(Article begins on next page)

23

THE COVID-19 INFODEMIC ON TWITTER

Dialogic contraction within the echo chambers

Marina Bondi and Leonardo Sanna

Introduction

Fake news and misinformation are a key topic when discussing social media analysis research. While they had also been discussed well before the Trump era (Marchi 2012; Saez-Trumper 2014; Frank 2015; Conroy et al. 2015; Zhou et al. 2015), they became a central concern after Trump's election as president of the United States in 2016. The majority of studies on the theme focused on the issue of fake news detection (Conroy et al. 2015; Pérez-Rosas et al. 2017; Shu and Liu 2019), while media studies also contributed a lot with their interdisciplinary perspective (Albright 2017; Vargo et al. 2018; Melchior and Oliveira 2022).

The centrality of these themes has manifested itself not only because of the rise of populist movements, but also because of the key role that information plays in contemporary society. Indeed, we might say that we now live in an information society (Floridi 2009) where the good functioning of us as social beings is ensured by the delivery of a huge amount of information at an incredible speed. The web and the social media have realised a second mass mediatic revolution after that experienced after World War II with the wide diffusion of television. Although the information society has relatively recent origins, as early as the 1980s there was talk of "information overload", with a focus on the cognitive distress brought about by continuous immersion in the media information flow (Wurman 1989).

The rise of digital platforms such as social media, however, has added several layers of complexity to the media environment, intensifying experiences of remediation (Bolter and Grusin 2000), that is, the transfer of typical features of traditional media to digital media and increasing the amount of information available. Notwithstanding this complexity, social media also eased access to this huge amount of information, while at the same time introducing new actors as gatekeepers, sometimes also very different from traditional mass media players. This is the case of emerging (sometimes also called "alternative") media, as well as automated content selection, namely, algorithmic personalization; theoretically, these have been called in to solve the cognitive overload problems caused to us by the enormous amount of information available.

In this context, we have recently experienced information hysteria phenomena for which the term "infodemic" has been coined (Asif et al. 2021; Patwa et al. 2021; Petropoulos and Makridakis 2020; Eysenbach 2020). This rapid spread of potentially harmful information is often fostered on

digital platforms. Linguists have paid increasing attention to this topic since 2016, as the Oxford dictionary chose "post-truth" as word of the year (McIntyre 2018). The term refers to "relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief".

Of course, this definition is rather problematic when dealing with language. In fact, in natural language, we do not have a clear-cut definition of what "objective facts" mean, as we can only deal with representations (or discourses) on objective facts or supposed-to-be objective facts. As rightly pointed out by Russell (1940), "truth and falsehood apply primarily to beliefs, and only derivatively to sentences as 'expressing' beliefs" (Russell 1940, p. 214). This helps us understand the foundation of misinformation spreading as a process of social transformation in which trust boundaries are being renegotiated. In Greimas, for example, the assumption is that communication is an exchange process where there are two kinds of "cognitive doing" (Greimas 1989, p. 659): the addresser realises a persuasive act while the addressee carries out an interpretive act. These two parts do not realise the communicative exchange by that very fact: they need a contract, which does not have a cognitive nature but rather fiduciary; Greimas calls it veridiction contract. We might argue that this perspective is almost too relativistic since, in its strictest interpretation, it would not allow the introduction of any concept of truth outside of a relationship of mutual trust between the two parties involved. Russell, on the other hand, provides a useful tool enabling us to identify textual affordances by highlighting the notion of verifiability as distinct from that of truth: "'true' is a wider concept than 'verifiable', and, in fact, cannot be defined in terms of verifiability" (Russell 1940, p. 227).

The concept of verifiability is crucial as it points to textual structures whose function might be to establish a trusting relationship. We might say then that, when dealing with journalism and fake news, "true" is everything that follows the correct methodology in the verification of sources, while "post-truth" is rather a text aiming to grasp the trust of their readers by other means rather than verifiability.

There are contexts where discourse, while still allowing for correct identification of sources and voices called into question, constructs trust and credibility mostly by emphasising shared identities and positions, usually in opposition to other views. This is what is observed online in echo chambers (Sunstein 2007), that is, particular social structures in which the views of others are systematically rejected and used instrumentally to support one's own beliefs in a polarised debate. Echo chambers are thus contexts in which the ideological-emotional plane plays a much more important role than the truth and verifiability of a text. The notion was introduced by Sunstein in terms of polarisation among groups:

> Group polarization is unquestionably occurring on the Internet. From the evidence thus far, it seems plain that the Internet is serving, for many, as a breeding group for extremism, precisely because like-minded people are deliberating with greater ease and frequency with one another, and often without hearing contrary views.
>
> (Sunstein 2007, p. 69)

An important epistemological distinction can be drawn between echo chambers and divergence of opinion, namely, epistemic bubbles (Nguyen 2020), by highlighting the active dimension of echo chambers in the veridiction process:

> Loosely, an epistemic bubble is a social epistemic structure in which some relevant voices have been excluded through omission. Epistemic bubbles can form with no ill intent,

through ordinary processes of social selection and community formation. We seek to stay in touch with our friends, who also tend to have similar political views. But when we also use those same social networks as sources of news, then we impose on ourselves a narrowed and self-reinforcing epistemic filter, which leaves out contrary views and illegitimately inflates our epistemic self-confidence. An echo chamber, on the other hand, is a social epistemic structure in which other relevant voices have been actively discredited.

(Nguyen 2020, p. 142)

It is thus not the simple opposition between two semantic frames that would define an echo chamber but rather the active rejection of the counterpart.

Echo chambers are often seen as dysfunctions of mass media communication in computational social science (Del Vicario et al. 2016b; Zollo et al. 2017; Di Marco et al. 2021). They are presented as structures that foster the polarisation of debates and the spreading of conspiracy theories. From a linguistic perspective, we might see echo chambers as ideological structures (Eco 1968) that emerge when ideological conflict happens (Rogers 2018). At a lexical level, this would rather produce effects on all those elements of evaluative language that we call dialogic elements, as the distinctive features of an echo chambers are ideological isolation and the extreme polarisation of two irreconcilable positions.

