



**HAL**  
open science

# Socio-semantic configuration of an online conversation space

Camille Roth, Iina Hellsten

► **To cite this version:**

Camille Roth, Iina Hellsten. Socio-semantic configuration of an online conversation space. Social Networks, In press, 10.1016/j.socnet.2022.06.007 . halshs-03835429

**HAL Id: halshs-03835429**

**<https://shs.hal.science/halshs-03835429v1>**

Submitted on 16 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

# Socio-semantic configuration of an online conversation space

## The case of Twitter users discussing the #IPCC reports

Camille Roth<sup>a,b,\*</sup>, Iina Hellsten<sup>c</sup>

<sup>a</sup>CNRS, Computational Social Science team, Centre Marc Bloch (CNRS/HU), Friedrichstrasse 191, D-10117 Berlin

<sup>b</sup>CAMS, Centre d'Analyse et de Mathématique Sociales (CNRS/EHESS), 54 bd Raspail, F-75006 Paris

<sup>c</sup>University of Amsterdam, ASCoR, Nieuwe Achtergracht 166, NL-1018 WV Amsterdam

---

### Abstract

In public debates, climate change communication tends to polarize into communities for and against the scientific basis of global warming. We analyze mention networks on Twitter around the publication of the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) Working Group 2 and 3 reports that were published in March-April 2014. Building upon earlier research into climate skepticism and polarization of climate change debate, we focus on the relative prominence of different types of Twitter user accounts, in terms of engagement with other users, and their alignments toward the scientific basis of climate change. We distinguish a “heart” actively discussing IPCC from a “shadow”, which more anecdotally mentions IPCC and is likely to correspond to the remainder of a public space minimally interested in IPCC-related reports. We develop an original network analysis framework that enables us to analyze and deconstruct the inner structure of this heart’s strongly intertwined engagement dynamics. Interesting observations relate to the position of critical users, who are in the minority, but are in relative terms most engaged with and most engaging with other users in this arena, while the media, casual users and governmental agencies occupy relatively less prominent positions. We further qualify the various structural positions by demonstrating that they correspond to different types of vocabulary specific to user types and positions. This socio-semantic approach may be generally helpful to disentangle semantic and structural polarization in online conversation spaces where opposing poles precisely appear to be mixing.

*Keywords:* socio-semantic networks, socio-semantic cohesiveness, climate change, twitter, polarization, position analysis, socio-semantics

---

### 1. Introduction

The last decade has witnessed a massive growth in the number of studies devoted to the informational and interactional structure of web communities, and their increasing contribution to the shape of the epistemic landscape of our connected societies. Debates regarding the specificities of the digital public space, especially with respect to its offline counterpart, have been striving: be it between scholars arguing whether the Internet is a haven for heterodox positions absent from the offline public space (Van Aelst and Walgrave, 2002) or academics discussing whether it reproduces classical power/authority positions and relations (Hindman et al., 2003; Gerhards and Schäfer, 2010) and, more broadly, to what extent the digitization of political discussion may facilitate or hamper democratic deliberation (Dahlgren, 2005; Shirky, 2011).

In the meantime, the issue of a possible balkanization of digital spaces (as early as Van Alstyne and Brynjolfsson, 1996) or the existence of so-called echo chambers (Sunstein, 2007) has been taking a tremendous momentum over the course of the last two decades. Assessing the importance and the effects of

polarization in this context increasingly relies on quantitative methods, borrowing either to social network analysis or automated content analysis, or both (Barberá et al., 2015; Conover et al., 2011; Bakshy et al., 2015). A typical approach consists in describing socio-semantic clusters groups of actors i.e., cohesive groups of interacting individuals who share similar properties (Adamic and Glance, 2005; Cattuto et al., 2009; Roth and Cointet, 2010; Lietz et al., 2014) — assuredly a robust sociological result replicated in many empirical fields way beyond web communities. Focusing on socio-semantic cohesiveness puts forward the configuration of groups with respect to one another, and exchanges between them. In other words, this literature principally sheds light on the relative arrangement of socio-semantic clusters: typically, how reds do, or do not, cite blues, talk with them, and about what (Etling et al., 2009; Himelboim et al., 2013; Hoffmann, 2014; Kaiser and Puschmann, 2017; Vicario et al., 2017; Gaumont et al., 2018).

However, socio-semantic cohesiveness on political issues appears to be rather weak when it comes to conversational networks as opposed to what may be denoted as affiliation networks. In the particular case of Twitter, for instance, affiliation links correspond to following other users or retweeting them without commenting, and they exhibit strong socio-semantic cohesiveness (Himelboim et al., 2013; Garimella et al., 2016a; Tyagi et al., 2020). By contrast, a significant amount of cross-

---

\*Corresponding author

Email addresses: roth@cmb.hu-berlin.de (Camille Roth),

I.R.Hellsten@uva.nl (Iina Hellsten)

cutting discussion between actors from opposing sides is observed with conversation links i.e., mentions, replies, quotes (Conover et al., 2011; Maireder and Ausserhofer, 2013; Pearce et al., 2014; Barberá et al., 2015; Williams et al., 2015; ?). This phenomenon is likely to undermine the relevance of focusing on cohesive groups per se when studying conversations rather than affiliations. In other words, the socio-semantic mixing of conversations blurs the structural boundaries between opposing poles of a debate — simply put, because conversations obey to a much less marked separation between the blues and reds, we contend that it may matter more to understand the differentiated roles and relative dominance of actors holding certain opinions within a rather mixed conversation space. This point appears to be particularly key for issues where online public spaces may act as a locus for the expression of ideas that remain under-represented in the offline public space (a feature identified a couple of decades ago already by Ayres, 1999). In this case, understanding which actors and discourses dominate online spaces is of particular importance, notably in comparison with the offline world.

Our contribution to the structural analysis of polarization thus departs from approaches centered on clusters. By contrast, we primarily aim at studying actor positions by acknowledging cohesiveness rather than describing it per se. Put differently, we aim at deconstructing the finer socio-semantic structure in the potential hairball formed by a conversation network interweaving opposing sides. Our approach ultimately aims at describing the dominance patterns of an online conversation space based on Twitter in both social and semantic terms. By “social” we essentially mean “structural” i.e., centrality in the underlying social network; by “semantic” we refer to the vocabulary i.e., the importance of some terms in the corresponding publications. We focus on “engagement”, which refers to two things: first, quite trivially, engaging with the issue i.e., the act of evoking the issue; and second, engaging with other users on the issue i.e., the act of mentioning another user while evoking the issue. We aim to understand which actors are dominant in terms of patterns of engagement with other users, and which poles they come from, in a context that features a lot of cross-cutting discussion and thus a structural intertwinement of opposing poles.

We empirically chose to focus on climate change, a typically polarized debate (Painter, 2011; Painter and Ashe, 2012) which now steadily takes place on social media platforms. In this respect, Twitter is no exception (Jang and Hart, 2015; Holmberg and Hellsten, 2016; Newman, 2017; Veltri and Atanasova, 2017; Vu et al., 2019) where conversations also appear to exhibit a significant amount of interaction between poles (Pearce et al., 2014). Besides, it constitutes an *in vivo* playground where both social networks and content may be observed within a given perimeter (for a recent review on online climate change debates, see Pearce et al., 2019). We carry out a socio-semantic positional analysis to characterize the alignments and utterances of users who occupy dominant and less dominant positions, in contrast to focusing principally on the most active or central actors (Newman, 2017; Wang et al., 2020) or on content dynamics (Jacques and Knox, 2016; Holmberg and Hellsten, 2016; Veltri

and Atanasova, 2017; Hellsten et al., 2019). In practice, we focus on a “heart” of Twitter users actively engaged in a specific climate change debate, related to the publication of an IPCC report, and aim at further describing its configuration in interactional and informational terms: which categories of users are the most or least engaged with within that heart (being mentioned by others), the most or least engaging (mentioning others), and which topics are being addressed by whom in which position. Formally, we proceed in three main steps. First, we define a perimeter of users and publications related to a specific climate change debate on Twitter. Second, we use a simple structural model to describe the possible existence of remarkable positions in the thus-defined space. Finally, we globally characterize the type of actors and vocabularies that are most dominant in these positions.

## 2. Empirical context

### 2.1. Structure of online debates on climate change

In a review article on online climate change communication, Schäfer (2012) discusses the relative prominence of different actors: scientists are less engaged while NGOs, in particular, are more prominent. While relatively little was known on the presence of corporations, politicians and small research institutes, a few studies showed think-tanks to be very active online. The more recent review paper by Fownes et al. (2018) details the variety of practices of NGO, political, or celebrity accounts in the climate change debate on Twitter. Nonetheless, in a previous study (Newman, 2017), the most engaged users were found to be non-elite users, followed by news media, advocacy organizations and governmental organizations while journalists and scientists were slightly less active. Wang et al. (2020) recently nuanced these findings, in the context of the debate on the UN COP21 conference, by further showing that NGO accounts were the most central in terms of betweenness, even if accounts of individuals still had the highest degree centrality. In a similar vein, Vu et al. (2019) found relatively low levels of interactivity and connectivity across NGOs who are members of Climate Action Network International.

