The Meaning of Links

On the interpretation of hyperlinks in the study of polarization in blogging about climate change

Dag Elgesem

Department of Information Science and Media Studies, University of Bergen, Norway, Dag.Elgesem@uib.no

Abstract

This article explores the potential and challenges of using hyperlinks as data through a study of polarization in English language blogs about climate change. The purpose of this research is to provide an interpretation of the meaning of the hyperlinks in climate change blogs by coding the functions that the links perform in the given blog posts. Beginning with a set of more than 500,000 blog posts about climate change, we focus on bloggers who actively link to highly visible sources that advocate, respectively, the denial or acceptance of the consensus view on anthropogenic climate change. We find that the bloggers in our sample predominantly link to sources that they agree with and that, if they link to a source with different opinions, the link is part of negative criticism of the targeted source. We argue that, by considering the functions of the links in the blog posts, we obtain a more nuanced understanding of the extent to which the discussion in the blogs is polarized.

Keywords: blogs, climate change, polarization, link studies

Introduction

Links are important as data in analyses of the social web (De Mayer, 2013). However, the social interpretation of links is highly context dependent and often underdetermined by the available data. Thelwall (2006) rightly pointed out that there can be no *general* theory of the semantics of hyperlinks, because the intentions behind the links can vary endlessly. The challenge, then, is to find methodologically sound ways to combine the analysis of patterns in the links with analyses of their functions. This article contributes to the theme of this special issue on "making sense of big and small data as onlife traces" by analysing the functions of links in blog posts' discussion of climate change. In particular, we investigate the role of links in the polarized discussion between those who accept and those who reject the idea of anthropogenic global warming. We find that most of the links are homophilic; that is, the function of the link is, in most cases, to endorse another blog, which shares the ideology of the linking blog. Furthermore, the

Elgesem, D. (2019). The meaning of links: On the interpretation of hyperlinks in the study of polarization in blogging about climate change. *Nordicom Review,* 40 (Special Issue 1): 65-78. doi:10.2478/nor-2019-0014.

function of links across the ideological divide is in most cases to back up criticism of the position of the linked blog. In this study, the linking practices of the climate change bloggers reflect and reproduce the polarized nature of the debate on climate change.

In a polarized situation, there is little common ground between the adversaries and therefore a scarce basis for a discussion in which the parties take the arguments of the other side seriously. In the literature, polarization refers to a *process* whereby people change to a more extreme variant of their original position as the result of deliberation.

The term "group polarization" refers to something simple: after deliberation, people are more likely to move toward a more extreme point in the direction to which the group's members were originally inclined. With respect to the internet and social media, the implication is that groups of like-minded people, engaged in discussion with each other, will typically end up thinking the same thing that they thought before – but in a more extreme form (Sunstein, 2017).

However, in most cases, as in the present case of blogging about climate change, we do not have access to the process through which polarization develops: we have access only to the *state* of a system of communication that is the result of that process. Neither do we have access to measurements of the extremeness of people's opinions at different points of time. Nevertheless, it makes sense to study the state of a system of communication from the perspective of polarization: we can meaningfully ask whether polarization is a plausible explanation for the properties of the system that we can observe. We will follow this strategy in this article.

There are different theories about the mechanisms underlying polarization. One explanation is that, in the course of discussions among a group of people who more or less agree on an issue, arguments in support of the favoured position will mostly emerge. Since the people involved in the discussion are presented with a limited pool of arguments that support one side of the issue, they end up becoming more strongly attached to the group's view. Polarization happens, according to this theory, because people become more entrenched in their own views and thereby more extreme (Fiske & Taylor, 2013; Isenberg, 1986).

According to another theory, polarization is more likely to happen in groups in which a political position is an important part of the group identity. In this theory, polarization is more likely to happen in a group in which people see themselves as having a shared identity by virtue of having a common cause. Adopting a more extreme view tends to strengthen the group identity and at the same time sharpen the opposition to other points of view (Fiske & Taylor, 2013; Sunstein, 2017).

A third explanation is that polarization is more likely to occur when people read or hear arguments that challenge their own views. On this account, polarization happens because people, when they encounter challenging arguments, start to activate counter-arguments to defeat the challenging arguments. By rehearsing counter-arguments, people tend to become more entrenched in their own views and thereby can move towards more extreme versions of their position (Lodge & Taber, 2013).