Since echo chambers are essentially the main digital structures in which disinformation and misinformation spread most easily (Del Vicario et al. 2016a), we believe that it is important to explore the role that social media may play in spreading misinformation by looking at evaluative language in their discourse. In order to explore this, we have focused on Twitter. If it is true that affordances like emojis, mentions (@), hashtags (#), retweeting are thought to facilitate dialogue and enhance "connectedness" among people, creating affiliation among people with similar ideological interests (Zappavigna 2011), Twitter has also often been described as having given birth to a new style of news coverage, where users tend to follow politically homogeneous clusters and are unlikely to be exposed to cross-ideological content (Himelboim et al. 2013). In general, there is a large consensus that Twitter may potentially contribute to partisan polarisation (Hong & Kim 2016). Twitter language has been critically explored for ideology construction and dissemination of racist and nationalist ideologies (Boukala & Dimitrakopoulou 2018; Bartlett et al. 2014; Wodak & Boukala 2015; Farkas et al. 2018; Chaudhry 2015). Trump has also been said to use it to establish credibility by denouncing media criticism (Ross & Rivers 2018) in a strategy that Lakoff (2017) defined in an interview as a strategy of "deflection", that is, attacking the "messenger" rather than responding to the accusations.

The focus of our analysis is on the language that manifests the writer's position. Focusing on evaluative language, we adopt Martin and White's (2005) appraisal framework, and in particular the concept of engagement, indicating the speaker's degree of commitment to what is being expressed and manifesting the attitudes projected by authors or speakers. The basic interplay is the one between two major discursive voices, *monogloss* and *heterogloss*. Monogloss voices present facts that tend to concede no room for the negotiation of meaning but rather to elicit confidence in the statement, by presenting themselves as self-evident. Heterogloss' voice, on the other hand, overtly presents propositions as one among others, using a variety of linguistic resources opening or closing options for dialogue (Martin & White 2005, p. 100). This may lead to opening up dialogic space for different positions and arguments advanced by interlocutors (dialogic Expansion) or rather to deflect alternative views (dialogic Contraction) (2005, pp. 102–104). An expression like "I believe that . . ." could then be taken to represent the speaker's acknowledgement of

alternative views (expansion), while an expression like "I know that . . ." signals that there is no interest in alternative views (contraction).

Our focus will be on heteroglossic forms of dialogism rather than on monogloss discourse. The presence of a monogloss voice in Twitter discourse would not in itself point to an echo chamber effect but rather to the centrality of positioning oneself on social media. By comparing tweets and news articles we have, for example, already noticed a major disproportion in the space given to reported discourse – inevitably heteroglossic – in the news as against direct claim-making on Twitter (Bondi & Sanna 2022). A focus on the dialogic dimension of Twitter, on the other hand, might provide insights into the nature of the dialogue that takes place on the social medium: how far do tweeters open or close dialogic space in their rhetoric? By looking at how different expressions are used to open up dialogic space or rather to deflect alternative views, we hope to be able to explore the nature of the dialogic space created.

The hypothesis that guides the study is that features of dialogic contraction may characterise Twitter discourse, thus confirming the opinionated, polarised, echo chamber effect that is often attributed to it. If it is true that Twitter now characterises itself as a news site, rather than as a social network, and many Twitter users get news on the site (especially for breaking news), our assumption is that the space given to dialogic contraction on Twitter may be wider than that provided by traditional journalism, even when considering that journalism tends to draw attention to issues of contention and report debates. The hypothesis is that, from a linguistic perspective, echo chambers manifest themselves in the dialogic dimension, restricting the space for external views. However, echo chambers work always in two directions: on the one hand, they restrict the dialogic space leading to ideological isolation, while on the other hand they actively discredit and refuse the counterpart. The study thus explores the hypothesis by comparing the use of dialogic contraction and expansion in Twitter and in news discourse.

The chapter is organised as follows. The next section illustrates the data and the methodology. The results of the overall quantitative analysis are presented in section 3, while section 4 focuses on the role of negation in dialogic contraction and section 5 centres on "know" and "think" as verbs of ideological positioning. Section 6 discusses the results of the different types of analyses and introduces some conclusions.

Data and methods

The study is based on two datasets, namely, a sample of the Coronavirus corpus (Davies 2019), available from English Corpora, and the COVID-19 Twitter dataset. The Coronavirus Corpus includes online newspapers and magazines in 20 different English-speaking countries. Although it is quite varied in sources, occasionally also including comments to the published articles, it is representative of the discourse of online news. It is built from a subset of the NOW Corpus (Davies 2016), a larger corpus of News collected from the daily scraping of more than 1,000 websites. All the articles containing at least two occurrences of the word "coronavirus", "COVID", or "COVID-19" are added to the Coronavirus Corpus.

On the other hand, the Twitter dataset corpus is a repository of an ongoing collection of tweet IDs associated with the COVID-19 outbreak, whose collection started at the end of January 2020. This corpus is collected by searching a set of COVID-keywords via Twitter's search API, as reported by Chen et al. (2020). As recommended by the authors (Chen et al. 2020), we used the software Hydrator to collect the full text of each tweet. Hydrator allowed us to collect the full text starting from the tweet IDs (a unique reference that is used to identify each tweet) provided

in our sample of data; in fact, as per Twitter's Terms and Conditions, the sharing of full text is not allowed, but only the sharing of Tweet IDs.

Our analysis is based, for both corpora, on a subset that includes the first seven months of the pandemic outbreak. This choice is dictated by the idea that, at that particular time, it should be possible to observe the emergence of misinformation phenomena and echo chambers, within the neurotic context of infodemics. For the Coronavirus corpus we included all the articles written between January 2020 and July 2020, while we extracted a sample of one million tweets for each month in the same range of time within our Twitter dataset.

From now on we will be referring to our corpora as the News Corpus, for the Coronavirus Corpus, and as the Twitter Corpus for our social media dataset.

The study combines the approach of corpus-assisted discourse analysis (Baker 2006; Baker et al. 2013; Bednarek 2008; Partington et al. 2013; Scott and Tribble 2006) with word embedding as a means to explore the dialogic dimension of echo chambers within the appraisal framework. In particular, corpus-assisted discourse analysis is used to explore the semantic preference and collocations of the appraisal markers in our corpora corpus using Wordsmith Tools 8 (Scott 2020). On the other hand, word embedding (Mikolov et al. 2013a, 2013b) is used to enhance the qualitative exploration of the pragmatic status of the most significant words. Word embedding is a machine learning technique used in natural language processing to create a computational semantic model. Given a word, the semantic model is able to infer words that are most likely to co-occur with the given term. In this study, we are using word embedding as a tool for qualitative exploration (Sanna and Compagno 2020), assuming that the probabilities of co-occurrence could be seen as a semantic frame (Fillmore 1976; Eco 1979) and that therefore this would allow us to infer their pragmatic dimension exploring their semantic preference (Sinclair 2004; Hunston 2007).

