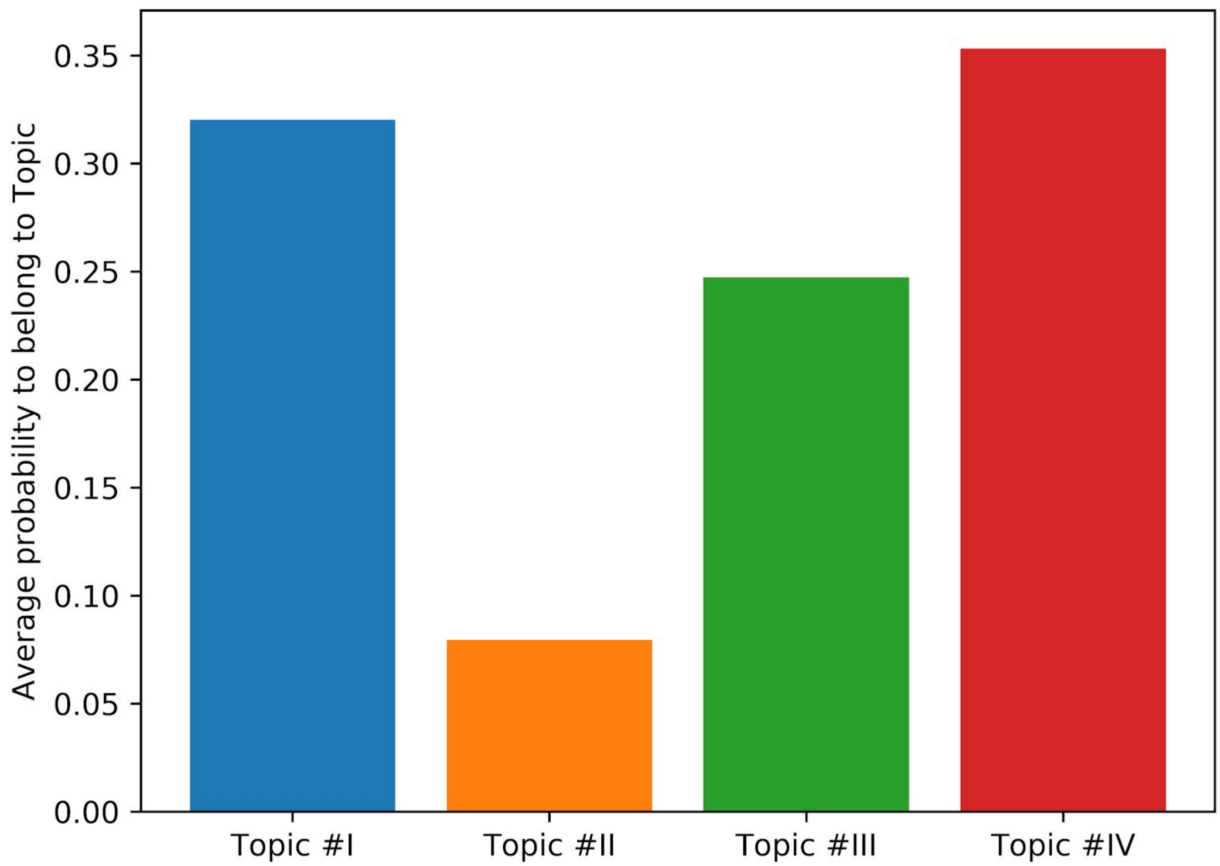
> **WAR**: fight (28.80%, 3033), fighting (10.19%, 1073), war (8.31%, 875), combat (5.84%, 615), threat (5.01%, 528), battle (3.96%, 417), front line (3.34%, 383), military (2.87%, 302), attack (2.82%, 297), peace (2.50%, 263), enemy (2.47%, 260), defeat (2.46%, 259), violence (1.91%, 201), attacks (1.39%, 146), struggle (1.30%, 137), weapon (1.11%, 117), soldiers (0.91%, 96), victory (0.79%, 83), troops (0.76%, 80), wars (0.74%, 78)

Words that were virtually absent (or had very limited usage) in the context of Covid-19 on Twitter were: armies (2x), battleground (2x), combatant (2x), combative (2x), conquering (2x), grenade (2x), gunfight (2x), hostility (2x), invader (2x), bombard (2x), belligerent (1x), disarmament (1x), guerrilla (1x), insurgency (1x), raider (1x), rebellion (1x), treaty (1x), minefield (1x), vanquish (1x), blitzkrieg (0x), vanquishment (0x), conquest (0x).