These explanations are not in conflict with each other, as the different mechanisms can be operative at the same time. The question here is whether these theories are plausible explanations for the patterns that we see in the linking in climate change blogs. We will return to this question in the discussion below.

Related work

A number of studies have documented the role of social networks and group affiliation in shaping political preferences and prejudices (e.g. Pettigrew, 1998; Sinclair, 2012). Studies of political communication in the online public sphere have added to this knowledge by analysing the polarization in the patterns of the links, for example Adamic (2008), Adamic and Glance (2005), Bakshy and colleagues (2015), Elgesem and colleagues (2015), Himelboim and colleagues (2013) and Yardi and Boyd (2010). The present article falls into this tradition, as we combine link analysis with coding the functions of the links in blog posts about climate change.

The discussion on climate change in parts of society is highly polarized, and this process is exacerbated by the involvement of powerful actors (Dunlap & McCright, 2011), characterized by conflicting framings of the problems (Hulme, 2009), and the contrarians' lack of trust in mainstream climate science is an important driver (Mann, 2012). Previous research has shown that the blogging on climate change is also highly polarized (Elgesem et al., 2015; Sharman, 2014).

The number of blogs discussing different aspects of climate change is high (Elgesem et al., 2015). Blogs are the main outlet for those who reject the consensus view on anthropogenic global warming (AGW), which receives relatively less attention in the mainstream media than the consensus view on climate change. One would expect the polarization in the blog posts to be visible in the link patterns. Previous research has shown that there are more links connecting blogs that share the same position on climate change than there are links connecting blogs with opposing positions (Elgesem et al., 2015). This seems to suggest that we should read a link as an endorsement of the views expressed by its target source. There are, however, problems with such an interpretation. First, as convincingly argued by Guera and colleagues (2013), the fact that a group of blogs systematically links to one group of sources and not another does not necessarily mean that it does not want to engage with a group that it does not link to – a simpler explanation could be that it is not aware of it. Second, the meaning of a link cannot be inferred from the type of the linked source alone: for example, a blogger can link to a source that she disagrees with either to criticize it or to recognize that the opponent has a valid point. Our solution to the first problem is to focus on sources that we have reason to believe active climate change bloggers know. To handle the second challenge, we analyse in detail the functions that the links perform in the blog post. In particular, we chart whether the links appear in the context of a criticism of the linked source, whether the linking blog and the linked source agree on the issue at hand or whether the linking is neutral.

Research questions

Our overarching research question is:

- To what extent are the patterns in the linking from blogs about climate change signs of polarization between the blogs that accept the consensus view on climate change and those that reject that view?

We address this overarching research question by breaking it down into four sub-questions:

- RQ1: What is the probability that a link to a source that rejects the mainstream view on climate change comes from a blog that itself rejects the mainstream view?
- RQ2: What is the probability that a link to a source that endorses the mainstream view on climate change comes from a blog that itself endorses the mainstream view?
- RQ3: What are the functions of the links that cross the ideological divide? Are they mainly formed in the context of criticism of the other position or as part of a dialogue?
- RQ4: What are the functions of links from blogs to sources that share the blog's position on climate change?

Note that, by taking as the starting point for our analysis all types of sources that the blogs link to, and not only blogs, we reject the idea of a separate "blogosphere" as unfruitful. Blogs link to all kinds of resources, and a blog will typically have many more links to the media, organizations and so on than to other blogs. However, blogs often have the role of alternative media: criticizing, amplifying and filtering information produced by the mainstream media and other sources (Bruns, 2005). A characteristic feature of climate change blogs is that, while most bloggers subscribe to the consensus on climate change, many blogs actively deny the consensus view. Moreover, while these sceptical bloggers define themselves and their position in terms of rejection of the consensus view, those who accept the consensus view do not define themselves as anti-sceptics. The situation is asymmetric: while the sceptics use their energy to fight against the consensus view on anthropogenic climate change, most non-sceptics do not spend much time quarrelling with the sceptics. There can be no doubt that it is the deniers of anthropogenic climate change who represent the extreme view on this issue. We thus have *asymmetric polarization*.