Notwithstanding, the literature consistently reports the existence of online polarization on this issue (Jang and Hart, 2015) which is reflected in the socio-semantic cohesiveness of the underlying networks: for instance, blog citation networks exhibit significant separation between actors holding opposite opinions (Elgesem et al., 2015). However, the situation appears to be more mixed in conversation networks (Williams et al., 2015). In Twitter in particular, while supporters and opponents of the scientific basis of climate change also forward (i.e., retweet) content posted by other, often like-minded, users (forming loose clusters of those critical, those supporting the scientific evidence, and neutral towards the scientific basis of climate change), a sizable share of them nonetheless appears to converse with users holding opposing views (Pearce et al., 2014; Tyagi et al., 2020).

Whatever the level of discussion between groups, it does not seem to strongly affect the semantics mobilized by users.

Indeed, polarized groups appear to use distinct words to discuss the issue: while Twitter users who support the scientific basis of climate change use “climate change” indeed, those skeptical of it rather use the term “global warming”, which appears to be further associated with hoax terminology (Jang and Hart, 2015; Shi et al., 2020). In addition, polarized groups label each other with terms such as “alarmists” and “warmists” or “deniers” (Howarth and Sharman, 2017). This hints at the interesting possibility that structural mixing (cross-cutting interactions) does not strongly affect semantic cohesiveness (polarized content production).

## 2.2. Focusing on the discussion of IPCC reports on Twitter

Reports of the Inter-governmental Panel on Climate Change (IPCC) Working Groups usually attract significant interest in offline and online public spaces (Hulme, 2009; O’Neill et al., 2015) as well as in the scientific realm, especially geophysical sciences and Western scientific communities, even though this skewness tends to have diminished recently (Vasileiadou et al., 2011). The publication of IPCC reports occurs roughly every six years (+/- one year) as “Assessment Reports”: the First Assessment Report (AR1) was published in 1990, AR2 in 1996, AR3 in 2001, AR4 in 2007 and AR5 in 2014. At the time of submission, AR6 was expected to be completed in 2022. In the Fifth Assessment Report (AR5), the IPCC Working Group 1 (WG1) report, *The Physical Science Basis*, detailed out the natural science basis for climate change, and was published in September 2013. The WG2 report on *Climate Change Impacts, Adaptation and Vulnerability* was published in March 2014, followed by the WG3 report on *Climate Change Mitigation* in April 2014.

We focus on the surge of Twitter discussions surrounding the publication of the latest AR5 IPCC reports in March-April 2014 (WG2 & WG3), not only because they are the most recent but also because this makes the definition of engagement much less ambiguous. To understand why, let us briefly evoke a few key technical aspects of Twitter. From its inception, the platform has allowed users to post short messages and share links (URLs). This quickly led to the emergence of original uses, whereby users could address another user by *mentioning* their username (preceded by an “at” sign, @), tag messages using keywords (or *hashtags*, preceded by a “hash” sign, #), and *retweet* i.e., retransmit another user’s post (initially by prefixing it with “RT”). These primordial affordances make Twitter a networked public space (e.g., Himelboim et al., 2017) where issue-specific publics form around hashtags (Bruns and Burgess, 2011) and directed relationships between users emerge around mentions and retweets (boyd et al., 2010). Further functions have been introduced through time, such as threads, replies and quotes (Garimella et al., 2016b). In this regard, focusing on IPCC reports in 2014 has several benefits with respect to defining a clear-cut perimeter and the engagement of its users. Indeed, over this time period, it is enough to target tweets that contain “IPCC” as users who discuss IPCC reports are likely to use this term, while it is also rarely related to something else.<sup>1</sup>

<sup>1</sup>We thus presume that there are few false negatives, and we observe very

Furthermore, at that time, mentions were one of the most obvious ways of engaging in a conversation with other users on Twitter — all the more that mentions generally span across opposing sides, unlike retweets. The inter-user social network that we construct in this paper is thus simply based on mentions: a link from user A to user B exists if A mentioned B using the “@” sign, it may or may not be reciprocated.

## 2.3. Data

*English-speaking tweets discussing “IPCC”.* We collected all tweets mentioning “IPCC” or using the “#IPCC” hashtag around the publication of the IPCC, AR5, WG2 and WG3 reports, respectively on March 31 and April 14, 2014. More precisely, tweets were collected using the Twitter API with Webometric Analyst (Thelwall, 2009) between March 11 and April 23, 2014. The period covers several weeks before and after the actual publication of IPCC reports for it is a highly anticipated event and generates a large amount of tweets prior to, during and after the event itself. This resulted in total of 259,610 tweets mentioning “IPCC” in any language. We proceed in several steps to further refine the case study perimeter. We controlled for the language used in the tweets, and in our sample 87% were in English. Previous research on Twitter additionally suggests that users generally address distinct audiences depending on the language they use (Kim et al., 2014). In this respect, we can reasonably consider that publication flows in distinct languages correspond to distinct discussion spaces, notwithstanding the fact that the small volume of non-English flows is likely to have a quite marginal effect on results.

In a nutshell, at least for languages where IPCC is indeed used as an acronym, we contend that most of the debate featuring “IPCC” occurs in English.<sup>2</sup> We eventually focus on the Twitter sphere made of English tweets, assuming that multilingual users address an English-speaking sphere when they write in English. We dismiss tweets posted in other languages and thus the other language-specific public spheres, whose respective sizes are comparably small in any case. The corpus is thus reduced to 212,429 tweets i.e., conserving slightly more than 80% of the initial data.

few false positives: manual inspection of 100 random tweets of our dataset reveals that all deal with the IPCC reports, except two tweets that use “IPCC” to refer to the British Independent Police Complaints Commission.

<sup>2</sup>IPCC is indeed an international keyword that may be used by speakers of distinct languages. Examining first the issue of multilinguism and using a language detection algorithm, we observed that 39,638 users out of 40,521 (i.e., 97.8%) only ever use one language for all their tweets on IPCC. Among these monolingual flows, 34,471 are in English i.e., about 87% of all users. We then find, in decreasing order, 6.7% of flows written only in Spanish, 2.7% for German, and even smaller amounts for Italian, Dutch, and French (possibly explained, in this last case, by the fact that IPCC is translated and usually referred to as GIEC in French). Among the 2,642 Spanish-speaking users, only 132 have published 5 tweets or more. For German-speaking users (the second group in terms of size), it goes down to 62; and so on for smaller language groups. This partition hints at the existence of separate linguistic spaces of discussion on Twitter, the bulk of it being dominated by English. The examination of multilingual user flows further substantiates this intuition. While this only concerns about 2% of users, in most cases, English appears alongside another language (either as a minority, roughly fifty-fifty, or majority language — these three typical situations being represented in relatively equal proportions).

*Defining a “heart” of users most engaged in the debate.* We further aim at delineating the perimeter of users who are the most engaged in IPCC discussions. To this end, we consider that the number of tweets actually written by a given user in this corpus is a good measure of engagement in the debate. Again, we do not take retweets into account as we assume that they correspond to a more casual (re)posting behavior which is more related to information relaying than direct expression and original content creation.<sup>3</sup> This reduces the dataset to 89,949 tweets, after excluding 122,480 retweets (which represent 57.7% of the English-speaking tweets). To avoid a partially circular argument, we also do not base the threshold on the number of mentions a user receives, since this precisely corresponds to the dependent variables in which our position analysis will be grounded, given that we will assess the relative levels of engagement of users with other users within a perimeter of users engaged in the debate (as in e.g., Hellsten et al., 2019). Simply put, engagement in the debate in terms of the number of original posts is distinct from engagement levels with other users in terms of mentions sent or received. As we shall see, we find many engaged users who emit or receive few or no mentions (roughly 80% of heart users mention or are mentioned by less than 10 other users from both the heart and the shadow).

Engagement with IPCC is typically heterogeneous, whereby many users post few tweets and few post many. This leads us to distinguish an IPCC “heart” audience from an IPCC “shadow” which more anecdotally evokes IPCC and is likely to correspond to the remainder of a public space more casually interested in IPCC-related reports. More precisely, we fix the threshold for the heart at 15 IPCC tweets over the whole observation period, which corresponds to the top 2% most engaged users among the 35k distinct authors of English-speaking IPCC tweets. This yields 629 “heart” nodes which produced about 27% of the total amount of tweets (24,396) i.e., a quarter of the dataset. While our study focuses on the heart, the wider peripheral audience defined as the shadow will still be used to inform us on the relative status of heart users.

*Manual identification of alignments.* We further aim at describing the *alignment* of the 629 heart user accounts towards the existence of global warming. We categorize them manually. We assume that they may be explicitly supportive of the scientific basis of the climate change or, to the contrary, affiliated with skeptical opinions on the issue; or finally that they may also express no specific or discernible opinion at all. At the account level, for heart users, it is generally possible to distinguish those who mainly express *supportive*, *critical* or *uncommitted* views, following the same categorization approach used

<sup>3</sup>Doing otherwise would additionally critically affect the structural analysis of mention networks: taking into account a retweet by A of B mentioning C would amount to considering that A mentioned C, whereas the original mention only goes from B to C (in the context of scientific paper networks, this would be equivalent to considering that a scientist who cites a given paper inherits, in turn, all the citation from that given paper). In other words, with respect to the extraction of mentions from tweets, we deem retweets to be second-order publications and we find it problematic to treat them on a par with original tweets.