The analysis was based on an adaptation of the list of appraisal markers made by Fuoli (2012, 2018). To minimise possible biases due to the nature of the original tagging process, we kept in our checklist only words that were not specifically related to Fuoli's type of data (corporate reports) but were generalisable to our own data, also avoiding multi-words expressions. This allowed us to have a list of common appraisal markers which could be used to compare and quantify the dialogic dimension in our two corpora. As we are interested in ideological structures that are used to express ideological positioning, the most appropriate type of appraisal to investigate is "Engagement", with markers of Contraction and Expansion.

Keeping this list in mind, we looked for the distribution of appraisal markers in both corpora. The idea behind this was to highlight relevant differences between the news and Twitter. We assume that Twitter is characterised by echo chambers but it is also perhaps a preferred space for discussion, and therefore what happens on Twitter is often quite relevant for public debate. If echo chambers do have peculiar characteristics such as the predominance of dialogic contraction, we might be able to highlight these differences using the News Corpus as a reference corpus.

Starting from the quantitative differences between the two corpora, we carried out a qualitative analysis of selected elements, with a view to identifying the lexico-semantic patterns that

Table 23.1 Corpus figures

	N. of Texts/Tweets	N. of Tokens	N. of Types
News Corpus	650,699	442,252,000	2,086,489
Twitter Corpus	7,000,000	152,468,080	2,638,855

surrounded them and that could help us reconstruct their engagement function, as well as their role in the argumentative dialogue underlying the text. In particular we explored the lexico-semantic dimension by looking at collocation and clusters, while we also used the wider context of concordances to explore the argumentative role of individual occurrences, to see whether these occurrences were actually pointing to instances of positioning with reference to the topic at issue (the debate over COVID). Finally, we also used word embedding to complement the analysis of selected lexical elements.

Analysis

A detailed comparison of the normalised occurrences of our markers, in alphabetical order, is presented in Table 23.2. The lexical elements present in both corpora are many and present significant differences. The higher figures of the pairs, highlighting the word forms that characterise the two corpora in the comparison, are foregrounded in bold.

As shown in Table 23.2, the engagement level is mostly higher – and unsurprisingly characterised by a wider range of word forms – in the news corpus. The Twitter corpus, on the other hand, features a marked preference for a few elements and a decided lack of argumentation markers.

Table 23.2 Normalised occurrence of appraisal markers on Twitter and in the news corpus

Token	Dialogic Function	PTTW Twitter	PTTW News
although	Contract	0.083	**0.762**
anticipate	Expand	0.005	**0.163**
assure	Contract	0.026	**0.113**
belief	Contract	0.044	**0.146**
believe	Contract/Expand	2.180	**2.817**
but	Contract	**8.555**	8.234
Clearly	Contract	0.049	**0.080**
confident	Contract	0.053	**0.367**
convinced	Contract/Expand	0.117	**0.226**
could	Expand	2.517	**5.269**
demonstrate	Contract	0.040	**0.154**
demonstrated	Contract	0.025	**0.164**
Despite	Contract	0.209	**0.658**
did	Contract	2.015	**3.111**
evidence	Contract	0.480	**0.750**
evident	Contract	0.023	**0.118**
expect	Expand	0.313	**0.989**
expected	Expand	0.294	**1.980**
found	Contract	0.778	**1.872**
However	Contract	0.216	**2.247**
Indeed	Contract	0.108	**0.326**
inevitable	Contract	0.068	**0.179**
knew	Contract	0.418	**0.438**
know	Contract	**4.178**	3.050
knowing	Contract	0.160	**0.269**
may	Expand	1.905	**4.689**

(Continued)

Proof

Marina Bondi and Leonardo Sanna

Table 2.3 (Continued)

Token	Dialogic Function	PTTW Twitter	PTTW News
might	Expand	0.983	**2.087**
Naturally	Contract	0.005	**0.025**
never	Contract	1.493	**1.560**
No	Contract	**2.242**	1.570
none	Contract	0.154	**0.349**
not	Contract	**12.694**	10.081
nothing	Contract	0.960	**0.797**
obviously	Contract	0.141	**0.325**
project	Expand	0.205	**0.630**
prove	Contract	0.124	**0.301**
recognising	Contract	0.006	**0.040**
recognise	Contract	0.062	**0.213**
reflect	Contract	0.064	**0.490**
reflected	Contract	0.010	**0.154**
reflecting	Contract	0.017	**0.112**
reflects	Contract	0.029	**0.171**
see	Contract/Expand	6.085	**7.162**
should	Expand	3.931	**4.396**
shows	Contract	0.719	**1.131**
stated	Expand	0.094	**0.682**
surely	Contract	0.107	**0.133**
think	Expand	**3.121**	2.403
understands	Contract	0.029	**0.130**
unthinkable	Contract	0.004	**0.028**
While	Contract	0.560	**2.393**
without	Contract	1.154	**2.847**
Yet	Contract	0.226	**0.361**

This emerges by comparing the normalised occurrences of the modal verbs ("could", "may") and other elements typical of the negotiation of the dialogistic dimension, namely, "however" and "although". Certainly, this could be a structural feature of the Twitter environment, where arguing is influenced by the character limit. Twitter also shows a prevalence of five word-forms, namely, a contrastive connector ("but"), two negations ("No" and "Not"), and two verbs ("know" and "think") that are clearly linked to the expression of cognitive processes.

We decided to leave aside a close analysis of "but", for its extremely complex nature, but there is no doubt that its role in "disclaiming" is central to acknowledging a multiplicity of positions, while at the same time rejecting one and taking a stance. We thus decided to take a closer look at the two negations and the two verbs, as they might be interesting markers of ideological conflict within the platform.