The data

We collected the data via the API of the Swedish company Twingly (twingly.com). The company indexes blogs and offers paid access to its database of blogs in a range of different languages, including English. One challenge with harvesting blogs is that no simple definition of a blog exists. Twingly, the company providing us with access to the data, uses the Wikipedia "definition" of a blog: "a discussion or informational website published on the World Wide Web consisting of discrete, often informal diary-style text entries ('posts'). Posts are typically displayed in reverse chronological order, so that the most recent post appears first, at the top of the web page."[1] The definition is not precise. For example, the distinction between a blog and an online magazine is not clear. Twingly keeps its operationalization of the definition secret. However, in the present study, our aim was to chart not the complete blogosphere but only a central part of it. Moreover, over many years, Twingly has developed through trial and error a method that produces results that seem to be compatible with its customers' concept of a blog.

We used a wide criterion to obtain a maximally inclusive set of blog posts and searched for posts that contained ("climate change" OR "global warming") published in the period between 1 May 2016 and 1 May 2017. The result was 570,498 blog posts in English. We used the following data about the posts: the date of publication, the blog URL, the title of the post, the text of the blog post, the URL of the post and the links from the post.

We wanted to identify blog posts that focus on climate change rather than those that merely mention the terms "climate change" or "global warming" in the context of other discussions. Based on previous experience and by reading a number of randomly selected posts, we found that some posts, although they contain an occurrence of one of the key phrases, are not really about climate change but only mention the search terms in the course of discussions of other issues. To increase the probability of including posts that focus on our topic, we chose to include only posts that have two occurrences of the key terms ("climate change" or "global warming"). This reduced the number of posts significantly to 189,297 blog posts. Moreover, since we wanted blogs that have a position on climate change, we chose to include only blog posts from blogs that have at least two posts about climate change (i.e. posts that have at least two occurrences of the two search terms). This reduced the number of relevant blog posts to 151,650, published by 14,422 different blogs.

Methods and analytical strategies

Our study used methods from quantitative text analyses to prepare the data, to conduct exploratory data analysis and to filter the material. We identified and characterized patterns in links, blogs and blog posts in the data with scripts in R developed by the author.[2] For the classification of blogs as, respectively, sceptical and non-sceptical, and for the classification of the function of the links, we used manual coding. Two coders were involved in the coding. The details will be explained later in this section.

The analytic strategy used in the argument consists of five steps:

First step: The linked sources

There were more than one million out-links from the 151,650 blog posts to different sources. However, when more than one out-link led from a blog post to the same source, we chose to count this as only one link between the post and the source. We excluded irrelevant links to ad servers and to blog platforms, for example blogger.com and wordspress.com. We ended up with 594,824 unique out-links from the blogs, targeting 67,277 unique sources.

Second step: The 100 most-linked sources

We sorted the sources by their in-degree and identified the 100 sources most frequently linked by the climate change blogs. This group consists of social media sites, mainstream media, organizations, alternative media and blogs. The blogs link to all types of sources on the web as part of their engagement with the issue of climate change. Therefore, we argue that, when analysing polarization around the issue of climate change among bloggers, we have to take into consideration the larger informational ecology of which the blogs are part.

However, in some cases, it is difficult to determine the precise position of a linked source. Let us illustrate this by considering the 10 most-linked sources (with the number of in-links in the parentheses):

1. www.twitter.com (15,166)
2. www.theguardian.com (11,666)
3. www.nytimes.com (10,222)
4. www.washingtonpost.com (8,728)
5. www.youtube.com (7,292)
6. www.facebook.com (7,171)
7. en.wikipedia.org (5,717)
8. www.huffingtonpost.com (4,530)
9. www.reuters.com (3,397)
10. www.cnn.com (3,378)

We can see that, among the ten most-linked sources, there are four social media platforms (1, 5, 6, 7), five mainstream media sites (2, 3, 4, 9, 10) and one alternative medium (8). What position do these outlets have on the question of anthropogenic climate change? The social media platforms have posts representing all variants of positions on anthropogenic climate change, so, with respect to them, there is no general answer. None of the mainstream media outlets, including Huffington Post, have a climate sceptical agenda. The newspaper *The Guardian* is the source among these ten that has most actively covered the science and politics of climate change and has taken the clearest stance in favour of the consensus view on anthropogenic climate change.