	<i>Alignment</i>	<i>Number</i>	<i>%</i>
(C)	<b>Critical</b>	60	9.5%
(S)	<b>Supportive</b>	229	36.4%
(U)	<b>Uncommitted</b>	340	54.1%

Table 1: Breakdown of “heart” user alignments.

by Pearce et al. (2014) or Williams et al. (2015). In practice, hand coding<sup>4</sup> has been performed by looking either at the self-description available on user profiles, or at the website their profile links to, where heart users often express an explicit stance towards the wider issue of climate change. When they do not, we used common knowledge on the actors: for instance, Greenpeace as an environmentally-oriented NGO that supports the scientific basis of climate change or, for an account called “no-consensus”, the fact that its declared homepage points to the website of an individual journalist who explicitly casts doubt on the supportive information on climate change. In all other cases, we tagged users as “uncommitted”. The relevance of this hand coding could be confirmed a posteriori through casual examination of IPCC-related tweets for about XX users of each category; for instance, here is an example of a tweet by a user tagged as critical to the scientific basis of human-induced climate change:

IPCC Insider Rejects Global-Warming Report - National Review Online (blog)

or, conversely, a tweet by a user coded as supportive of the scientific basis of human-induced climate change:

#IPCC: Climate change is everywhere. That means we need to take action. Now.

Users whose profile are not clearly falling into one of these two categories were tagged as “*uncommitted*” – we observe that their IPCC-related tweets mainly provide flat information related to the IPCC report publication:

Climate scientists meet in Japan for IPCC’s WG2 report - Top Asia-Pacific News 25.03.14

See Table 1 for the respective breakdowns of user types and alignments. As it turns out, critical users constitute a bit less than 10% of the heart.

### 3. Configurations and positions

*Structural model.* We use and extend a structural model that had been introduced by Cardon et al. (2011) in the analysis of blogspace. This model is based on a so-called “A1-D4” matrix and allows for the description of position classes defined with respect to the relative attention that nodes give and receive from

<sup>4</sup>We favored this approach over an automated method for we believe it achieves a higher precision, knowing that the coding of several hundreds of accounts by hand remained tractable — by contrast, using, say, sentiment analysis, would have been likely to classify user attitudes (expressing mostly negative or positive opinions) without necessarily telling us which side of the debate they were on.

inside and outside of a target group of actors (a specific group of interest). The group perimeter, that we call “heart”, can be defined semantically: for instance, users interested in a given topic vs. users who are not. Here, within the IPCC debate, we base the notion of heart on a certain engagement level: highest engagement in the debate vs. casual engagement.

Irrespective of this choice, the model is based on a *primary dichotomy* distinguishing the heart from its outside, which we call “shadow”. This configures a dichotomy between actors who are engaged, concerned, involved with a given topic vs. less engaged, less concerned, less involved ones. It further relies on a *secondary double dichotomy* differentiating actors of the heart who (a) are the most engaged with by, or are engaging the most with, actors who belong (b) to the heart or to its shadow. This makes it possible to jointly characterize engagement activity (mentioning) and engagement authority (being mentioned) with respect to actors who are engaged or less engaged with the topic. This model is most relevant to describe the internal structure of a target group (the heart) in terms of distinct attentional positions with respect to a broader environment (both the heart and the shadow). It sheds light on structural positions in a manner distinct from cluster analysis, especially when the group under study exhibits some amount of structural cohesiveness. We extend the application domain of the model by (i) adopting a heart/shadow boundary that is defined inside a semantically-defined target group, rather than between them, (ii) performing a two-level analysis of the over-/under-representations of the structural positions revealed by the model inside that focus group, and (iii) by further qualifying their semantics at both levels.

More concretely, heart actors may receive more or less mentions from actors of the heart or of the shadow, and address more or less mentions towards heart or shadow actors. This defines four types of links: (i) incoming links from the heart itself, (ii) incoming from the shadow, (iii) outgoing to the heart itself, (iv) outgoing to the shadow. This double dichotomy is illustrated on Fig. 1. In what follows, we denote by “*activity*” the fact that users establish links, and by “*authority*” the fact that they receive them.

For each of these link categories, we anticipate and confirm degree centrality to follow an heterogeneous distribution i.e., a significant number of nodes having low centralities and fewer nodes having higher ones. We further use this heterogeneity to distinguish the most central nodes from the other ones. To this end, we set a target threshold  $p < 0.5$  such that “top” nodes in a given link category are *at most* among the fraction  $p$  of the most central nodes in that category, and the “bottom” nodes are the rest. Nodes around the threshold can exhibit identical centralities, so that in practice we consider the largest value  $p' \leq p$  which does not require to arbitrarily split a class of nodes with identical centralities between the top and the bottom categories. Empirically, actual thresholds for each categories generally only slightly diverge from the target  $p$ . In this paper, we use a classical  $p = 0.2\%$  target threshold (20/80%), even though using  $p = 0.1$  or  $p = 0.3\%$  does not qualitatively change most of the results (see footnote 7).

Eventually, each node belongs exclusively either to the top

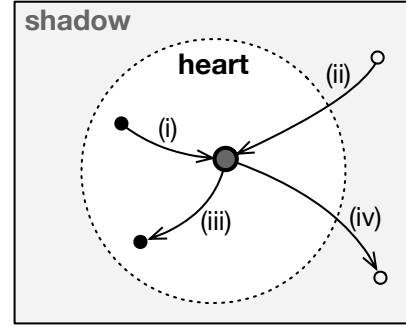


Figure 1: Illustration of the four types of incoming/outgoing links with respect to the heart/shadow dichotomy: (i) and (ii) are incoming links respectively from the heart and the shadow, while (iii) and (iv) are outgoing links resp. to the heart and the shadow.

or to the bottom for each link category. We denote this as *dominance* in a type of link. This also yields a model featuring 4x4 possible positions in a regular fashion, depending on whether a node is dominant in each of the four link categories. An illustration is provided on Fig. 2. For instance, the top-left slot (A1) is populated by nodes which are in the top category everywhere, while the bottom-right slot (D4) gathers nodes which belong to the bottom for all link categories. One of the first uses of this regular model is to provide a simple way to discuss which nodes enjoy which type of authority and are involved in which type of activity. We discuss the meaning of the various slots further below.

*Remarkable positions.* Besides, not all positions are equally likely, i.e. not all slots are equally populated. This is first due to the fact that “top” and “bottom” are not evenly splitted, by virtue of  $p$  being strictly smaller than 0.5. For instance, since the A1 slot gathers only nodes which are in the top 20% for each link categories, its a priori expected population is likely to be smaller than that of D4 which gathers nodes in the bottom 80% for all link categories. More generally, even if the chances of belonging to the top or bottom for each link category were uncorrelated, for a given slot and thus a given top/bottom combination, the chance of belonging to a slot would be equal to the product of the chances of belonging to the top ( $p$ ) or bottom ( $1 - p$ ) for each link category.<sup>5</sup> This yields the a priori matrix of probabilities pictured on Fig. 3.

Second, dominance in a link category is in practice generally correlated with dominance in other categories — internal and external authority and activity appear to be correlated with one another. The model thus makes it possible to assess whether some type of position could be remarkable in itself. Some positions are going to be over- or under-populated with respect to the baseline expectations for the uncorrelated case shown on

<sup>5</sup>Formally, we may describe the type of a node by, say, a tuple  $(t_i)_{i \in \{1, \dots, 4\}} \in \{0, 1\}^4$  where each value equals 1 or 0 depending on whether that node is in the top or bottom for link category (i), (ii), (iii) or (iv), respectively. For instance, the type of nodes in slot B2 is denoted by  $(0, 1, 0, 1)$ . In the absence of correlation, the probability of belonging to type  $t$  is thus  $\prod_{i \in \{1, \dots, 4\}} p^{t_i} (1 - p)^{(1 - t_i)}$ .