A closer study of concordances first of all confirmed the impression that many of the specific elements highlighted are very flexible in terms of their actual value: they do open up to heteroglossia, but whether they are used to enhance the possibility of a continued negotiation of meanings or to fend off that negotiation largely depends on context. Moving

Proof

away from the proposition to the entire clause complex or to the wider context of the debate helps to see the role they play as responses, either retrospectively (responding to previously expressed opinions) or prospectively (anticipating the interlocutor's response and including counter-responses).

The role of negation in dialogic contraction

Two of the five words identified as characterising Twitter discourse are "not" and "no". If the case of "no" is rather illustrative of a very versatile element which can be found in many different contexts, "not" is more clearly a verb-modifying adverb marking negative polarity, potentially making it easier for us to understand how it is used as a form of dialogic contraction in our corpus in explicit forms of denial. Because of its function in polarity, we can in fact be more precise in our exploration and directly select the most significant verbs that occur with "Not" in our corpus.

We selected the five most significant ones according to t-score (Oakes 1998; Hunston 2002), that is, the verb "to be" in the present tense ("is not"), the verb "to do" in the present tense ("does not"), the verb "should", the verb "to have" in the present tense ("have") and finally the verb "to be" in the past tense ("were not"). In general, it might be noticed that modals are frequently used in their negative form in the corpus, as the modals attested (in order of frequency *will, should, would, could, must, might, can't/cannot* and *shall*) represent approximately 8% (49897/ 625719) of the occurrences of *not*. In terms of collocational strength, however, *should* is definitely the one that most predictably collocates with *not*.

What emerges from an observation of the strongest collocates is that when "not" is used to negate a verb in the third person singular, it is largely part of dialogic contraction processes that point directly to the topics at issue (>90% of cases out of 200 random collocations for both), as clearly shown in the sequence of denials in Example 1:

(1) Just a reminder: Gates is NOT a doctor Gates is NOT an epidemiologist Gates was NOT elected.

In very few other cases, there is no explicit denial, for example when the negation is embedded in a relative clause that simply selects a specific scenario, as in Example 2:

(2) Close the borders. Any citizen not back already frankly needs to accept they should have be back earlier.

In the case of "does not" and "is not", the negation is used ~~in~~ most ~~cases~~ to deny epistemic validity to voices outside the dialogue or the direct counterpart we are addressing (Example 1), sometimes also with a direct reference to veridical aspects:

(3) 5G does not spread coronavirus.
(4) Information circulating on social media on a confirmed case of COVID 2019 Corona Virus is not true.

In the proposed examples, negation of the verbal element is used to position oneself at one of the two poles of the debate, simultaneously narrowing the dialogic space.

Apart from the most frequent collocations with the verb forms mentioned (*is/does/should/have/were*), we further explored the use of the negative element in itself. The study of a randomly selected set of 200 occurrences of "not" can offer insights on its use. The choice of limiting the analysis to 200 occurrences is often the standard in corpus linguistics when numbers are very high (see Groom 2010, pp. 64–65 for a discussion of the measure). Almost 70% of the occurrences in this small random sample (138/200, i.e. 69%) are actually direct forms of denials (Example 5), whereas another 4.5% (9/200) are weakened forms of dialogic contraction, as they are presented in conjunction with an epistemic modal signalling dialogic expansion (Example 6). Other cases, such as those of a few rhetorical questions (6/200, i.e. 3%) might also be considered forms of dialogic contraction, though in the form of a strong (affirmative) claim, implied by the negative question (Example 7):

(5) Seriously people – STOP BUYING MASKS! They are NOT effective in preventing general public from cat.
(6) It might not be a regular respiratory virus.
(7) So yall are NOW washing your hands bc the coronavirus outbreak??? Was this not a thing for yall before??

Of the rest, about 11% (22/200) are forms of reported discourse (mostly reporting decisions) (Example 8), whereas 6% (12/200) are found in conditional clauses presenting a possible case, qualifying a direct or indirect claim (Example 9), 7.5% (15) are found in expressives, manifesting a personal state of mind, which is not in itself subject to any verifiable truth claim (Example 10), and 4% (8/200) are found in imperatives expressing recommendations (Example 11), and another one was an example of a recommendation with the use of "why not":

(8) Trump Administration Says Planned Parenthood Will Not Receive Coronavirus Aid.
(9) Don't worry about the coronavirus, you'll be fine if you're not elderly or vulnerable.
(10) It's 40 days into stay-at-home and I'm not sure I'm entirely sane.
(11) Remember that Covid-19 takes a week to two weeks to incubate. do not let the media fool you when they say these 19,000 cases.

A further level of analysis might want to investigate the nature of the claims and counterclaims involved in the occurrences of "not", looking at their argumentative structures. Given the fragmentary nature of the data, this is of course much more complex to ascertain, let alone quantify. The overall impression, however, is clearly that there are very few cases where the denial is part of a main rejection claim that is also supported by verifiable arguments (Example 12). In some other cases, denials are an unsupported (and unverifiable) subclaim, supporting another conclusion (Example 13), but in most cases they are bare contradictory rejections (Example 14 and 15).

(12) COVID cases are rising and people are getting sick. This is NOT the time to stop social distancing.
(13) China is doing everything right at the moment. Restrictions on travel and major gatherings! The coronavirus could not be anticipated to travel so far so quickly, people didn't know they were infected!
(14) No it's not, pandemic was never what it was reported.
(15) The news regarding the negative test of the coronavirus patient is false. He has not been discharged. He did test positive.

Table 23.3 Top ten three-word clusters of "No" on Twitter

N	Cluster	Freq.
1	THERE IS NO	17,498
2	THERE ARE NO	4,394
3	NO HEALTH CARE	4,257
4	OUT WITH NO	4,179
5	NO TUITION REFUND	4,151
6	WITH NO TUITION	4,148
7	UNEMPLOYMENT NO HEALTH	4,143
8	NO SOCIAL DISTANCING	3,893
9	NO ONE IS	3,386
10	NO AVAILABLE VACCINE	2,554

Table 23.4 Top ten three-word clusters of "No" in the news

N	Cluster	Freq
1	THERE IS NO	49,950
2	THERE S NO	20,773
3	THERE ARE NO	15,623
4	THERE WAS NO	15,421
5	WILL BE NO	6,890
6	THERE WILL BE	6,562
7	WE HAVE NO	6,559
8	THERE WERE NO	6,387
9	IS NO LONGER	6,109
10	NO MORE THAN	5,795

It is worth spending a few more words on the use of "No", the other negative element highlighted in Table 23.2. It may be interesting to look at the top ten clusters in which it occurs in both corpora, to compare Twitter discourse and news discourse.