In the following, we will distinguish between sources that are central to the sceptical discourse on climate change and blogs that are central to the non-sceptical discourse on climate change. In the first group, we included, first, sources that actively reject the consensus view on climate change. There are two sources of this type among the 100 most-linked sources:

40. www.breitbart.com (1,416)
48. www.WattsUpWithThat.com (1,290)

WattsUpWithThat.com is the most visible and well-known blog in the sceptical community (Elgesem et al., 2015; Sharman, 2014), while breitbart.com is a leading outlet at the extreme right, also with a clear sceptical agenda. We used these two highly visible sources to represent the sceptical discourse on climate change and to investigate the extent to which a link to these sceptical sources is a signal that the linking blog is a sceptic.

Most of the other 97 sources are not part of the sceptical discourse. As representatives of the non-sceptical discourse, we selected eight sources, which are all both clearly in support of the consensus view on anthropogenic climate change and the most extensively linked representatives of different types of blogs. The eight sources (listed by their rank of popularity and with the number of in-links in the parentheses) are:

2. www.theguardian.com (11,666)
29. www.climatecentral.org (1,854)
35. insideclimatenews.org (1,529)
41. climate.nasa.gov (1,407)
45. www.unfccc.int (1,353)

51. www.ipcc.ch (1,134)

95. www.scepticalscience.com (721)

96. www.climatechangenews.com (719)

All of these sources are highly visible sites for information about the consensus view on climate change. We saw above that theguardian.com is the second-most-linked source in the whole corpus. While www.unfccc.int is the site of the UN activity on climate change, www.ipc.ch is the site of the IPCC. www.scepticalscience.com is devoted to the correction of sceptical claims about anthropogenic climate change. We chose to consider it as part of the non-sceptical discourse, because it actively promotes the mainstream view and the main output is explanations of "what peer reviewed science has to say about global warming".[3] www.climatecentral.org, insideclimatenews.org and www.climatechangenews.com are popular sites for the communication of mainstream climate science. We used these eight highly visible non-sceptical sources to investigate the extent to which a link to a source in the non-sceptical discourse is a signal that the linking blog is a non-sceptic.

In addition, we included in the analysis one source among the top 100 that clearly shares the mainstream view but is devoted to criticism of the sceptics:

81. www.desmogblog.com (836)

This anti-sceptical source engages actively in the sceptical discourse by conducting investigative journalism on "global warming misinformation campaigns".[4] It publishes reports by journalists and scholars aimed at exposing and criticizing the deniers. We investigated whether this source has a role in the blogging about climate change that differs from that of the sceptical sources, on the one hand, and from the other non-sceptical sources, on the other.

Third step: Selecting blogs

We used the three groups of representatives of, respectively, the sceptical, the mainstream and the anti-sceptical discourses as starting points for selecting the blogs to investigate further. There are 2,706 links from the blog posts in our corpus to the sceptical discourse, published on 861 different blogs. On the side of the mainstreamers, there are 20,383 blog posts with links to the eight non-sceptical sources, published by 4,338 different blogs. Thirdly, there are 836 links to the anti-sceptical source desmogblog.com, published by 457 different blogs. We will use these groups of blogs, blog posts and links to address RQ1 and RQ2.

Fourth step: Sampling linking blog posts for manual analysis

We manually coded two random samples of blog posts that linked to, respectively, the three sceptical sources and the eight non-sceptical sources. This should enable us to see whether the linking blog posts share the position of the sources to which they link. We performed the selection in two steps. First, we identified the blogs that have more than five links to at least one of the three groups of sources. This meant that we focused on bloggers who engage actively with the sources. From the posts published by the blogs with at least five links to either the sceptical sources or the anti-sceptical source

desmogblog.com, we randomly selected one hundred blog posts for manual analysis. We also analysed manually one hundred randomly selected posts from the blogs that have at least five links to the non-sceptical sources. This means that we coded posts from approximately 10 per cent of the blogs linking to either the sceptical or the anti-sceptical sources but that we coded only 4 per cent of the posts from the non-sceptical blogs.

Fifth step: Manual coding of the blog posts and a selection of links in them

We manually coded these 200 blog posts for their position on anthropogenic climate change. We followed this procedure: 1) use CtrF and search for the key terms "climate change" and/or "global warming" on the page; 2) determine the post's position on climate change by reading the paragraphs including the search terms. If the post is clearly critical of the consensus view on anthropogenic warming, classify it as a sceptic. If it discusses climate change without questioning the reality of it, classify it as a non-sceptic. Classify the blog post as undecided if an opinion is not discernible. The coding involved two coders – one coding the whole sample and the other coding 20 per cent for control. We obtained agreement of 95 per cent, which is good (Lombard et al., 2002). Only one blog post was classified as undetermined by one of the coders, but, after closer inspection and discussion, it was coded as a sceptic.