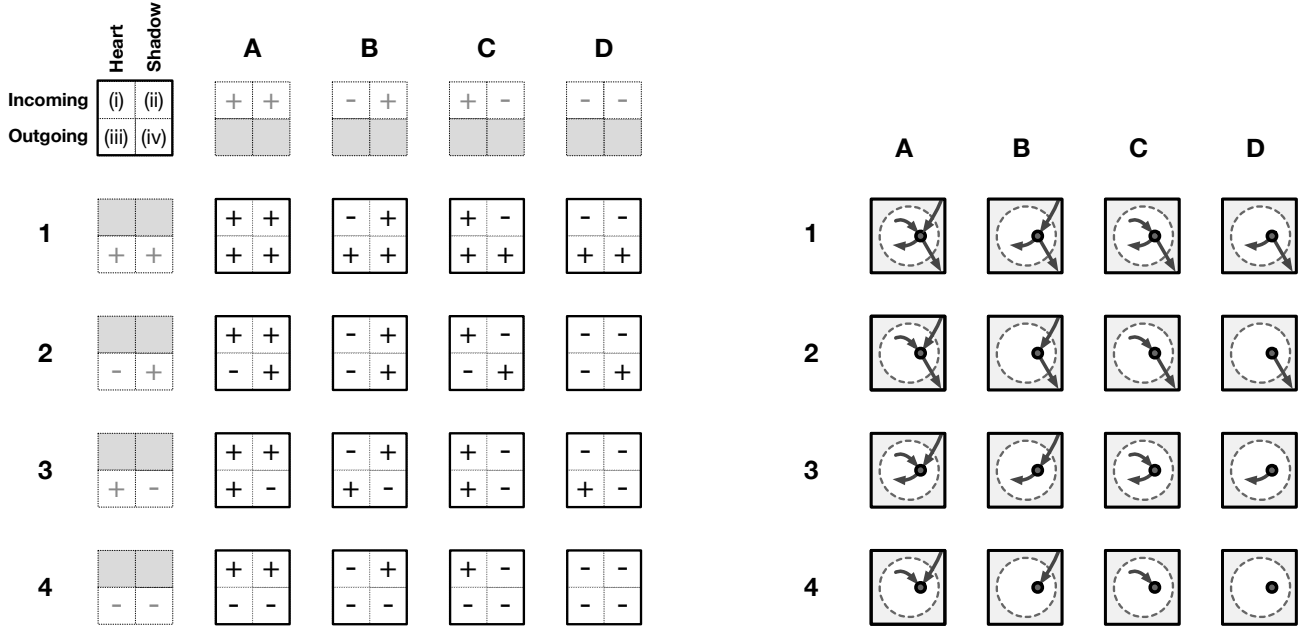


Figure 2: The “incoming/outgoing” and “heart/shadow” double dichotomy defines 4x4 possible positions, depending on whether a node is in the top (+) or bottom (-) of degree centralities in each of the four link category. It yields a regular model of combinations as shown on the *left* panel. A1, for instance, corresponds to nodes belonging to the top for all categories, while D4, on the contrary, denotes nodes in the bottom everywhere. An equivalent graphical illustration is provided on the *right* panel, where arrows are pictured only when a position gathers top nodes for the corresponding link category.

	A	B	C	D
1	$p^4$ 0.16%	$p^3(1-p)$ 0.64%	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%
2	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%
3	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%
4	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%	$p(1-p)^3$ 10.2%	$(1-p)^4$ 41%

Figure 3: Prior probabilities in the absence of correlations between the various link categories if considering exactly  $p\%$  for the top nodes of each category. In gray, illustration of expected population proportions for each slot when fixing  $p = 0.2$ .

Fig. 3. Discrepancies between actual and expected populations hint at further examining the corresponding slots to describe whether they indeed correspond to remarkable node properties (e.g., types of actors, their vocabularies, etc.).

*Positions and vocabulary.* We principally rely on these imbalances to discuss the socio-semantic configuration of the heart. To this end, we propose furthermore to study the vocabulary typical of slots and alignments. We proceed with a simple natural language processing protocol entirely aimed at exhibiting the terms that are both proper and characteristic of the various slots. We first isolate common words mentioned in posts.<sup>6</sup> This yields a large number of unique single word lemmas (9,205). We thus further need to isolate the most meaningful words i.e., typical of the respective slots or alignments. To this end, we introduce a simple scoring method inspired by the classical “tf.idf” weighting scheme (Salton and Buckley, 1988) to extract terms which are characteristic of the subcorpus corresponding to a certain category, say alignments or slots. We indeed assume that some terms will be significantly more present in the vocabulary of nodes of a certain category than in the rest of the corpus.

In all generality, we compute the *term frequency* of a term  $w$  in a category  $c$  by computing the proportion  $\text{tf}_c(w) \in [0, 1]$  of users of category  $c$  who use term  $w$ . We then calculate the relative frequency  $\text{rf}_c(w)$  of that term with respect to the other

<sup>6</sup>We use Natural Language Toolkit (NLTK) 3.2.4 for python, apply a part-of-speech (POS) processing to focus on nouns and verbs only, and eventually find lemmas using the WordNetLemmatizer.

categories,  $rf_c(w) = \frac{tf_c(w)}{\langle tf_{c'}(w) \rangle_{c' \in \mathcal{C}}}$  where  $\mathcal{C}$  is the set of categories (alignments, slots, or combinations of both). This ratio indicates how much a term is specific to a category. We typically consider its logarithm, which is a traditional approach to account for the wide range of orders of magnitude of such ratios. We then obtain an estimate of the importance or relevance of term  $w$  to category  $c$  by defining its score as its pure frequency in category  $c$  weighted by the logarithm of its relative frequency:

$$s_c(w) = tf_c(w) \cdot \log rf_c(w) \quad (1)$$

The detection of the vocabulary specific to some slots or groups will heavily rely on this score.

## 4. Results

We obtain the global heart structure as depicted in Fig. 4.<sup>7</sup> We immediately observe that slots belonging to column “A” (i.e. top cited nodes both from the heart and the shadow), to line “1” (i.e. top active nodes both towards the heart and the shadow) as well as slot “D4” are the only over-represented slots.

The other slots are under-represented, a finding consistent with the typical structure observed in cases corresponding specifically to online deliberative spaces when adopting a purely semantic understanding of the heart (Cardon et al., 2011). The central slots of the matrix (B2-C2-B3-C3) are almost non-existent (especially C3 which corresponds to a heart-centric pattern where activity and authority relate exclusively to the heart with no regard to the shadow). The same goes for the central slots of line “4” (i.e., users who are not dominant for both types of activity may only be dominant in terms of authority when it comes from both the heart and the shadow) and column “D” (i.e., users who are not dominant for both types of authority may only be dominant for both types of activity when it goes to both the heart and the shadow).

By contrast, the over-representation of line “1”, column “A” and slot D4 suggests remarkable configurations and combinations of authority and activity. We thus focus on these slots and further group and distinguish them depending on whether there is a strong combination of authority and activity (A1-A2-A3, B1 and C1), only authority (A4), only activity (D1), and noth-

ing at all (D4).<sup>8</sup> The first group is principally exemplified by A1 (which is most significantly over-represented) and corresponds to a situation where nodes are authoritative in the heart and shadow while being strongly active: Twitter accounts in slot A1 combine all four types of dominance, while the other slots combine three of the four, we denote this group with the term “**star**”. For instance, this slot comprises both prominent supportive accounts, such as Climate Reality, founded by Al Gore (@climatereality), prominent critical accounts (often of individual users), and uncommitted such as @RTCCnewswire (now @ClimateHome) that forwards information for and against human-induced climate change. The second group (A4) only exhibits authority dominance but not activity dominance, a situation which denotes prestige without necessarily engaging intensively in the conversation, we denote this group with the term “**famous**”. This group comprises supportive accounts that belong to, for instance, Greenpeace, while critical accounts consists of prominent accounts such as Watts-up-with-that, that explicitly promotes climate change denial and was founded by Anthony Watts. Neutral accounts in the A4 group are mainly news media, for instance The Guardian or WMONews, that belongs to the World Meteorological Organization. The third group (D1) is exactly the contrary of famous or star accounts: heart users who address the heart and the shadow without getting much attention in return, and we denote it with the term “**curious**”. Eventually, the fourth group (D4) combines the absence of all four types of dominance, and we denote it with the term “**silent**”; it is the vast majority of Twitter users in this case. The remaining slots (B2-C2-D2-B3-C3-D3-B4-C4) are under-represented and are grouped under the label “**absent**” (see also an illustration of these groupings on Fig. 6). B3 and C2, in particular, correspond to seemingly paradoxical configurations, where nodes are either strongly mentioned by the shadow only while addressing only the heart (B3) or, vice versa, mentioning the shadow only while being addressed by the heart only (C2).

*Positions and alignment.* We further examine, at the macro-level, the type of alignment predominating in each of these categories. More precisely, we first check whether some slots are disproportionately hosting nodes aligned with a critical (C), supportive (S) or uncommitted (U) discourse. This yields the three representativity matrices shown on the left of Fig. 5. Each matrix describes how much a certain type of alignment is found in each slot with respect to the expected proportion under an independence hypothesis between the slot and node alignment. For instance, while critical nodes represent 9.5% of the corpus, and should thus make for 1.6 nodes of the 17 nodes of slot A1, they actually constitute 29.4% of that slot (5 nodes) i.e. an over-representation ratio of a factor more than 3. On the whole, we see that critical nodes are rather over-represented in the star slots, supportive accounts make the bulk of the silent and relatively silent slots (in the bottom right corner), while uncommitted actors are rather more present in the bottom left quadrant

<sup>7</sup>Note that we also ran the model using  $p = 0.1$  and  $p = 0.3$  and obtain qualitatively similar results. The main matrix on Fig. 4 goes essentially in the same direction for both  $p = 0.1$  and  $p = 0.3$ , with respectively only one (B1) and two (B1 and C2) slots going in an opposite direction (from over to under-representation) and which are among the few slots where the empirical ratio is not statistically significantly different from the expected value. The matrices on Fig. 5 also behave similarly. The populations are smaller for these sub-matrices (even for  $p = 0.3$ ) and generally not large. We observe slightly more variation (from under to over representation or vice versa): excluding slots occupied by one or zero node (admittedly a very small number), we find that on average 2.2 slots per matrix behave differently than for  $p = 0.2$ , generally on weakly populated (central) slots. As a result the matrices on Fig. 6 are globally stable, except for the central “absent” group which features several very weakly populated slots and a mixed combination of dominant attitudes S/U/C (very much like it is for  $p = 0.2$ ).