The data highlight the importance of negative existential patterns. As these existential clauses might be used to contract the enunciation space within the discourse in forms of *denial* (Martin and White 2005), we should not be surprised by their abundance in a context such as the pandemic since most of the discourses are initially centred on the account of the existence of the virus. The most obvious context is one of dialogic contraction, supporting a discourse of denial of the exist-ence of the virus or of the danger of the virus. The word form "no", however, might also be used in other positions, where it is less likely to be a good indicator for dialogic contraction, or at least not explicitly related to the issue that is object of debate.

Taking this possibility into account, we took a sample of 200 occurrences of the top-occurring clusters in both corpora, namely, "There is", to evaluate how many times it was actually used to introduce dialogic contraction in relation to the topic. Again, in the Twitter corpus we found a more marked conflict, with a greater use of "No" as contractor of dialogic space, with 48 occurrences (25%) signalling forms of positioning by denial on the COVID issue, vs 30 occurrences (15%), in the news. In both cases, however, the vast majority of our top-occurring clusters do not act as markers of specific ideological positioning, that is, they do not constitute claims about the (non-)

existence of COVID but may for example simply provide supporting evidence (e.g. "There was no immediate reaction from the White House"). This confirms that our bottom-up approach to appraisal requires analysis of the wider context and might not provide enough elements for any automatic identification of positions in a debate.

Indeed, as suggested by Hunston (2004), we are dealing with a subject that is rather difficult to grasp in a quantitative evaluation since the restriction of dialogic space can also occur by means of allusions that are expressed with complex sequences of words on the discursive level. When these do not occur with regularity, they are difficult to detect by studying the collocations of lexical elements. Moreover, many lexical items on which we are basing our observations on appraisal can certainly occur in contexts where their function is not directly related to the object of debate.

Nonetheless, it is interesting to select some examples from the top occurring cluster in Twitter, that is, "There is no", to illustrate the variety of elements that can be denied. See for example how the expression is used on Twitter to deny the existence of the virus (16), to sum up a position or report it (17), and to reject the validity of a policy (18).

(16) COVID 19 = Exosomes naturally found in all cells. Cells excrete in times of stress or illness There is no virus. Only flu etc.

(17) #CureCancer_By_TrueWorship There is no such disease which cannot be cured by the devo-tion of Sant Rampal Ji whether it is corona virus or cancer. All diseases can be removed, but true devotion.

(18) There's no point to let unlimited #coronavirus infected patients to enter Hong Kong when we are already running out of medical resources.

It appears that forms introducing a proposition (rather than just the existence of an entity or pro-cess), like Examples (17) and (18), are particularly apt at taking distance from this proposition and acting therefore as forms of denial from the point of view of engagement, but also as disagreement from the point of view of argumentative dialogue. In the specific case of the COVID infodemic, forms of denial of existence were also central to the debate, when directly related to the existence of the virus, for example.

It is important to notice, then, that the role of the single markers in determining a contraction of the dialogic space relies on the lexico-grammatical patterns involved, beyond the simple presence of a word. The relevance of these contractions is also determined by the link to specific arguments in the debate. The distinguishing features of Twitter discourse, however, appear to be in line with the definition of echo chambers, as the key strategy is seemingly the rejection of the opponent's view by mere denial.

"Know" and "think" as verbs of ideological positioning

The other two words characterising Twitter discourse are the verbs "think" and "know". While sharing a semantic relationship (i.e. representing a cognitive process), the verbs seem to play dif-ferent roles in the engagement system, as typically representative of expansion and contraction in combination with the first person singular ("I think"/"I know").

Looking at a small excerpt of 200 concordances for each of these word forms, we notice that "think" is used to introduce a personal standpoint (i.e. "I think") in 30% of the cases, while in the rest of the cases it is used to introduce external voices. The wider context shows, however, that this leads mostly to contexts of contraction (>90%), where external voices are introduced to be

discredited, especially when formulating rhetorical questions to negate the counterpart voice ("Do you think you know COVID so well?"), thus confirming a dominant contraction function.

The use of "know" shows a more complex dialogistic dimension, as in the vast majority of the cases (>80%). It is used in combination with negative polarity to express lack of knowledge ("I did not know that this happened") in ways that – in the wider context of the argument – do in fact cast doubt on what others have said. Similarly, when the process is attributed to others ("they don't know if the COVID cases will double"), the argumentative function appears to be that of discrediting external voices.

Beyond the local most obvious heteroglossic function, then, it is important to study the lexico-grammatical patterns around the node word and, even more clearly, the argumentative role that the proposition plays in the debate.

The verb "to know" is obviously linked to information and all the actions connected to bringing new information and knowledge to the debate. However, looking at its general collocates in the whole corpus, it is clear that on Twitter the situation is quite peculiar. Indeed, the most significant verbal element that occurs in the immediate context is a negation, that is, "don't", which obviously adds to the frequency of negations with "no" and "not" that we have already seen in the previous subsection. The significance of the collocates is calculated using t-score (Oakes 1998; Hunston 2002). One of the most recurring patterns is "I don't know who needs to hear that", influenced also by many retweets. What is evident, exploring the concordances (and in particular the propositions introduced by "know"), is that in these negative contexts "know" is almost always used as a means to express a standpoint and very rarely with epistemic intent.

(19) I don't know if I want my hair to come from China. I'm scared.
(20) I don't know if I'm keen to trust data in the middle of a pandemic where adequate testing hasn't been accomplished and attributable death totals are questionably.
(21) I don't know if I've seen anyone say this but my heart goes out to the people of China right now.

As seen in examples (19–20–21), in its most used form, the verb is used with a clear intent to position and almost never as a real recourse to epistemic modalities. A marked willingness to position oneself emerges from its use regardless of debating the veracity of what the stance presupposes. In the examples shown, the potential discussion of an epistemic status in the introductory clause (seemingly opening dialogic space) is only instrumental to strong positioning (narrowing dialogic space) in the second part of the sentence.