We also coded one link in each of the selected blog posts to determine whether there was correspondence between the positions of the linking post and the linked source. For each post, we selected the third link in the text of the post and followed it to its target. We coded whether the target shared the linking post's position on climate change, whether the text around the link expressed agreement with the target and what function the link had in the text of the linking post. For the coding of the function of the link in the text, we used a simplified version of a scheme developed for the coding of references in academic articles (Abu-Jbara et al., 2013). Our scheme uses these types to chart the functions that a link can have:

- Criticizing – positive or negative
- Comparison – contrast the author's work with that of another article
- Use – use of method or ideas from the cited article
- Substantiating – the results in the cited article are used to substantiate claims
- Basis – the cited article is used as a starting point or motivation
- Neutral (other) – neutral description of the work in the cited article (or does not fit into any of the categories above)

Again, two persons were involved in these coding tasks. In the first task, classifying the target as a sceptic or non-sceptic, we again obtained high agreement of 95 per cent among the coders. We also obtained high agreement, 90 per cent, on the task of determining whether the text of the target agrees with the linking blog post. However, it turned out to be much more difficult to use the scheme for classifying references in academic journals to blog posts about climate change. We often found it impossible to say whether the linking expresses positive criticism, the use of ideas, substantiation of a point of view or a basis for the argument. The agreement among the coders was less

than 50 per cent. However, if we made a crude three-way distinction between linking in the context of negative criticism of the target linked to, linking in the context of an endorsement of the view expressed by the target and neutral links, we obtained a much higher percentage of agreement (95 per cent).

Results

Note that the results reported below are tentative, since we coded a rather small number of items. In particular, we coded only a small set of links out of the very large number of links in the corpus.

As mentioned above, we sampled for manual analysis posts from blogs with at least five links to one of the three groups of sources. There were 158 blogs with more than five links to the sceptical sources, 58 blogs with more than five links to the anti-sceptical source and 728 blogs with more than five links to the non-sceptical sources. We observed that 52 (ca. 90 per cent) of the 58 blogs with more than five links to the anti-sceptics also have more than five links to the sceptics. In contrast, only 75 (ca. 10 per cent) of the 728 blogs with more than five links to the non-sceptics also have more than five links to the sceptics. There is also a large overlap between the anti-sceptics and the non-sceptics, as 51 (ca. 87 per cent) of the 58 blogs with more than five links to the anti-sceptics also have more than five links to the non-sceptics. This suggests that the anti-sceptic source has a role as a provider of input to the blogging on climate change that is different from that of both the sceptical and the non-sceptical sources. Our first two research questions are discussed below.

- RQ1: What is the probability that a link to a source that rejects the mainstream view on climate change comes from a blog that itself rejects the mainstream view?

Note that, since the set of blogs linking to the anti-sceptical source is almost completely included in the set of blogs that link to the sceptical sources, we found it legitimate to answer this first research question using a random sample of posts linking to the sceptical sources. Our manual coding of the 100 posts linking to either the two sceptical sources or the anti-sceptical source showed that 79 of these linking posts are sceptical while 13 are non-sceptical (eight links to the posts are inactive and we could not code them). Moreover, four of the 13 non-sceptical posts link to the anti-sceptical source desmogblog.com, while none of the sceptical posts link to this source. This suggests that, in this limited data set, there is a high probability that a link to one of the sceptical sources comes from a blog that is itself a sceptic.

- RQ2: What is the probability that a link to a source that endorses the mainstream view on climate change comes from a blog that itself endorses the mainstream view?

Here, our coding of the posts linking to one or more of the non-sceptical sources showed that all 84 of them are themselves non-sceptic (16 addresses were inactive and we could not code them). This suggests that there is a very high probability that a blog post from an active climate blogger to one or more of the highly visible non-sceptical sources is itself expressing a non-sceptical position. Despite the limited size of the samples, the patterns suggest that a link to either the sceptics or the non-sceptics is a signal of the position of the linking blog.