<sup>8</sup>Note that our results hold should we opt for a different breakdown, for instance by distinguishing full stardom in A1 from strongly authoritative slots (A2-A3-A4), strongly active slots (B1-C1-D1), and neither active nor authoritative (D4).



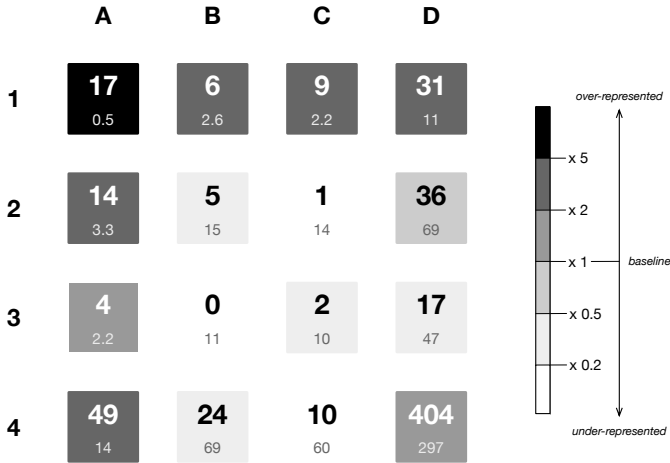


Figure 4: Representativity matrix. Each matrix cell is filled with the actual number (top and large figure) and expected number (small, bottom figure) of nodes falling in the corresponding slot. The ratio between the former and the latter describes the over- or under-representation of a slot, corresponding here to its background color: the darker, the more over-represented and thus remarkable. (Actual thresholds for each link category are respectively  $p_{(i)} = 0.17$ ,  $p_{(ii)} = 0.19$ ,  $p_{(iii)} = 0.14$  and  $p_{(iv)} = 0.19$ , corresponding to cut-off degree centralities of 4 (i), 7 (ii), 6 (iii) and 7 (iv). Extremely strong statistical significance of the divergence from the expected values for most cases ( $p < 0.01$ , multinomial proportion test), the only exception being the weakly populated slot B2, B3 [ $p < 0.05$ ] and A3 [ $p > 0.1$ ]).

(around the “famous” slot). To put it briefly, these uneven distributions indicate that the remarkable slots are also occupied by diverse actors. The right side of Fig. 5 summarizes this finding by showing the most over-represented alignment for each slot. This is all the more visible when considering the above-mentioned groupings, where critical alignment, while in small minority in the corpus, is much more present than it should be in the most dominant slots, whereas the uncommitted alignment makes much of the famous (and absent slots), and the supportive alignment rather tends to overpopulate the silent majority.

The combination of the structural positions (Figure 2) and the type of alignment reveals that each position consists of a set of different types of actors. In the “star” positions (A1 to A3 and B1 and C1), the main supportive actor types are individuals linked to environmental NGOs (such as Oxfam or ClimateReality), critical individuals (such as a member of Climate Sceptic Party, or Prof. Richard Tol), and uncommitted large news media (such as RTCCNewsWire). In the “curious” user group (D1), the most common supportive users are individuals tweeting on their personal account (such as a self-named communication consultant), and organizations linked to legal issues (such as Environmental Justice Foundation, and environmental law organizations). Uncommitted users here include, for example, a geological consortium, citizen scientists, and more broadly movements related to natural resources, while critical voices in this position are mainly individuals. In the “famous” position (A4) the supportive groups are large environmental organizations (such as individual accounts linked to Greenpeace Nordic, Global Climate and Health Alliance, and the main ac-

count of Oxfam), the uncommitted voices include for instance an editor of a major Australian newspaper and an individual weather forecaster. The critical accounts consist of larger organizations such as ClimateDepot and ClimateRealists. This corresponds to fame with little conversational activity, typical of actors with prior offline authority. Finally, the “silent” (D4) group consists of supportive small newsfeeds, mainly focusing on environmental issues, and individuals such as ‘concerned parents’.

*Socio-semantic alignment.* In summary, critical Twitter users over-occupy the most prominent structural positions, while those supportive over-occupy ‘silent’ and less ‘vocal’ positions in the climate change debate. What type of discourse is involved in each of these positions and possibly dominates these conversations, are there groups of issues which are typical of groups of positions? We are eventually interested in appraising the amount of either shared or disjoint vocabulary, whatever the level of discussion between user groups: while polarized user groups appear to use distinct words to discuss a same issue, are there concepts that are used by both poles? Is vocabulary further influenced by the identified structural positions, which could thus affect the semantics mobilized by users? Finally, to what extent do uncommitted users have a specific vocabulary on their own, or a vocabulary that resembles other categories?

In practice, we analyze the vocabulary employed by users in the various slots and for the various alignments using the score introduced at the end of section 3. We examine content at the level of tweets i.e., we focus on the words that users employ in their posts on IPCC. We do so with respect to three partitions of categories: (a) alignment (i.e. critical, supportive, uncommitted), (b) groupings of over-represented slots (i.e. star, famous, curious, silent), and (c) a combination of both for the most significant slots for each alignment, i.e. *critical*  $\times$  *star*, *famous*  $\times$  *uncommitted*, and *silent*  $\times$  *supportive*. For each partition, we first compute the top  $t$  words for each category, arbitrarily fixing  $t = 15$ . The vocabulary extracted through this procedure is both relevant (all words relate to some extent specifically to the IPCC discussion) and consistent (there are only 51 distinct unique words out of a theoretical maximal value of 150 words, with significant overlap across several categories). For this reason, we consider these 51 words to be a reliable *reference vocabulary* from which to base the remainder of our analysis. We finally compute scores for all categories and partitions using this vocabulary. Results are gathered on Table 3.

From there, we perform two distinct, complementary types of analyses for each partition. First, we look at the overlap and differences between vocabularies among categories. Let us focus on alignments. We represent the vocabulary specific to each of these three categories by using a Venn diagram, as shown on the top panel of Fig. 7. We observe a strong discrepancy in the terms used, on the one hand, by critical users and, on the other hand, by supportive and uncommitted users, who seem to have a significant common vocabulary. Furthermore, while the latter users seem to focus on the effects and policy impacts of the IPCC reports, the former appear to essentially discuss scientific matters — be it science, scientists (including

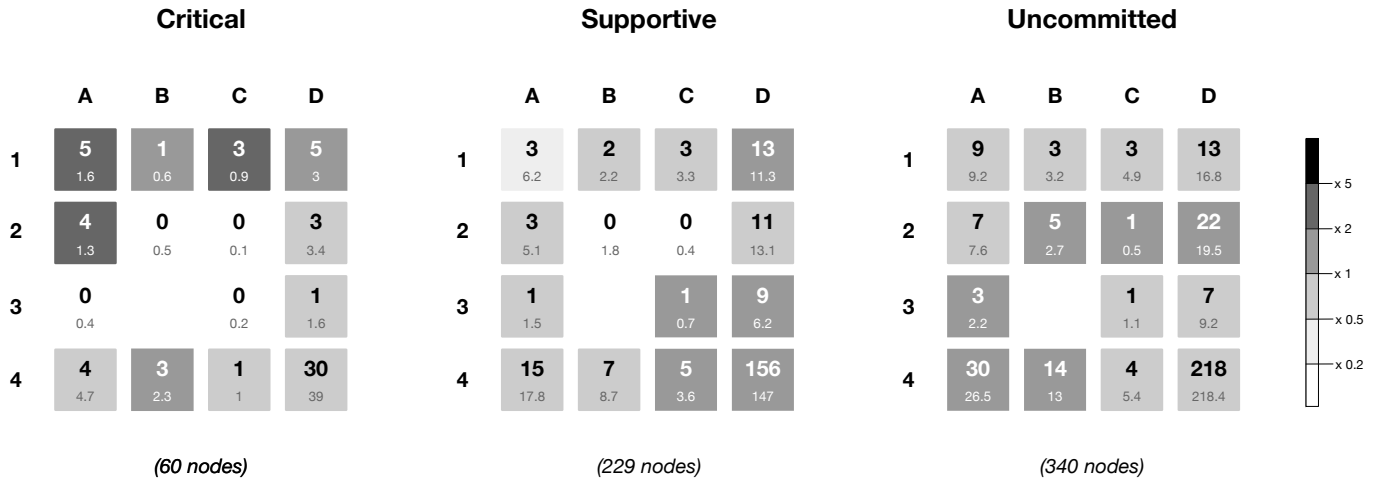


Figure 5: Over- or under-representation of discourse types for each slot. The three matrices on the describe how each type of alignment (C, S, or U) is more or less present than should be expected if position did not matter, i.e. if it were uniformly represented in each slot.

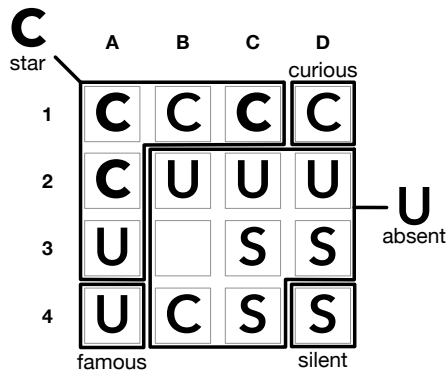


Figure 6: This global matrix describes which type of alignment is the most over-represented for each slot (bold characters additionally indicate a ratio of overrepresentation above 2). Dominant types are also figured at the level of slot groups.