Special mention should be made of the locution "Don't know who needs to hear that, but", which is perhaps the clearest example of the use of the negation of the verb "to know" to narrow dialogic space. What this type of formulation implies is a shared knowledge, presenting an epistemic validity that is so shared as to be self-evident; this allows for very strong positioning, as in the case of examples (22) and (23), typically representing the two main positions in this debate, Democrats and Republicans.

(22) I don't know who needs to hear this, but unlike other developed countries, the U.S. really has not flattened the curve.
(23) I don't know who needs to hear this, but Joe Biden is compromised with China. They know Hunter's secrets.

Very similar conclusions might be reached for the verb "to think", often regarded as potentially opening dialogic space. Certainly the verb already naturally appears in contexts where personal

beliefs and convictions are expressed, as it is used to refer to our opinions in our cognitive sphere. On the other hand, it is interesting to note how the presence of this cognitive sphere tends to raise the level of ideological confrontation somewhat, especially in negative forms.

(24) You know, I don't think it's too complicated to wear a mask.
(25) You don't think the new world order would exploit something like covid-19 to subjugate us all into slavery, do you?
(26) I don't think coronavirus is caused by 5G. I think it's a complete and total hoax, and there is no new strains of illness.
(27) Raise your hand if you don't think covid-19 is serious. Now, use it to slap yourself because you're an idiot.
(28) Do you think your representative should be getting paid to stay at home and telework? Because that's what Democrats believe.
(29) Do you think Trump woke up today with an ache in his heart for all of the people suffering in the world. Prayed with Melania and her son over breakfast for the families effected by Coronavirus, and then went straight to his office to read the updated briefings to keep us safe?

There are indeed, as in the case of Examples 24 and 25, more or less direct attacks on the other party, either by means of strong criticism (26) or turpiloquy (27), or through an interesting use of rhetorical questions (25). Examples 28 and 29 are in fact questions that reveal ideological positioning. In other words, they are fake questions manifesting denial of the belief attributed to the other party. In this game of extreme positioning, typical of echo chambers, the counterpart is not regarded as legitimate, and their positions are in fact systematically denied at the semantic and morphosyntactic level.

Overall then, these two verbs are used with similar functions on Twitter, that is, representing a precise standpoint or belief. Looking at the word embedding model of both verbs in our corpus (Figure 23.1), we notice that the most similar word for "think" is "know" and vice versa, meaning that they hold a strong semantic and pragmatic relation. Indeed, in both semantic spaces, we do have the word "believe", confirming the hypothesis of representing a precise standpoint.

The semantic frame of both verbs also provides confirmation of their slightly different functions: "think" is used to take a strong position that might also end in an explicit attack to the counterpart, while "know" is used to delimit the ideological space with respect to ideological positioning and is often modulated with a negation, with the aim of excluding from the dialogic space anything that does not adhere to the proposed ideological positioning. The frames of "think" and "know" seem to show overlap of lexical elements afferent to the cognitive ("believe", "understand") and demonstrative areas ("see", "actually"), while differing markedly in other elements. In the case of "know" there are indeed elements related to epistemic modes and knowledge sharing, although we have seen that these are modulated instrumentally to take a position within the debate. In the case of "think", on the other hand, what characterises its pragmatic status is precisely the exaggeration of positioning, with related attack to the other side.

Discussion and conclusions

Our analysis has clearly shown important characteristics of Twitter discourse by contrasting it to traditional journalistic discourse. Twitter certainly cannot be regarded as an informational space, but it is clear that it has now been chosen as the preferred platform for discussion of texts with informational content. However, this discussion does not take place with the tools of
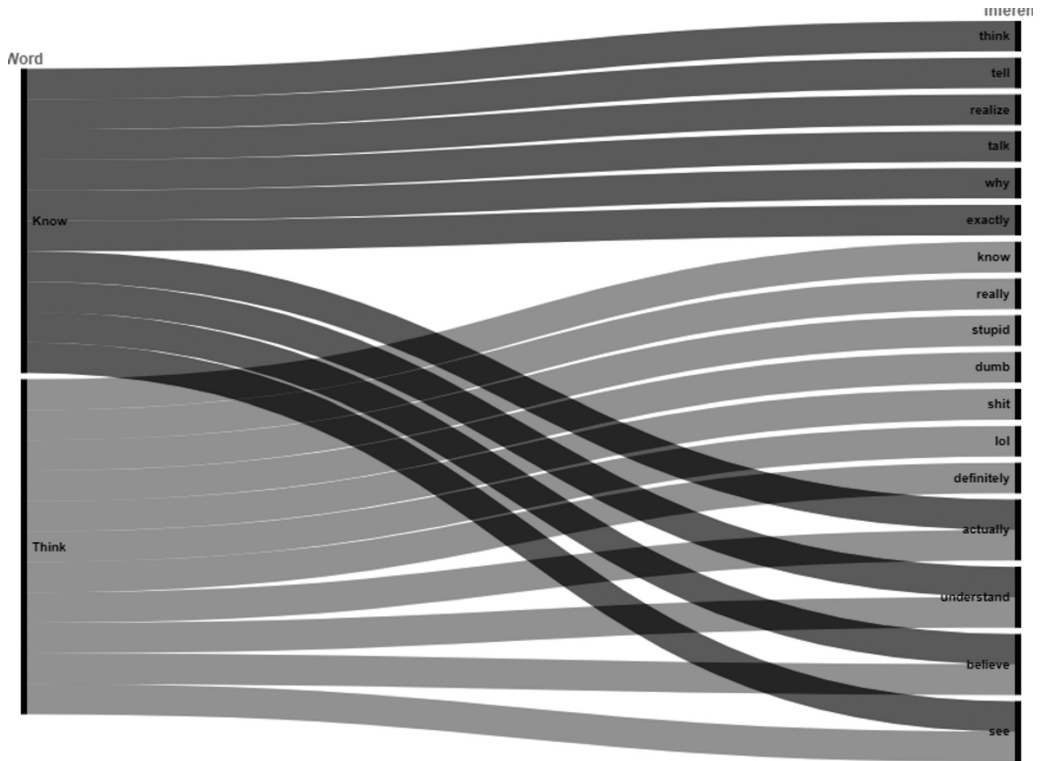
Figure 23.1 Semantic preference visualised in our word embedding model

argumentation but rather with those of positioning. This positioning occurs either by explicit means, with a clear stance of the author of the tweet, or by implicit means, mainly by positioning oneself in a contradictory position to the interlocutor. Nonetheless, the contradiction at the lexical level results in a much deeper division at the semantic level, where there is not a simple negation of the thesis of the other but a contrary positioning. In this case, the contradiction expressed by the negation is instrumental in actually presenting the values of only one of the two parties involved as evident.