## LDA Topic prediction of WAR tweets

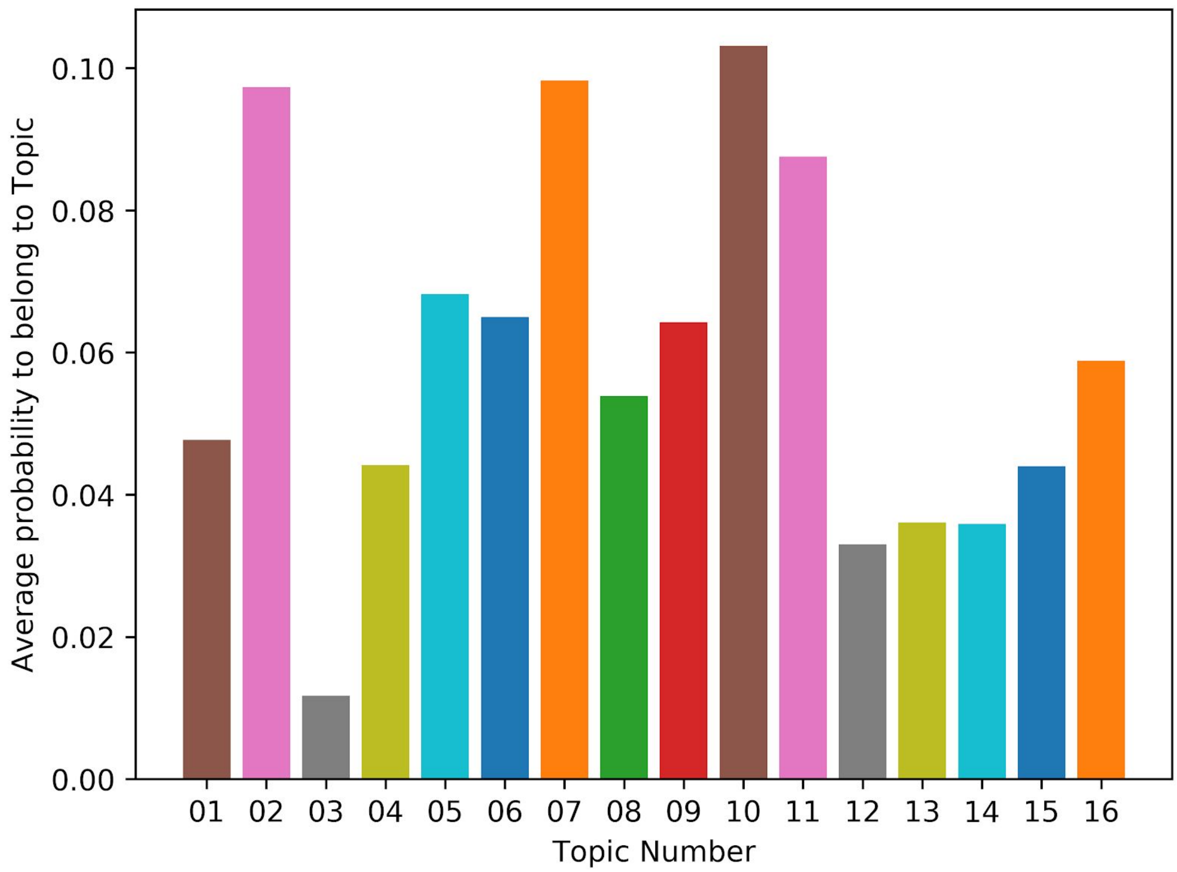
The LDA model can predict the probability of a document to belong to a certain topic of the corpus. We therefore used this prediction method to investigate what topics are relevant for those tweets that feature WAR terms. For this, we tokenize all tweets that contain at least one WAR term and use both of our LDA models to suggest which of the four and sixteen topics those WAR related tweets most likely belong to. For the four topic model we can see a distribution in Fig 4.