Richard Tol, outspokenly skeptical of the importance of climate change), models, together with the expression of a sentiment of alarmism. This phenomenon is even stronger when combining alignment with the positions where they are most over-represented, as shown on the bottom panel of Fig. 7. Here, no vocabulary is shared at all between the critical users occupying the top star positions, who further focus on scientific matters (including stronger terms such as belief, fraud, and even lie), and the other prototypical group of users for supportive or uncommitted discourses. Remember that manual coding was performed by inferring an account’s stance towards climate change from self-presentations and external knowledge on the users or organizations an account belongs to. As a result, a few terms such as “alarmism” which, as noted before, is often indicative of climate skeptical discourse, can unsurprisingly be expected to be specific of critical users. Notwithstanding, other terms could be a priori attributed to any pole, or at least be a common ground for all groups — for instance, “earth” and “planet”; “government” and “science”; or even “belief” and “lie”. This is

Slot group	Most typical terms
<i>star</i>	agw, alarmist, ar5, author, belief, co2, data, global, model, paper, prediction, science, tol, weather
<i>famous</i>	adaptation, continent, emission, food, fuel, hope, impact, security, water
<i>curious</i>	assessment, cost, earth, graph, ocean
<i>silent</i>	catastrophe

Table 2: Terms which are most specific to topological categories among the reference vocabulary of 51 words. (Note: “agw” stands for “Anthropogenic Global Warming”, “ar5” relates to the so-called fifth Assessment Report, and “tol” typically refers to Richard Tol.)

however not the case. Furthermore, we observe that supportive and uncommitted share a lot of terms. This additionally hints at the similarity of discourses between these latter users: they appear to largely share vocabulary with users coded as supportive i.e., who display an explicit support to the reality of climate change.

Second, we examine the semantics induced by the structural partition defined by the topological model per se, especially at the level of slot groups — star, famous, curious, and silent. Since this partition features four categories, for the sake of clarity we do not use Venn diagrams (which become unwieldy beyond three categories). Rather, we describe the most specific terms for each category.<sup>9</sup> To this end, we use the list on Table 3 to extract terms whose score deviates the most for a given category with respect to the other. More precisely, we assign each term to the category where it obtains the highest score. This way, each term belongs to one and only one category; overlaps are not possible anymore. This yields the lists of Table 2. Here, the partitioning is based on the structural model and is thus topological. It indicates topics most typi-

<sup>9</sup>Similarly, for the sake of clarity, we do not carry out the present analysis on the previous partitions — interested readers may nonetheless look further at Table 3 to examine further the specifics of these results.

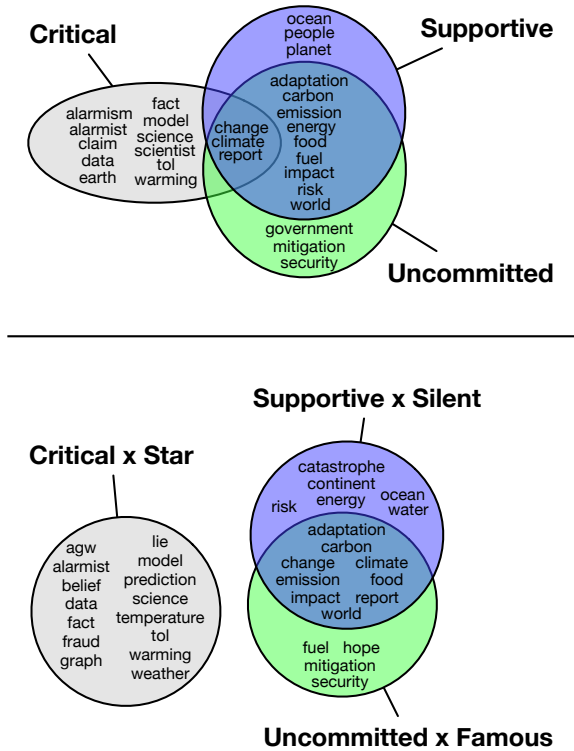


Figure 7: Venn diagram of the top 15 terms of each alignment category (*top*) and of each alignment category combined with the most representative slot grouping (*bottom*).

cally debated in each slot group, from dominant to less dominant positions, independently of their respective composition in terms of critical, supportive or uncommitted users (even though we know that there is an uneven distribution thereof in each slot group). We notice that the dominant positions in terms of activity and authority (star) are involved in a discussion based on science, while the authoritative-only positions (famous) are comparatively more focused on political issues, consequences and possible actions. Most interestingly, the other positions do not seem to exhibit a specific vocabulary of their own: this is most prominent in the case of the “silent” slot D4 whose only specific term is “catastrophe”, even though it is the most populated. Moreover, this slot corresponds to nodes which are least dominant in conversational terms (mentioning fewer others and being mentioned by fewer actors) but not necessarily in publication terms, since by definition of the heart, they also publish above the threshold of engagement with the topic.

To summarize, there are two broad classes of vocabularies representing the different ways of framing the debate, one typical of critical users, the other one more representative of both supportive and uncommitted users. This hints at the possibility that uncommitted and supportive users are affiliated with the same framing of the climate change debate. Moreover, we learn that there is little semantic mixing at this alignment level (Fig. 7-bottom); all the more for users occupying positions most typical of their alignment (e.g., critical stars). Finally, the most dominant structural positions appear to be nonetheless charac-

terized by a vocabulary typical of critical users.

## 5. Discussion and Conclusion

Building upon earlier studies of social networks of actors on Twitter, and research into the active or influential Twitter users, we proposed a novel methodological approach for a more detailed view on critical, neutral and supportive actors, and their relative position in the debate around the IPCC AR5 report on Twitter. Online citation networks have typically been shown to be structured in separate clusters that roughly correspond to opposing sides of a debate (Adamic and Glance, 2005; Conover et al., 2011; Kaiser and Puschmann, 2017; Gaumont et al., 2018). In the context of Twitter, retweets indeed configure reference spaces which unsurprisingly mimic the various poles of the epistemic spaces of actors. The sub-networks induced by focusing on each of the various sides (Shi et al., 2020) or on the most visible actors only (Newman, 2017; Wang et al., 2020) may not provide a comprehensive insight on the integrated discussion space gathering all sides, at various levels of visibility and activity. This is all the more the case in mention networks i.e., conversation networks, where homophily is notoriously weak and conversations are intertwined between opposing poles (Lietz et al., 2014; Pearce et al., 2014) such that structural boundaries are blurred by socio-semantic mixing.

By recursively analyzing who occupies the dominant and less dominant positions of mention networks in terms of activity, authority, and content, we find a series of results that contrast in several ways with the state of the art. Traditionally, the communication literature on climate change has long established that critical voices are more represented in mainstream and print media than their demographic weight and scientific consensus would suggest (Hulme, 2009; Painter and Ashe, 2012). Online discussion spaces, on the other hand, have so far been described to be rather dominated, *on average*, by neutral (Pearce et al., 2014) or supportive users (Williams et al., 2015), or supportive frames (O’Neill et al., 2015). This average or aggregate perspective masks however the inner structure of such conversational playgrounds. Our results show that while critical voices are in significant numerical minority in our focus perimeter consisting of Twitter users who are the 2% most engaged in the discussion of the IPCC reports, in terms of mentions and thus direct engagement with other users, they occupy much more a “star” position than expected. The supportive majority of users are rather over-represented in the so-called “silent” part of the heart. For one, the results have implications for climate change communication in showing that, in general, those critical toward the scientific basis of climate change hold disproportionately more prominent structural positions in Twitter conversations. The relative prominence of minority opinions in digital public space also calls for further research into the relative positions of users holding different opinions, beyond the climate change debate. One of the limitations of the present contribution lies in the statistical significance of some slots, where the number of users is rather small, and a replication effort on larger and further case studies would be much welcome.