In fact, it is precisely on the polarity of "not" and "no" that epistemic modalities are articulated on Twitter, whereas recourse to modal verbs – capable of expressing different nuances with respect to the construction of the discursive space – are comparatively very limited. Twitter positioning is prevalently and explicitly concerned with emphasising contrasting positions. The presentation of this contrast, however, is nothing but a means of reinforcing one's own positioning, with a systematic contraction of the dialogic space actually leading to the negation of dialogue.

This is a typical feature of echo chambers, where polarisation is expressed precisely in this dual nature of ideological reinforcement through the denial of the other. This denial is perhaps something more than a simple divergence of opinion, probably because of the nature of the ideological conflict. As shown previously, the articulation of epistemic modalities onto negation produces, at a lexical level, a contradiction. However, this negation is nothing but a way to present a reinforced standpoint.

Marina Bondi and Leonardo Sanna

The simple denial of other positions is actually a type of polarisation that we might describe as weak, as it subordinates one's own point of view to what is being denied (Lakoff 2004). However, when denial is used, as in our case, with the specific intent of denying the epistemic validity of others in order to reinforce one's own, we are faced with something different. In this case we are in front of a real relation of contrariety, which is then expressed through an articulation of semantic frames that goes beyond the ordinary axiological level, that is, the investment of values on the ideological level, what the ideology in question considers as positive or negative. In the case of echo chambers, we have something slightly different, which may be considered stronger in fostering the polarisation of the debate. Simply put, the main difference between a disagreement and the polarisation of positions in an echo chamber is the robustness with which conflicting axiological planes are contrasted. This robustness might be explored and, partially, measured looking at the key markers of dialogic contraction.

Our work has thus confirmed that individual markers of dialogic construction, even when clearly pointing to denial, are potential markers but not an automatic measure of dialogic contraction. It is always necessary to look at least at lexico-grammatical structures to confirm that there is indeed denial and to look at the wider context to understand the pragmatic and argumentative function of the utterance. Focusing on Twitter as a networking site has also shown that denial and rejection of the opponent's discourse are a key element of Twitter discourse. This of course in itself points to the notion of echo chamber, but not directly to that of disinformation. It is our contention, however, that this kind of discourse paves the way for disinformation, by privileging bare rejection of counterclaims, with no actual argument brought about to support one's own claim or the refutation of the counterclaim.

References

Albright, J. (2017) 'Welcome to the era of fake news', *Media and Communication*, 5(2), pp. 87–89.

Asif, M., Zhiyong, D., Iram, A. and Nisar, M. (2021) 'Linguistic analysis of neologism related to coronavirus (COVID-19)', *Social Sciences & Humanities Open*, 4(1), p. 100201.

Baker, P. (2006) *Using Corpora in Discourse Analysis*, London: Continuum.

Baker, P., Gabrielatos, C. and McEnery, T. (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*, Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511920103

Bartlett, J., Reffin, J., Rumball, N. and Williamson, S. (2014) 'Anti-social media', *Demos*, pp. 1–51.

Bednarek, M. (2008) *Emotion Talk across Corpora*, Basingstoke: Palgrave Macmillan.

Bolter, J. D. and Grusin, R. (2000) *Remediation: Understanding New Media*, Cambridge, MA: MIT Press.

Bondi, M. and Sanna, L. (2022) 'Exploring the echo chamber concept. A linguistic perspective', in M. Demata, V. Zorzi and A. Zottola (eds.) *Conspiracy Theory Discourses*, Amsterdam: Benjamins, pp. 143–167.

Boukala, S. and Dimitrakopoulou, D. (2018) 'Absurdity and the "blame game" within the Schengen area: Analyzing Greek (social) media discourses on the refugee crisis', *Journal of Immigrant & Refugee Studies*, 16(1–2), pp. 179–197.

Chaudhry, I. (2015) '# Hashtagging hate: Using Twitter to track racism online', *First Monday*, 20(2).

Chen, E., Lerman, K. and Ferrara, E. (2020) 'Tracking social media discourse about the Covid-19 pandemic: Development of a public coronavirus twitter data set', *JMIR Public Health and Surveillance*, 6(2), p. e19273.

Conroy, N. K., Rubin, V. L. and Chen, Y. (2015) 'Automatic deception detection: Methods for finding fake news', *Proceedings of the Association for Information Science and Technology*, 52(1), pp. 1–4.

Davies, M. (2016) *Corpus of News on the Web (NOW)*, Available online at www.english-corpora.org/now/

Davies, M. (2019) *The Coronavirus Corpus*, Available online at www.english-corpora.org/corona/

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G. and Quattrociocchi, W. (2016a) 'The spreading of misinformation online', *Proceedings of the National Academy of Sciences*, 113(3), pp. 554–559.

Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G. and Quattrociocchi, W. (2016b) 'Echo chambers: Emotional contagion and group polarization on facebook', *Scientific Reports*, 6(1), pp. 1–12.

Di Marco, N., Cinelli, M. and Quattrociocchi, W. (2021) 'Infodemics on Youtube: Reliability of content and echo chambers on Covid-19', *arXiv preprint arXiv:2106.08684*.

Eco, U. (1968) *La struttura assente [The Absent Structure]*, Milan: Bompiani.

Eco, U. (1979) *The Role of the Reader*, Bloomington: Indiana University Press.

Eysenbach, G. (2020) 'How to fight an infodemic: The four pillars of infodemic management', *Journal of Medical Internet Research*, 22(6), p. e21820.

Farkas, J., Schou, J. and Neumayer, C. (2018) 'Platformed antagonism: Racist discourses on fake Muslim Facebook pages', *Critical Discourse Studies*, 15(5), pp. 463–480.

Fillmore, C. J. (1976) 'Frame semantics and the nature of language', *Annals of the New York Academy of Science*s, 280(1), pp. 20–32.