**Fig 4: LDA-predicted average probability of WAR term contributing to one of 4 topics.**




For the sixteen topic model, we can see a distribution in Fig 5. This image shows that lexical units belonging to the WAR domain and therefore tweets that relate to the WAR frame are most likely to be found in tweets that belong to topics IV and I, and partly III (in the macro distinction of topics) and in tweets that belong to topics 2, 7 and 10 in the fine-grained distinction.

**Fig 5: LDA-predicted average probability of a WAR term contributing to one of 16 topics.**




## Study 2: WAR Framing Discussion

The results show that 4.66% of all tweets contain war-related terms and are therefore likely to frame the discourse around Covid-19 metaphorically, in terms of a literal war. While it is hard to evaluate in absolute terms the impact that this frame has on the overall discourse around Covid-19, we will see in Study 3 how the WAR frame compares to the usage of 3 other figurative frames, as well as to a literal frame, and we will resume the discussion of this aspect in the General Discussion section.

The specific words within the WAR frame that appear to be used in the tweets are “fight” “fighting”, the very same word “war”, “combat”, “threat”, and “battle”. All these words carry a very negative valence, of course, and denote aspects of the war that relate to actions and events. This is probably due to the stage of the pandemics that we are in, that is, the emergency situation, and the related urgent need to take action and confront the negative situation. We cannot exclude that this tendency may change, once the pandemic moves into a different stage. In particular, it could be the case that when the emergency has passed, and we will move toward the next phase, in which we will leave the peak of the emergency, the most frequent words used in relation to the WAR frame might relate to the identification of strategies to keep ourselves safe and to defend our community from potential new attacks.

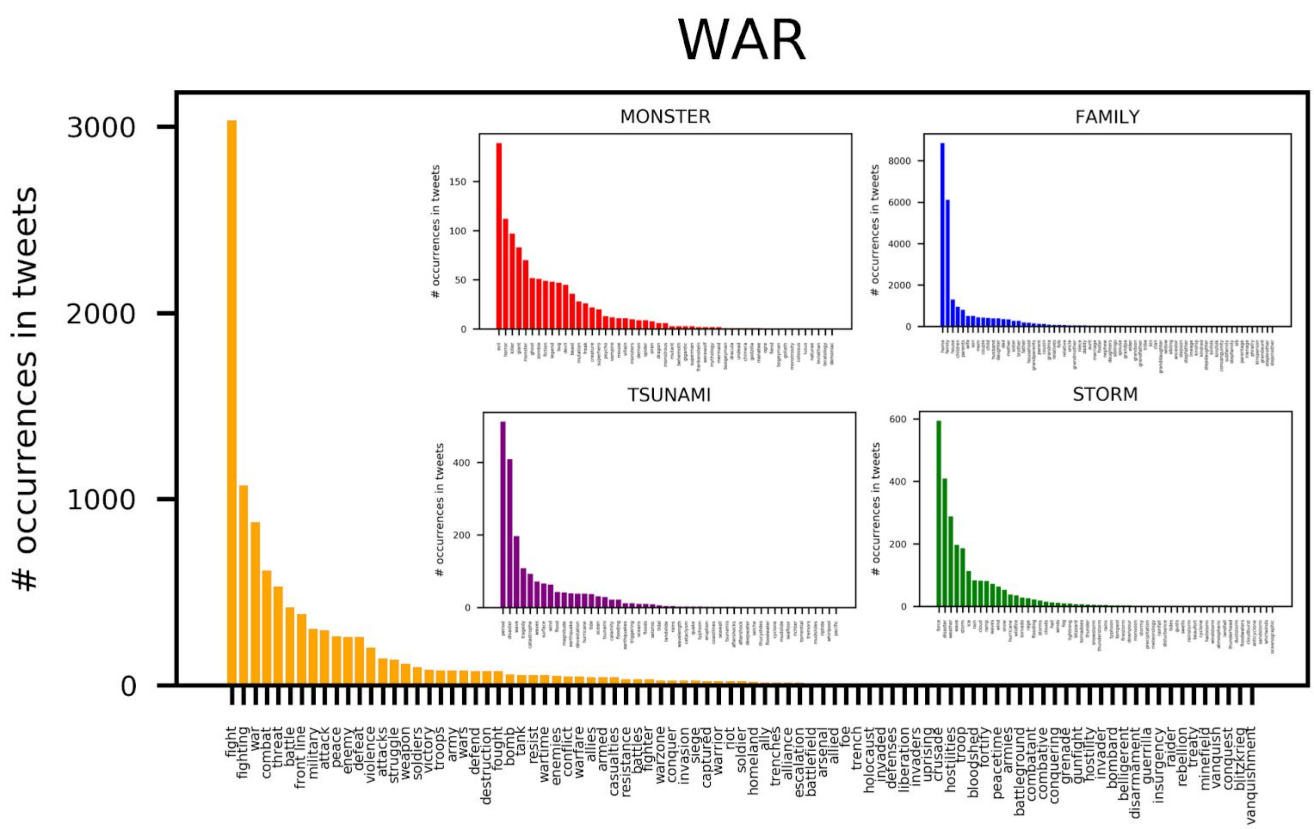
In relation to the topic modelling of the war-related tweets, we showed that tweets that feature war-related terms are most likely to belong to topics IV, I and III, rather than to topic II. Interestingly, topic IV addresses aspects related to the reactions to the epidemics, including the measures proposed by the governments and taken by the people, such as self-isolating, staying at home, protecting our bodies and so forth. Our analysis therefore suggests that using war-related words is a communicative phenomenon that we use to express aspects of the Covid-19 epidemic related to the measures needed to oppose (fight!) the virus. Moreover, tweets that feature war-related words are also often classified within the topics I and III, which include the aspects related to communications and reports about the virus, and politics. We interpret these results arguing that public communications and political messages are likely to frame the discourse in the WAR framing. Finally, it might not come as a surprise the fact that topic II, which encompasses aspects of the discourse related to the familiar sphere, the community and the social compassion, does not relate well with the tweets containing war terms.