	C	S	U	star	famous	curious	silent	star × C	silent × S	famous × U
<i>adaptation</i>	0.01	<b>0.38</b>	<b>0.3</b>	0.39	<b>0.49</b>	0.22	0.17	0.06	<b>0.34</b>	<b>0.72</b>
<i>agw</i>	0.41	0.0	0.0	0.34	0.0	0.04	0.01	<b>1.23</b>	0.0	0.0
<i>alarmism</i>	<b>0.57</b>	0.01	0.03	0.19	0.05	0.0	0.08	0.75	0.02	0.01
<i>alarmist</i>	<b>0.42</b>	0.0	0.0	0.24	0.02	0.0	0.01	<b>0.89</b>	0.0	0.0
<i>ar5</i>	0.23	0.3	0.17	<b>0.61</b>	0.26	<b>0.39</b>	0.08	0.41	0.15	0.24
<i>assessment</i>	0.09	0.08	0.12	0.13	0.05	<b>0.4</b>	0.08	0.1	0.07	0.08
<i>author</i>	0.29	0.14	0.23	<b>0.65</b>	0.16	0.09	0.16	0.52	0.15	0.28
<i>belief</i>	0.11	0.0	0.0	0.17	0.0	0.0	0.0	<b>1.03</b>	0.0	0.0
<i>carbon</i>	0.06	<b>0.43</b>	<b>0.36</b>	0.36	<b>0.35</b>	0.18	<b>0.33</b>	0.27	<b>0.47</b>	<b>0.42</b>
<i>catastrophe</i>	0.06	0.29	0.19	0.07	0.16	0.16	<b>0.3</b>	0.02	<b>0.38</b>	0.18
<i>change</i>	<b>0.44</b>	<b>0.69</b>	<b>0.62</b>	<b>0.64</b>	<b>0.62</b>	<b>0.57</b>	<b>0.63</b>	0.69	<b>0.71</b>	<b>0.62</b>
<i>claim</i>	<b>0.57</b>	0.0	0.02	0.23	0.04	0.1	0.03	0.68	0.0	0.0
<i>climate</i>	<b>0.62</b>	<b>0.69</b>	<b>0.66</b>	<b>0.69</b>	<b>0.66</b>	<b>0.67</b>	<b>0.66</b>	0.72	<b>0.69</b>	<b>0.64</b>
<i>co2</i>	<b>0.45</b>	0.11	0.11	0.4	0.02	0.14	0.1	0.77	0.1	0.0
<i>continent</i>	0.01	0.33	0.12	0.04	0.28	0.17	0.18	0.03	<b>0.35</b>	0.07
<i>cost</i>	0.2	0.25	0.2	0.25	<b>0.42</b>	<b>0.42</b>	0.12	0.36	0.19	<b>0.46</b>
<i>coverage</i>	0.01	0.05	0.13	0.32	0.24	0.1	0.01	0.11	0.01	<b>0.53</b>
<i>data</i>	<b>0.57</b>	0.01	0.03	0.35	0.01	0.09	0.04	<b>0.82</b>	0.01	0.0
<i>earth</i>	<b>0.54</b>	0.14	0.05	0.21	0.12	0.27	0.11	0.49	0.13	0.01
<i>emission</i>	0.07	<b>0.47</b>	<b>0.38</b>	0.4	<b>0.51</b>	<b>0.38</b>	<b>0.24</b>	0.15	<b>0.4</b>	<b>0.84</b>
<i>energy</i>	0.11	<b>0.39</b>	<b>0.36</b>	<b>0.45</b>	0.18	0.31	<b>0.31</b>	0.36	<b>0.36</b>	0.2
<i>fact</i>	<b>0.51</b>	0.04	0.04	0.37	0.17	0.13	0.03	<b>1.03</b>	0.02	0.11
<i>food</i>	0.12	<b>0.45</b>	<b>0.33</b>	0.23	<b>0.56</b>	0.21	<b>0.34</b>	0.34	<b>0.43</b>	<b>0.53</b>
<i>fraud</i>	0.34	0.0	0.0	0.18	0.0	0.0	0.01	<b>0.93</b>	0.0	0.0
<i>fuel</i>	0.05	<b>0.36</b>	<b>0.26</b>	0.28	<b>0.37</b>	0.3	0.18	0.29	0.3	<b>0.4</b>
<i>global</i>	0.41	0.15	0.11	<b>0.41</b>	0.06	0.14	0.18	0.65	0.16	0.0
<i>government</i>	0.22	0.11	<b>0.25</b>	0.33	0.23	0.03	0.19	0.34	0.12	0.31
<i>graph</i>	0.19	0.0	0.0	0.16	0.01	0.2	0.0	<b>0.87</b>	0.0	0.01
<i>hope</i>	0.04	0.08	0.08	0.1	0.33	0.01	0.04	0.0	0.07	<b>0.37</b>
<i>impact</i>	0.15	<b>0.49</b>	<b>0.34</b>	0.39	<b>0.59</b>	<b>0.38</b>	<b>0.28</b>	0.21	<b>0.41</b>	<b>0.6</b>
<i>lie</i>	0.22	0.05	0.0	0.15	0.01	0.08	0.03	<b>0.83</b>	0.08	0.0
<i>mitigation</i>	0.04	0.28	<b>0.32</b>	0.35	0.29	0.23	0.15	0.06	0.19	<b>0.45</b>
<i>model</i>	<b>0.63</b>	0.0	0.02	0.26	0.06	0.06	0.02	<b>0.82</b>	0.0	0.02
<i>ocean</i>	0.04	<b>0.36</b>	0.16	0.06	0.27	<b>0.43</b>	0.19	0.02	<b>0.33</b>	0.15
<i>paper</i>	0.13	0.02	0.04	<b>0.46</b>	0.09	0.0	0.0	0.55	0.0	0.1
<i>people</i>	0.22	<b>0.33</b>	0.18	<b>0.48</b>	<b>0.33</b>	<b>0.38</b>	0.18	0.57	0.28	0.28
<i>planet</i>	0.04	<b>0.38</b>	0.14	0.26	0.2	0.24	0.15	0.14	0.31	0.08
<i>policy</i>	0.18	0.13	0.17	<b>0.41</b>	0.16	<b>0.39</b>	0.07	0.69	0.1	0.25
<i>prediction</i>	0.36	0.01	0.01	0.32	0.02	0.04	0.01	<b>0.92</b>	0.0	0.01
<i>report</i>	<b>0.62</b>	<b>0.7</b>	<b>0.68</b>	<b>0.71</b>	<b>0.69</b>	<b>0.6</b>	<b>0.68</b>	0.71	<b>0.7</b>	<b>0.67</b>
<i>risk</i>	0.06	<b>0.55</b>	<b>0.38</b>	<b>0.41</b>	<b>0.35</b>	<b>0.39</b>	<b>0.35</b>	0.23	<b>0.53</b>	0.33
<i>science</i>	<b>0.65</b>	0.27	0.25	<b>0.75</b>	0.22	<b>0.4</b>	<b>0.27</b>	<b>0.88</b>	0.27	0.15
<i>scientist</i>	<b>0.56</b>	0.25	0.23	<b>0.52</b>	0.28	<b>0.51</b>	<b>0.25</b>	0.81	0.23	0.13
<i>security</i>	0.06	0.29	<b>0.28</b>	0.17	<b>0.35</b>	0.26	<b>0.25</b>	0.13	0.31	<b>0.49</b>
<i>summary</i>	0.23	0.13	0.19	0.35	0.2	0.26	0.08	0.28	0.08	0.3
<i>temperature</i>	0.17	0.09	0.05	0.17	0.15	0.08	0.06	<b>0.83</b>	0.08	0.13
<i>tol</i>	<b>0.44</b>	0.01	0.02	0.29	0.03	0.01	0.03	<b>0.96</b>	0.01	0.01
<i>warming</i>	<b>0.6</b>	0.22	0.18	<b>0.46</b>	<b>0.36</b>	0.19	<b>0.24</b>	<b>0.99</b>	0.19	0.18
<i>water</i>	0.03	0.25	0.07	0.11	0.14	0.12	0.13	0.24	<b>0.32</b>	0.19
<i>weather</i>	0.17	0.03	0.08	0.3	0.12	0.02	0.03	<b>0.82</b>	0.01	0.17
<i>world</i>	0.22	<b>0.61</b>	<b>0.43</b>	<b>0.52</b>	<b>0.51</b>	<b>0.51</b>	<b>0.39</b>	0.72	<b>0.57</b>	<b>0.57</b>

Table 3: Scores for the 51 words which appear at least once among the top 15 words for the ten node categories of interest: either critical (C), supportive (S), uncommitted (U); star, famous, curious, silent; or critical star, famous uncommitted, silent supportive. (Figures in bold indicate words which indeed appear in the top 15 for the corresponding category.)

Furthermore, the semantic analysis of the words used by those critical and those supportive of climate change shows two different types of discussions: critical Twitter users discuss climate change in terms of beliefs (instead of scientific facts), fraud and lies (questioning the scientific models and facts presented in the IPCC report). Those supportive tend to focus more strictly on the content of the IPCC report, in particular words related to the topical focus of the WG2 and WG3 reports on

adaptation and mitigation, respectively. While the online discussion of climate change has already been studied by focusing on top-cited (O’Neill et al., 2015; Newman, 2017) or specific content (Martin and Rice, 2014; Jacques and Knox, 2016; Tyagi et al., 2020), we demonstrate how it concurrently depends on the structural position and semantic alignment of actors. We notably underline how several terms are specific of particular dominant positions and minority alignments.

More broadly, the socio-semantic positional analysis that we proposed in this paper makes it possible to contrast the structural dominance and the semantic cohesiveness of opposing sides in contexts where socio-semantic clusters precisely appear to be intertwined, as is often the case in Twitter conversation networks. As such, we contend that our approach may be generally helpful to disentangle semantic and structural polarization in online conversation spaces, and to show how semantics depend not only on poles but also on their structural positions.