Floridi, L. (2009) 'The information society and its philosophy: Introduction to the special issue on "the Philosophy of Information, its Nature, and future developments"', *The Information Society*, 25(3), pp. 153–158.

Frank, R. (2015) 'Caveat lector: Fake news as folklore', *Journal of American Folklore*, 128(509), pp. 315–332.

Fuoli, M. (2012) 'Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports', *Discourse & Communication*, 6(1), pp. 55–81.

Fuoli, M. (2018) 'A stepwise method for annotating appraisal', *Functions of Language*, 25(2), pp. 229–258.

Greimas, A. J. (1989) 'The veridiction contract', *New Literary History*, 20(3), pp. 651–660.

Groom, N. (2010). 'Closed-class keywords and corpus-driven discourse analysis', in *Keyness in Texts*, eds. M. Bondi and M. Scott, Amsterdam: Benjamins, pp. 59–78.

Himelboim, I., McCreery, S. and Smith, M. (2013) 'Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter', *Journal of Computer-Mediated Communication*, 18(2), pp. 40–60.

Hong, S. and Kim, S. H. (2016) 'Political Polarization on Twitter: Implications for the use of social media in digital governments', *Government Information Quarterly*, 33(4), pp. 777–782.

Hunston, S. (2002) *Corpora in Applied Linguistics*, Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139524773.

Hunston, S. (2004) 'Counting the uncountable: Problems of identifying evaluation in a text and in a corpus', *Corpora and Discourse*, 9, pp. 157–188.

Hunston, S. (2007) 'Semantic prosody revisited', *International Journal of Corpus Linguistics*, 12(2), pp. 249–268.

Lakoff, G. (2004) *Don't Think of an Elephant! Know Your Values and Frame the Debate*, White River Junction, VT: Chelsea Green Publishing.

Lakoff, G. (2017, January 13) 'A taxonomy of Trump tweets [On the Media]', *WYNC*, Available online at www.wnyc.org/story/taxonomy-trump-tweets/

Marchi, R. (2012) 'With Facebook, blogs, and fake news, teens reject journalistic "objectivity"', *Journal of Communication Inquiry*, 36(3), pp. 246–262.

Martin, J. R. and White, P. R. (2005) *The Language of Evaluation*, London: Palgrave Macmillan.

McIntyre, L. (2018) *Post-truth*, Cambridge, MA: MIT Press.

Melchior, C. and Oliveira, M. (2022) 'Health-related fake news on social media platforms: A systematic literature review', *New Media & Society*, 24(6), pp. 1500–1522.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013b) 'Efficient estimation of word representations in vector space', *arXiv preprint, arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. and Dean, J. (2013a) 'Distributed representations of words and phrases and their compositionality', *Advances in Neural Information Processing Systems*, 26, pp. 3111–3119.

Nguyen, C. T. (2020) 'Echo chambers and epistemic bubbles', *Episteme*, 17(2), pp. 141–161.

Oakes, M. P. (1998) *Statistics for Corpus Linguistics*, Edinburgh: Edinburgh University Press.

Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse*, Amsterdam: Benjamins.

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A. and Chakraborty, T. (2021) 'Fighting an. infodemic: Covid-19 fake news dataset', in *International Workshop on Combating on Line Hostile Posts in Regional Languages during Emergency Situation*, Cham: Springer, pp. 21–29.

Pérez-Rosas, V., Kleinberg, B., Lefevre, A. and Mihalcea, R. (2017) 'Automatic detection of fake news', *arXiv preprint arXiv:1708.07104*.

Petropoulos, F. and Makridakis, S. (2020) 'Forecasting the novel coronavirus COVID-19', *PLoS ONE*, 15(3), p. e0231236.

Rogers, R. (2018) 'Otherwise engaged: Social media from vanity metrics to critical analytics', *International Journal of Communication*, 12, pp. 450–472.

Ross, A. S. and Rivers, D. J. (2018) 'Discursive deflection: Accusation of "fake news" and the spread of mis- and disinformation in the Tweets of President Trump', *Social Media & Society*, 4(2), pp. 1–12.

Russell, B. (1940) *An Inquiry into Meaning and Truth*, London: George Allen and Unwin LTD.

Saez-Trumper, D. (2014), 'September. Fake tweet buster: A webtool to identify users promoting fake news on Twitter', in *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, New York: ACM Digital Library, pp. 316–317.

Sanna, L. and Compagno, D. (2020) 'Implementing eco's model reader with word embeddings. An experiment on Facebook ideological bots', in *JADT 2020: 15th International Conference on Statistical Analysis of Textual Data*. Available online at https://hal.science/hal-03144105.

Scott, M. (2020) *Word Smith Tools* (Version 8.0), Stroud: Lexical Analysis Software.

Scott, M. and Tribble, C. (2006) *Textual Patterns*, Amsterdam: Benjamins.

Shu, K. and Liu, H. (2019) 'Detecting fake news on social media', *Synthesis Lectures on Data Mining and Knowledge Discovery*, 11(3), pp. 1–129.

Sinclair, J. M. (2004) *Trust the Text: Language, Corpus and Discourse*, London: Routledge.

Sunstein, C. (2007) *Republic.com. 2.0*, Princeton, NJ: Princeton University Press.

Vargo, C. J., Guo, L. and Amazeen, M. A. (2018) 'The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016', *New Media & Society*, 20(5), pp. 2028–2049.

Wodak, R. and Boukala, S. (2015) '(Supra)national identity and language: Rethinking national and European migration policies and the linguistic integration of migrants', *Annual Review of Applied Linguistics*, 35, pp. 253–273.

Wurman, R. S. (1989) *Information Anxiety*, New York: Doubleday.

Zappavigna, M. (2011) 'Ambient affiliation: A linguistic perspective on Twitter', *New Media & Society*, 13(5), pp. 788–806.

Zhou, X., Cao, J., Jin, Z., Xie, F., Su, Y., Chu, D., Cao, X. and Zhang, J. (2015) 'May. Real-time news certification system on sina weibo', in *Proceedings of the 24th International Conference on World Wide Web*, New York: ACM Digital Library, pp. 983–988.

Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Havlin, S. and Quattrociocchi, W. (2017) 'Debunking in a world of tribes', *PLoS ONE*, 12(7), p. e0181821.