The fine-grained analysis into 16 topics shows some interesting trends as well. In particular, tweets containing war-related terms are particularly well represented in topics 2, 7, and 10. Topic 2 seems to relate to online learning and education, Topic 7 encompasses aspects related to the treatment of the virus, with words such as “workers”, “health”, “care”, “help”, “thank”, “need”, “support”. Similarly, topic 10 relates to the diagnostics and treatment of the virus, with words such as “positive”, “death”, “cases”, “tested”, “people”, “confirmed”. Therefore, as the MetaNet WAR metaphor suggests, and as we described in the Theoretical Background of this paper, it is the discourse around the disease treatment and its diagnostics that are likely to be framed figuratively in terms of a war. Conversely, topic 3, which is characterized by words like “friends”, “share”, “trying”, “family”, “time” and therefore addresses intimate social relations and personal affective aspects related to Covid-19, is not related to the WAR frame: tweets addressing these aspects do not employ military lexical units.

## Study 3: Alternative Framing Results

In general, the terms of the alternative framing STORM were found in 2394 tweets (1,17% of all tweets). The alternative framing MONSTER was found in 1077 tweets (0.53% of all tweets). The alternative TSUNAMI was found in 1905 tweets (0.93% of all tweets). The literal frame used as control, that is the FAMILY frame, was found in 21452 tweets (10.53% of all tweets). We then looked at the distribution of the frequencies by which the terms within each framing were used and observed that they all tended to follow Zipf distributions (see Fig 6, where the term “fight” from the WAR frame has more than 3,000 occurrences in the tweets). In other words, within each frame there were few words used very frequently, but many words were rarely used. Moreover, although this is not visible on the plot in Fig 6, in the online repository we stored the full list of lexical units within each frame. Among the top ranked ones for the FAMILY frame we found “home”, “family”, “house”, “children”, “parents”, “wife”, “son”, and “mom”. For the STORM frame among the most frequently used lexical units we found “force”, “disaster”, “wave”, “storm”, “cloud”, and “waves”. For the MONSTER frame among the most frequently used lexical units we find “evil”, “horror”, “killer”, “giant”, “monster”, “ghost”, “zombie”, “devil”, “beast”, “mutation” and “freak”. Finally, for the TSUNAMI frame among the most frequently used lexical units we found “wave”, “waves”, “tide”, “tsunami”, and “flooding”.