Let us now turn to the analysis of the function of the links from the blog post that we coded. We will first look at the *first group* of blog posts, that is, posts selected from blogs that have at least five links to either the sceptical sources or the anti-sceptical source. The coding found that there are 79 sceptical posts and 13 non-sceptical posts in this group. As explained above, in step five, we checked one out-link (the third) in each post and coded the position of the source that it links to and whether the linking post agrees with the views of the targeted source, criticizes it negatively or is neutral. Table 1 below shows the roles of the links in this first group:

Table 1. *The functions of links from the first group of blog posts (frequencies)*

Link from:	Criticizing non-sceptic	Criticizing sceptic	Endorsing non-sceptic	Endorsing sceptic	Neutral to non-sceptic	Neutral to sceptic
Sceptic post (79)	16	0	0	47	16	0
Non-sceptic post (13)	0	1	12	0	0	0

Note that the links charted here are not necessarily links to any of the highly visible, sceptical sources with which we started. Table 1 shows the function of the third link in each of the blog posts that we coded, as explained in step five. The table shows that more than half of the links from the sceptical blog posts are to another sceptic. Furthermore, all but one link from the non-sceptical blog posts are to another non-sceptic (four of the 12 endorsing links are to the anti-sceptic desmogblog.com). We also see that there are 32 links from sceptical blog posts to non-sceptical sources: 16 are criticizing the source and 16 are neutral to the source itself. The 16 neutral links from sceptics to non-sceptical sources are interesting. First, eight of these linked sources were unclassified, because they do not express a position on climate change. In Table 1, we group these together with eight sources that do have a position on climate change (we thereby obtain 16 neutral links to non-sceptics). Reading the blog posts in which the 16 links appear, however, we learned that they express dismissive criticisms of the consensus view and climate change. In these cases, even if the links are not part of criticism of the linked source, the neutral link is targeting the source to back up criticism of the non-sceptics. Table 1 also shows that in no cases does a sceptical blog post endorse the views of a non-sceptical blog post or criticize a sceptical blog post. The same holds true for the linking practices of the 13 non-sceptical blogs posts.

The differences in the distribution of the different types of links between blog posts and sources are obviously significant: a chi-square test of the distribution of, respectively, critical and endorsing links from the sceptical and non-sceptical blog posts to sceptical and non-sceptical sources emphasizes this point (X-squared = 76, p-value < 2.2e-16).

Table 2 below shows the results of the same analysis for the *second group* of blog posts, namely posts published by blogs with at least five links to the highly visible non-sceptical sources. The linking practices of the non-sceptical posts show that it is clear and simple:

Table 2. *The functions of links from the second group of blog posts (frequencies)*

Link from:	Criticizing non-sceptic	Criticizing sceptic	Endorsing non-sceptic	Endorsing sceptic	Neutral to non-sceptic	Neutral to sceptic
Sceptic posts (0)	0	0	0	0	0	0
Non-sceptic posts (81)	0	1	54	0	26	0

Again, the links charted in Table 2 are not links to the highly visible non-sceptical sources. We randomly chose 100 posts from the blogs that link to the non-sceptical sources. It turned out that 19 of the selected posts are inactive. Table 2 shows that half of the links express endorsement of another non-sceptic and that the only link to a sceptic is an expression of negative criticism of that target. The non-sceptics targeted with neutral links are sources that do not discuss climate change, the reading of the posts showed. The distribution of the links indicates that the tone of the discussion is less polemic than that of the sceptical blog posts.

We concluded above that there is a high probability that a blog that frequently links to one of the groups of highly visible sources shares the position of that group. The results above suggest a more general conclusion that active climate bloggers mostly link to sources with which they agree and, if they link to sources with which they disagree, the point is to criticize. The number of data points to back up this conclusion is small, but the tendency is clear enough to make the claim plausible.

We are now in a position to address our third research question:

- RQ3: What are the functions of the links that cross the ideological divide? Are they mainly formed in the context of criticism of the other position or as part of a dialogue?

The answer that we can provide based on the samples that we have analysed is clear: the links that cross the divide between the positions show few signs of a dialogue. The links from a blog post to a source holding a position different from that of the linking post are all critical of the linked source, and there is no negative criticism of a source that shares the position of the linking post.

Let us turn to the fourth research question:

- RQ4: What are the functions of links from blogs to sources that share the blog's position on climate change?