## References

- Adamic, L.A., Glance, N., 2005. The political blogosphere and the 2004 U.S. election: divided they blog, in: *LinkKDD '05: Proc. 3rd Intl. Workshop on Link discovery*, ACM. pp. 36–43.
- Ayres, J.M., 1999. From the streets to the internet : The cyber-diffusion of contention. *Annals of the American Academy of Political and Social Science* 566, 132–143.
- Bakshy, E., Messing, S., Adamic, L.A., 2015. Exposure to ideologically diverse news and opinion on facebook. *Science* 348, 1130–1132.
- Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R., 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 26, 1531–1542.
- boyd, D., Golder, S., Lotan, G., 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter, in: *2010 43rd Hawaii International Conference on System Sciences*, IEEE. pp. 1–10.
- Bruns, A., Burgess, J., 2011. The use of twitter hashtags in the formation of ad hoc publics, in: *Proceedings of the 6th European consortium for political research (ECPR) general conference 2011, The European Consortium for Political Research (ECPR)*. pp. 1–9.
- Cardon, D., Fouetillou, G., Roth, C., 2011. Two paths of glory – structural positions and trajectories of websites within their topical territory, in: *Proc. ICWSM 5th Intl Conf Weblogs Social Media*, AAAI. pp. 58–65.
- Cattuto, C., Barrat, A., Baldassarri, A., Schehr, G., Loreto, V., 2009. Collective dynamics of social annotation. *PNAS* 106.
- Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., Menczer, F., 2011. Political polarization on twitter, in: *Proc. 5th ICWSM Intl. Conf. Weblogs and Social Media*, AAAI. pp. 89–96.
- Dahlgren, P., 2005. The internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication* 22, 147–162.
- Elgesem, D., Stskal, L., Diakopoulos, N., 2015. Structure and content of discourse on climate change in the blogosphere: The big picture. *Environmental Communication* 9, 169–188.
- Etling, B., Kelly, J., Faris, R., Palfrey, J., 2009. Mapping the Arabic blogosphere: Politics, culture, and dissent. Technical Report 2009-06. Berkman Center Research Publication.
- Fownes, J.R., Yu, C., Margolin, D.B., 2018. Twitter and climate change. *Sociology Compass* 12, e12587.
- Garimella, K., De Francis Morales, G., Gionis, A., Mathioudakis, M., 2016a. Quantifying controversy in social media, in: *Proc. 9th WSDM Intl Conf Web Search and Data Mining*, ACM. pp. 33–42.
- Garimella, K., Weber, I., De Choudhury, M., 2016b. Quote RTs on Twitter: Usage of the new feature for political discourse, in: *Proceedings of the 8th ACM Conference on Web Science*, pp. 200–204.
- Gaumont, N., Panahi, M., Chavalarias, D., 2018. Reconstruction of the socio-semantic dynamics of political activist twitter networks—method and application to the 2017 French presidential election. *PLOS ONE* 13, e0201879.
- Gerhards, J., Schäfer, M.S., 2010. Is the internet a better public sphere? comparing old and new media in the us and germany. *New Media & Society* 12, 143–160.
- Hellsten, I., Jacobs, S., Wonneberger, A., 2019. Active and passive stakeholders in issue arenas: A communication network approach to the bird flu debate on twitter. *Public Relations Review* 45, 35–48.
- Himmelboim, I., McCreery, S., 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, 154–174.
- Himmelboim, I., Smith, M.A., Rainie, L., Shneiderman, B., Espina, C., 2017. Classifying twitter topic-networks using social network analysis. *Social Media + Society* 3, 1–13.
- Hindman, M., Tsiouliouklis, K., Johnson, J.A., 2003. Googlearchy: How a few heavily-linked sites dominate politics on the web, in: *Proc. Annual Meeting of the Midwest Political Science Association*, pp. 1–33.
- Hoffmann, I., 2014. The Populist Networks. spotlight Europe 2014/02. Bertelsmann Stiftung.
- Holmberg, K., Hellsten, I., 2016. Twitter campaigns around the fifth ipcc report: Campaign spreading, shared hashtags, and separate communities. *SAGE Open* 6, 1–7.
- Howarth, C., Sharman, A., 2017. Influence of labeling and incivility on climate change communication. Technical Report. Oxford Research Encyclopedia, Climate Change. Doi:10.1093/acrefore/9780190228620.103.382.
- Hulme, M., 2009. Mediating the messages about climate change. Reporting the IPCC fourth assessment in the UK print media, in: Boyce, T., Lewis, C. (Eds.), *Climate Change and the Media (Global Crises and the Media)*. Peter Lang, pp. 117–128.
- Jacques, P.J., Knox, C.C., 2016. Hurricanes and hegemony: A qualitative analysis of micro-level climate change denial discourses. *Environmental Politics* 25, 831–852.
- Jang, S.M., Hart, P., 2015. Polarized frames on “climate change” and “global warming” across countries and states: Evidence from Twitter big data. *Global Environmental Change* 32, 11–17.
- Kaiser, J., Puschmann, C., 2017. Alliance of antagonism: Counterpublics and polarization in online climate change communication. *Communication and the Public* 2, 371–387.
- Kim, S., Weber, I., Wei, L., Oh, A., 2014. Sociolinguistic analysis of twitter in multilingual societies, in: *Proc. HT'14 25th Intl Conf Hypertext and social media*, ACM. pp. 243–248.
- Lietz, H., Wagner, C., Bleier, A., Strohmaier, M., 2014. When politicians talk: Assessing online conversational practices of political parties on twitter, in: *Proc. ICWSM 8th Intl. Conf. Weblogs and Social Media*, AAAI. pp. 285–294.
- Maireder, A., Ausserhofer, J., 2013. Political discourses on twitter: Networking topics, objects and people, in: Weller, K., Bruns, A., Burgess, J., Mahrt, M., Puschmann, C. (Eds.), *Twitter and Society*. Peter Lang, New York, NY, pp. 305–318.
- Martin, N., Rice, J., 2014. Rebalancing climate change debate and policy: An analysis of online discussions. *Environmental Policy and Governance* 24, 338–350.
- Newman, T.P., 2017. Tracking the release of IPCC AR5 on Twitter: Users, comments and sources following the release of the working group 1 summary for policymakers. *Public Understanding of Science* 26, 815–825.
- O’Neill, S., Williams, H.T., Kurz, T., Wiersma, B., Boykoff, M., 2015. Dominant frames in legacy and social media coverage of the IPCC fifth assessment report. *Nature Climate Change* 5, 380–385.
- Painter, J., 2011. Poles Apart: The International Reporting of Climate Scepticism. Oxford: Reuters Institute for the Study of Journalism.
- Painter, J., Ashe, T., 2012. Cross-national comparison of the presence of climate scepticism in the print media in six countries, 2007–10. *Environmental Research Letters* 7, 044005.
- Pearce, W., Holmberg, K., Hellsten, I., Nerlich, B., 2014. Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group I report. *PLOS ONE* 9.
- Pearce, W., Niederer, S., Özkula, S.M., Sánchez Querubín, N., 2019. The social media life of climate change: Platforms, publics, and future imaginaries. *Wiley Interdisciplinary Reviews: Climate Change* 10, e569.
- Roth, C., Cointet, J.P., 2010. Social and semantic coevolution in knowledge networks. *Social Networks* 32, 16–29.
- Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management* 24, 513–523.
- Schäfer, M.S., 2012. Online communication on climate change and climate politics: a literature review. *Wiley Interdisciplinary Reviews: Climate Change* 3, 527–543.
- Shi, W., Fu, H., Wang, P., Chen, C., Xiong, J., 2020. Climatechange vs. global-warming: Characterizing two competing climate discourses on twitter with semantic network and temporal analyses. *International journal of environmental research and public health* 17, 1062.
- Shirky, C., 2011. The political power of social media : Technology, the public sphere, and political change. *Foreign Affairs* 90, 28–41.

- Sunstein, C.R., 2007. *Republic.com 2.0*. Princeton University Press.
- Tyagi, A., Babcock, M., Carley, K.M., Sicker, D.C., 2020. Polarizing tweets on climate change, in: *Social, Cultural, and Behavioral Modeling*, Springer. pp. 107–117.
- Van Aelst, P., Walgrave, S., 2002. New media, new movements ? the role of the internet in shaping the 'anti-globalization' movement. *Information, Communication & Society* 5, 465–493.
- Van Alstyne, M., Brynjolfsson, E., 1996. Electronic communities: Global village or cyberbalkans?, in: DeGross, J., Jarvenpaa, S., Srinivasan, A. (Eds.), *Proc 17 ICIS Intl Conf Information Systems*, pp. 80–98.
- Vasileiadou, E., Heimeriks, G., A., P., 2011. Exploring the impact of the ipcc assessment reports on science. *Environmental Science & Policy* 14, 1052–1061.
- Veltri, G.A., Atanasova, D., 2017. Climate change on twitter: Content, media ecology and information sharing behaviour. *Public Understanding of Science* 26, 721–737.
- Vicario, M.D., Zollo, F., Caldarelli, G., Scala, A., Quattrociocchi, W., 2017. Mapping social dynamics on facebook: The brexit debate. *Social Networks* 50, 6 – 16.
- Vu, H.T., Do, H.V., Seo, H., Liu, Y., 2019. Who leads the conversation on climate change? A study of a global network of NGOs on Twitter. *Environmental Communication* 14, 450–464.
- Wang, X., You, Y., Lin, L., 2020. Tweeting the United Nations Climate Change Conference in Paris (COP21): An analysis of social network and factors determining the network influence. *Online Social Networks and Media* 15, 100059.
- Williams, H.T., McMurray, J.R., Kurz, T., Lambert, F.H., 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global environmental change* 32, 126–138.