**Fig 6: Five histograms depicting the occurrences of terms for each frame within the corpus.**




Also, the total number of words within each frame was different, with the WAR frame featuring more words than the other figurative frames. In order to compare how frequently was the WAR frame used in Covid-19 discourse,compared to other possible figurative frames (and a literal frame), it was necessary to have lists of lexical units of the same length because longer lists could have yielded larger numbers of tweets in the corpus than shorter lists, in principle. Therefore, we decided to evaluate two subsets of the term lists for each framing, setting two cutoffs at N=30 and N=50 terms on each list. In this way, we only considered the top 30 and then 50 most relevant (i.e., most frequently used) terms within each frame. We then compared the number of tweets featuring words from these lists which were now comparable in length.

Table 2 reports the number of tweets featuring at least one lexical unit related to a frame, and the general percentage of tweets in the corpus that can be related to these frames. Results show that the literal frame FAMILY is substantially more frequently used in the discourse on Covid-19 than the figurative frames. However, among figurative frames, the WAR frame covers a higher portion of the tweets in our corpus than the other figurative frames. The table also shows that there is no substantial difference between the coverages of the corpus obtained using the 30 words and the 50 words lists of lexical units for each frame.

**Table 2: Proportions of tweets that contain at least one of the terms from each of the frames with term list size N=30 and N=50.**

| Frame | # of tweets with at least 1 word from 30-item list | Percentage of tweets over the whole corpus (30 terms) | # of tweets with at least 1 word from 50-item list | Percentage of tweets over the whole corpus (50 terms) | Total Tweets |
|---|---|---|---|---|---|
| **WAR** | 8969 | 4.40% | 9407 | 4.62% | 203756 |
| **FAMILY** | 21201 | 10.41% | 21450 | 10.53% | 203756 |
| **STORM** | 2380 | 1.17% | 2394 | 1.17% | 203756 |
| **MONSTER** | 1065 | 0.52% | 1077 | 0.53% | 203756 |
| **TSUNAMI** | 1897 | 0.93% | 1905 | 0.93% | 203756 |

## Study 3: Alternative Framing Discussion

Our results show that the literal frame used as control (FAMILY) covers a wider portion of the tweets in the corpus while the figurative ones cover substantially less tweets. This is not particularly surprising, as previous literature shows that metaphor-related words cover only a percentage of the discourse, and that literal language is still prevalent. However, we show that in the Covid-19 discourse the WAR frame is particularly well represented, compared to other figurative frames that we operationalized in this set of studies, that is the MONSTER, the STORM and the TSUNAMI, which appeared to be relevant frames associated with the discourse around diseases and viruses according to the Metaphor Magnet.

Within the FAMILY (literal) frame, the top words (i.e., most frequent words) that are used in the tweets denote family members and family relations. Within the STORM frame, words that suggest the most frequently used words seem to denote concrete entities that can be typically observed within a storm scenario. In general, from a qualitative observation of the words that belong to these scenarios and appear to be used on Twitter to frame the discourse around Covid-19, it emerges quite clear the fact that different frames are used to tackle different aspects associated with Covid-19. Words in the STORM and in the TSUNAMI frames seem to relate to events and actions associated with the arrival and spreading of the pandemic. Words within the MONSTER framing, instead, are mostly nouns and can be arguably used to frame the discourse about the behavior of the virus, in a rather personified way, which is loaded with emotional content and extremely negative valence. This phenomenon, overall, supports the idea that different frames may be apt to elaborate the discourse around different aspects, related to a topic.

# Conclusion

In this study we explored the discourse around Covid-19 in its manifestation on Twitter. We addressed three specific research questions: 1. What are the topics around which the Twitter discourse revolves, in relation to Covid-19; 2. To what extent the pervasive WAR framing is used to model the Covid-19 discourse on Twitter, and specifically in relation to which topics does this figurative framing emerge; 3. To what extent does the WAR framing compare to other potentially relevant figurative framings related to the discourse on viruses, and to the literal framing FAMILY.