It turned out to be difficult to answer this question. Table 2 shows that most of the links are to sources that share the position of the linking blog post: there are 47 links from sceptic blog posts to sceptic sources and 90 links from non-sceptic blog posts to the non-sceptic source. Our conclusion after attempting to code the function of these links with the scheme described above (see the seventh step in our analytical strategy) was that the bloggers do not use links in the same way as references in academic articles are used. The coders found it particularly difficult to decide whether the function of a given link is to substantiate an argument, to indicate the use of ideas from the source, to provide access to the basis of the argument or to criticize the source positively. However, if we collapsed these functions (substantiation, use, basis and positive criticism) into something like "endorsement", we were able to agree completely on which links from a blog post endorse a source.

The second factor that made the comparison with referencing in academic articles problematic was that some blog posts criticize representatives of the other position without linking to the target of the criticism. This is true in particular of the sceptical blog posts. All of the sceptical blog posts include criticism of the consensus view on climate change, but only a few links target representatives of that view. Most of the links are to sources that back up the criticism. We found the same tendency with some of the non-sceptical blog posts but not to the same extent.

The third feature that made it difficult to compare links to academic references is that some blog posts do not contribute anything new but only republish the entire text from another blog. We saw, for example, that some of the sceptical blogs republish material originally published by WattsUpWithThat.com, the most visible critic of mainstream climate science. This practice is part of a role that many blogs choose to play as *amplifiers* of information produced by others (Bruns, 2005; Elgesem et al., 2015).

Discussion

Let us return to our overarching research question:

- To what extent are the patterns in the linking from blogs about climate change signs of polarization between blogs that accept the consensus view on climate change and those that reject that view?

Above we described three theories about why polarization develops: 1) a limited pool of arguments, 2) confrontation by challenging arguments and 3) the possibility that opposing another group strengthens group identity. In the analysis, we saw some evidence of all three mechanisms. The analysis above showed that there is a clear tendency for sceptical bloggers to link to the sceptical discourse on climate change and non-sceptic bloggers to link to the mainstream discourse. We thus obtained evidence of both groups accessing and sharing content with which they agree, and, if they engage with sources representing the other side, this occurs in the context of negative criticism of the other side. These patterns are signs of polarization. However, the linking to the anti-sceptical source desmogblog.com seems to follow a different pattern, as an equal proportion of links to this anti-sceptical source comes from blogs that also follow the sceptical sources and blogs that follow the non-sceptical sources.

We also found that the bloggers engage with sources on the other side not to have a dialogue but, instead, to dismiss it. This suggests the second mechanism behind polarization, that is, that people become more entrenched in their own position because they rehearse arguments in its favour when confronted by views that challenge theirs. Again, it is plausible to suggest that this practice is a sign of polarization.

The third explanation for why polarization develops is that it strengthens group identity. The deniers have an identity as a group because they define themselves as the opposition to the mainstream view on climate change. We see clear signs of this with the sceptical blogs, in which all of the posts are dedicated to the rejection of the consensus view. The non-sceptics, in contrast, are mostly concerned with other climate-related issues than the sceptical opposition. We thus seem to find evidence of what we called asymmetric polarization above: the sceptics hold the extreme position and they mainly drive the polarization.

Conclusion

The aim of this article was to undertake an exploratory study of how and the extent to which the patterns in links from climate change blogs carry information about polarization. Our conclusion is that, given the limited number of data points, there is evidence of polarization in the patterns of links from the blogs that actively engage with the topic of climate change.

Furthermore, our analysis suggests that blogs about climate change do not use links in the same way as references in academic articles. The explanation for this difference between blogs and academic articles is, we suggest, that blog posts are typically opinion pieces and not written to contribute original knowledge to a community of informed peers. In a polarized context like the debate on climate change, the posts typically try to convince their readers that only one position makes sense – the one that the blogger holds. This is not surprising, perhaps, but it should make us aware that we have to be careful about making assumptions about the role of links in blogging.

The third conclusion is that, to understand the role of the links in the discussion on climate change, we need to understand the function of the links in the texts: the structure of the graph of hyperlinks is not sufficient. Our study suggests that analyses of the ideological affiliations of linked blog posts can provide insights into the mechanisms of polarization at the micro level. In combination with text analysis, links can therefore be very valuable sources of information about the ideological landscape of online debates. To determine the ideological function of links, it is however necessary not only to identify the ideological profile of the linking and the linked blog posts but also to classify the role of the link in the embedding text. This is challenging to perform reliably with automated methods and big data.