Through the implementation of three studies we answered these three research questions and discussed our specific findings. In general, we found that the topics around which most of the Twitter discourse revolves, in relation to Covid-19, can be labelled as *Communications and Reporting*, *Community and Social Compassion*; *Politics* and *Reacting to the epidemic*. A more fine-grained analysis brings to light topics related to the treatment of the disease, mentioning people involved in this operation such as doctors and nurses, and topics related to the diagnostics of the virus. In our second study we found that these specific topics appear to be those in which the WAR framing is particularly relevant: most lexical units within the WAR frame are found in tweets that get automatically classified within the specific topics of virus treatment and diagnostics. Moreover, in study 2 we observed that there is a very little number of lexical units related to war that are very frequently used, while the majority of war-related words are not used to frame the discourse around Covid-19. The more frequently used words refer to actions and events, such as “fighting”, “fight”, “battle”, and “combat”. As we anticipated in study 2, this might be a peculiarity of the stage of the pandemic we are currently living, which is the peak of the emergency. We do not exclude that with the development of the epidemics and the passage to the next phase (i.e., leaving the peak) also the most frequent words used within the WAR frame will change, to exploit new aspects of this frame that are relevant to the new situation. Finally, in study 3 we compared the frequency by which the WAR frame, the FAMILY literal frame and three other figurative frames are used. We found that while the FAMILY literal frame used as control covers a wider portion of the corpus, among the alternative figurative frames analysed (MONSTER, STORM and TSUNAMI), the WAR frame is the most frequently used to talk about Covid-19, and thus, arguably, the most conventional one, as previous literature also suggests.

Taken together our results confirm the pervasiveness of the WAR frame also in the discourse on Covid-19, as previous literature would have predicted, given the frequent use of this frame in discourses on diseases and viruses. However, we have also found that this frame is used specifically to talk about specific aspects of the current epidemic, such as its treatment and diagnostics. Other aspects involved in the epidemic are *not* typically framed within a WAR. This point is particularly important. The WAR frame, like any other frame, is useful and apt to talk about some aspects of the pandemic, such as the treatment of the virus and the operations performed by doctors and nurses in hospitals, but not to talk about other aspects, such as the need to feel our family close to us, while respecting the social distancing measures, or the collaborative efforts that we should undertake in order to #flattenthecurve, that is, diluting the spreading of the virus over a longer period of time, so that hospitals' ICU departments can work efficiently without getting saturated by incoming patients. In this sense, future studies could focus on the systematic identification of alternative figurative framings actually used in the Covid-19 discourse to tackle different aspects of the epidemic, but could also focus on the generation of additional frames, which can help communities to understand and express aspects of this situation that cannot be expressed by the WAR frame. A collection of different frames and metaphors that tackle different aspects of the current situation, or a *Metaphor Menu* (http://wp.lancs.ac.uk/melc/the-metaphor-menu/), as Semino and colleagues proposed in relation to cancer discourse [15], is arguably the most desirable set of communicative tools that, as language, communication, and computer scientists, we shall aim to construct in these current times, as a service to our communities.

# Acknowledgments

The authors would like to thank all doctors, nurses, health-care workers, grocery store workers and anyone else at the frontline of this epidemic.

# References

[1] Flusberg SJ, Matlock T, Thibodeau PH. War metaphors in public discourse. Metaphor and Symbol 2018 33.1. 1-18. doi: 10.1080/10926488.2018.1407992

[2] Culotta A. Towards Detecting Influenza Epidemics by Analyzing Twitter Messages. Proceedings of the First Workshop on Social Media Analytics. SOMA'10. New York, NY, USA: ACM; 2010. p. 115–122. doi: 10.1145/1964858.1964874.

[3] Gomide J, Veloso A, Meira W Jr, Almeida V, Benevenuto F, Ferraz F, et al. Dengue Surveillance Based on a Computational Model of Spatio-temporal Locality of Twitter. Proceedings of the 3rd International Web Science Conference. WebSci'11. New York, NY, USA: ACM; 2011. p. 3:1–3:8. doi: 10.1145/2527031.2527049.

[4] Pruss D, Fujinuma Y, Daughton AR, Paul MJ, Arnot B, Szafir DA, Boyd-Graber J. Zika discourse in the Americas: A multilingual topic analysis of Twitter. PloS one, 2019 14(5). doi: 10.1371/journal.pone.0216922

[5] Paul MJ, Dredze M, Broniatowski D. Twitter Improves Influenza Forecasting. PLoS Currents Outbreaks. 2014. doi: 10.1371/currents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117

[6] Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining Search, Social Media, and Traditional Data Sources to Improve Influenza Surveillance. PLOS Comput Biol. 2015;11(10):e1004513. doi: 10.1371/journal.pcbi.1004513

[7] Ji X, Chun SA, Geller J. Monitoring public health concerns using Twitter sentiment classifications. IEEE International Conference on Healthcare Informatics; 2013. pp. 335-344.

[8] Smith M, Broniatowski DA, Paul MJ, Dredze M. Towards Real-Time Measurement of Public Epidemic Awareness: Monitoring Influenza Awareness through Twitter. AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content; 2016.

[9] Mollema L, Harmsen IA, Broekhuizen E, Clijnk R, De Melker H, Paulussen T, et al. Disease detection or public opinion reflection? Content analysis of tweets, other social media, and online

newspapers during the measles outbreak in the Netherlands in 2013. Journal of Medical Internet Research. 2015; 17(5): e128. doi: 10.2196/jmir.3863

[10] Lazard AJ, Scheinfeld E, Bernhardt JM, Wilcox GB, Suran M. Detecting themes of public concern: A text mining analysis of the Centers for Disease Control and Prevention's Ebola live Twitter chat. American Journal of Infection Control. 2015. doi: 10.1016/j.ajic.2015.05.025

[11] Tran T, Lee K. Understanding citizen reactions and Ebola-related information propagation on social media. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016. pp. 106-111.

[12] Miller M, Banerjee T, Muppalla R, Romine W, Sheth A. What Are People Tweeting About Zika? An Exploratory Study Concerning Its Symptoms, Treatment, Transmission, and Prevention. JMIR public health and surveillance. 2017; 3(2):e38. doi: 10.2196/publichealth.7157

[13] Vijaykumar S, Nowak G, Himelboim I, Jin Y. Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter. American Journal of Infection Control. 2018; doi: 10.1016/j.ajic.2017.10.015

[14] Sontag S. Illness as Metaphor. Allen Lane, London. 1979.

[15] Semino E, Demjén Z, Demmen J, Koller V, Payne S, Hardie A et al. The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals:

a mixed methods study. BMJ supportive & palliative care 2017 7.1: 60-66. doi: 10.1136/bmjspcare-2014-000785

[16] Hendricks, R. K., Demjén, Z., Semino, E. and Boroditsky, L. Emotional implications of metaphor: Consequences of metaphor framing for mindset about cancer, Metaphor and Symbol, 2018 33:4. 267-279. doi: 10.1080/10926488.2018.1549835

[17] Karlberg M, Buell L. Deconstructing the 'war of all against all': The prevalence and implications of war metaphors and other adversarial news schema in TIME, Newsweek, and Maclean's. Journal of Peace and Conflict Studies 2005. 12.1: 22-39.

[18] Dodge EK, Hong J, Stickles E. MetaNet: Deep semantic automatic metaphor analysis. Proceedings of the Third Workshop on Metaphor in NLP. 2015.

[19] Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D. Fake news on Twitter during the 2016 US presidential election. Science, 2019. 363(6425), 374-378.

[20] Williams ML, Burnap P, Sloan, L. Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. Sociology, 2017. 51(6), 1149-1168.

[21] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of machine Learning research, 2003 3(Jan), 993-1022.

[22] Stone B, Dennis S, Kwantes PJ. Comparing methods for single paragraph similarity analysis. Topics in Cognitive Science 2011. 3(1), 92-122. doi: 10.1111/j.1756-8765.2010.01108.x

[23] Le Q, Mikolov T. Distributed representations of sentences and documents. In International conference on machine learning 2014 (pp. 1188-1196).

[24] Liu H, Singh P. ConceptNet—a practical commonsense reasoning tool-kit. BT technology journal, 2004 22(4), 211-226.

[25] Choi WG, Lee KS. Conceptual Representation for Crisis-Related Tweet Classification. Computación y Sistemas. 2019 23(4).

[26] Veale T. A service-oriented architecture for metaphor processing. Proceedings of the Second Workshop on Metaphor in NLP. 2014. pp. 52-60.