The study of course has limitations. One of them is the relatively small number of the randomly selected blog posts and links that we coded manually. In addition, in the analysis above, we coded the position of the blogs that link to the two groups of sources, but we did not code the function of the links from these blogs to the two groups of highly visible sources. It could have strengthened the argument if we could have coded the function of the links from the selected blogs to the two groups of, respectively, sceptical and non-sceptical sources.

In summary, since the tendency in the linking patterns and the functions of the links is so clear, we believe that our argument draws a suggestive picture despite these limitations.

Notes

1. https://en.wikipedia.org/wiki/Blog
2. Using in particular the R package dplyr.
3. "The goal of Skeptical Science is to explain what peer reviewed science has to say about global warming. When you peruse the many arguments of global warming skeptics, a pattern emerges. Skeptic arguments tend to focus on narrow pieces of the puzzle while neglecting the broader picture." From the "About" section of https://www.skepticalscience.com/about.shtml.
4. "The DeSmogBlog Project began in January 2006 and quickly became the world's number one source for accurate, fact based information regarding global warming misinformation campaigns". From the "About" section of https://www.desmogblog.com/about.

References

Abu-Jbara, A., Ezra, J. & Radev, D. (2013). Purpose and polarity of citation: Towards NLP-based bibliometrics. In *Proceedings of NAACL-HLT* (pp. 596-606).

Adamic, L. (2008). The social link. In J. Turow & L. Tsui (eds.). *The hyperlinked society* (pp. 227-248). Ann Arbor, MI: University of Michigan Press.

Adamic, L. & Glance, N. (2005). The political blogosphere and the 2004 election: Divided they blog. In *Proceedings of the 3rd international workshop on link discovery, KDD 2005* (pp. 36-43). Chicago, Illinois: ACM

Bakshy, E., Messing, E., Adamic, L. et al. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348: 1130-1132.

Bruns, A. (eds.) (2005). *Gatewatching. Collaborative online news production*. New York: Peter Lang.

Dunlap, R. E. & McCright, M. J. (2011). Organized climate change denial (pp. 144-160). In J. S. Dryzek, R. B. Norgaard & D. Schlosberg (eds.), *The Oxford handbook of climate change and society*. New York: Oxford University Press.

De Mayer, J. (2013). Towards a hyperlinked society: A critical review of link studies. *New Media and Society*, 15(5): 737-751.

Elgesem, D. et al. (2015). Structure and content of the discourse on climate change in the blogosphere: The big picture. *Environmental Communication*, 9(2): 169-188.

Elgesem, D. (2017). Polarization in blogging about the Paris meeting on climate change. In G.L. Ciampaglia et al. (eds.) SocInfo 2017, Part I, Lecture Notes in Computer Science (NCS 10539). 1-23. New York: Springer.

Fiske, S. T. & Taylor, S. E. (2013). *Social cognition – From brains to culture*. Los Angeles: SAGE.

Guera, P. H. C. et al. (2013). A measure of polarization on social media networks based on community boundaries. In *Proceedings of the seventh international AAAI conference on weblogs and social media*. Retrieved from http://www.aaai.org/Library/ICWSM/icwsm13contents.php [Accessed 2018, March 28].

Himelboim, I., McCreery, S. & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. *Joutnal of computer-mediated communication*, 18(2): 40-60.

Hulme, M. (2009). *Why we disagree about climate change*. Cambridge: Cambridge University Press.

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of personality and Social Psychology*, 50: 1141-1151.

Lodge, M. & Taber, C. S. (2013). *The rationalizing voter*. Cambridge: Cambridge University Press.

Lombard, M. et al. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, 28: 587-604.

Mann, M. E. (2012). *The hockey stick and the climate wars*. New York: Columbia University Press.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49: 65-85.

Sharman, A. (2014). Mapping the climate sceptical blogosphere. *Global environment change*, 26: 159-170.

Sinclair, B. (2012). *The social citizen: Peer networks and political behavior*. Chicago: Chicago University Press.

Sunstein, C. (2017). *#republic*. Princeton: Princeton University Press.

Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology*, 57(1): 1-147.

Turner, J. C. et al. (1987). *Rediscovering the social group: A self-categorization theory*. New York: Basil Blackwell.

Yardi, S. & Boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5): 316-327.