ARTICLE

Credibility of climate change denial in social media

Abhishek Samantray [ID] [1]* & Paolo Pin [2,3]

ABSTRACT   Public perception about the reality of climate change has remained polarized and propagation of fake information on social media can be a potential cause. Homophily in communication, the tendency of people to communicate with others having similar beliefs, is understood to lead to the formation of echo chambers which reinforce individual beliefs and fuel further increase in polarization. Quite surprisingly, in an empirical analysis of the effect of homophily in communication on the level of polarization using evidence from Twitter conversations on the climate change topic during 2007–2017, we find that evolution of homophily over time negatively affects the evolution of polarization in the long run. Among various information about climate change to which people are exposed to, they are more likely to be influenced by information that have higher credibility. Therefore, we study a model of polarization of beliefs in social networks that accounts for credibility of propagating information in addition to homophily in communication. We find that polarization can not increase with increase in homophily in communication unless information propagating fake beliefs has minimal credibility. We therefore infer from the empirical results that anti-climate change tweets are largely not credible.

[1] IMT School for Advanced Studies Lucca, Lucca, Italy. [2] Department of Economics and Statistics, University of Siena, Siena, Italy. [3] BIDSA and IGIER, Bocconi University, Milan, Italy. *email: abhishek.samantray@imtlucca.it

Introduction: polarization of beliefs, homophily in communication, information credibility

Polarization of beliefs is about the existence of opposing beliefs within large sections of the society (Bail et al., 2018; DiMaggio et al., 1996; Baldassarri and Gelman, 2008; Prior, 2013). In situations, like climate change action, where unanimous belief can drive the required collective steps, polarization can be a hurdle and may lead to socially undesired actions (Iyengar and Westwood, 2015; Iyengar et al., 2012; McCright and Dunlap, 2010; Center, 2014). Public understanding of climate change has remained polarized (Moore et al., 2019; McCright et al., 2014; Hamilton, 2011). Polarization of beliefs can be affected not only by the nature of users' tendencies like homophily in communication about the reality of climate change, but also by certain nature of the information itself. Credibility of information is one of such properties of information. We discuss these determinants of polarization below.

Homophily in communication can affect the polarization of beliefs. Homophily in communication is a key tendency of users in information propagation media like Twitter and other platforms: people tend to communicate with others who hold similar beliefs (Barberá, 2015; Conover et al., 2011). Although homophily can facilitate the flow of information (Jackson and López-Pintado, 2013), it can lead to polarization (Bessi et al., 2016) as well and decrease the general ability of the society to learn the truth (Taylor et al., 2018; Golub and Jackson, 2012). Communicating only with people who share similar ideology or opinion restricts beliefs and prevents learning the truth (Madsen et al., 2018; Acemoglu et al., 2011, 2010). For example, homophily among liberals and conservatives in political blogs links (Adamic and Glance, 2005), i.e., the tendency of liberal and conservative blogs to link primarily within their separate communities, leads to echo chambers (Jasny et al., 2015; Garrett, 2009; Flaxman et al., 2016), and hence beliefs can be polarized due to such exposure to selective information (Bakshy et al., 2015). Similarly, homophily in opinion exchanges in social media (e.g., via publicly visible replies and mentions in Twitter) can reinforce beliefs within various sections of the network due to selective exposure. Previously echo chambers have been observed for climate change discussions on social media (Williams et al., 2015; McCright and Dunlap, 2010). Such patterns when repeat themselves in various parts of the network can lead to polarization of beliefs.

Credibility of information is also an important factor to create polarization, especially in online media where usually there exists information from several sources (Greer, 2003) that propagate controversial beliefs. Credibility of information is the precision of information, it signifies how certain the information is and helps to assign a certain level of trust to the information (Kiousis, 2001; Hovland and Weiss, 1951). The credibility of information that propagates in a social network is a critical factor that can shape the beliefs: if incoming information from a person's social network carries no credibility, then it is less likely to be incorporated in to the belief of the person (Pornpitakpan, 2004; Heesacker et al., 1983). Hence, negative consequences that may arise due to the spread of fake information in social networks depend on information credibility, thereby making it a factor that can induce polarization. The importance of precision of misinforming signals has also been highlighted recently by Allcott and Gentzkow (2017). Credibility is a dimension of information which is independent of veracity of the information. For instance, a particular (unintended) fake story can be more credible if the information source is highly reputable (and hence the fake story inherits the credibility of the source) compared to the case when the same story comes from a less credible information source. It is the former case that has the potential to change beliefs in the society and change the existing levels of polarization concerning the truth

of its story. Credibility of information in social media, generally speaking, may be ascribed due to several factors including reputation of information source, number of verified facts cited along with the information, mentioning opinion leaders and others (Markham, 2006; Westerman et al., 2014). Lastly, credibility of a communicated information is independent from the preciseness with which the information source believes the information. For example, a certain propaganda house which believes strongly about a story need not be able to spread such beliefs in the society through various communications because such communications may lack the degree of credibility required to generate substantial change in beliefs in society.

Twitter has become a modern platform for news dissemination and opinion exchanges, and is widely adopted by many users worldwide. We believe discussions and information propagation that happens on such a platform has potential to shape beliefs at a large scale. Important topics like climate change are also discussed on such social media platforms. With fake information on many topics becoming prevalent on widely adopted social media platform like Twitter, it is probably not surprising that there are several tweets both in favor and against the statement that climate change is a real concern. Fake information regarding such an important issue as climate change can pose a collective hurdle for the society at large if such information becomes highly credible among users of the platform. In this study, we infer the credibility of anti-climate change opinions on Twitter using (i) an empirical analysis that investigates the effect of homophily in communication patterns on the polarization of beliefs, and (ii) the predictions of a model of polarization of beliefs that jointly accounts for the roles of information credibility and homophily in communication. For the empirical analysis, we use tweets about climate change topic during 2007–2017. We rely on opinion exchanges among climate change believers and skeptics made via mentioning (and replying) others in tweets as the communication pattern. (Replies in Twitter tag the user names at the beginning of tweets, while user mentions can be made anywhere within the text of tweets). This is because retweets do not contain new opinions and are mere repetitions or broadcast of opinions expressed in original tweets, whereas mentions contain explicit referencing of other users and hence convey exchange of opinions targeted towards users being mentioned.

Methods

Data. We use tweets from the online social network site Twitter. Tweets are the messages posted by users on the platform. Tweets can contain hashtags, mentions to other users, external links in addition to text. A tweet posted by a user can be retweeted, replied and liked by other users. The tweets from 2007 to 2017 were collected based on a search filter that each tweet contains at least one of the following words: 'climate change', '#climatechange', 'global warming', '#globalwarming'. The search was performed via Tweeter's public advanced search page. This collection of tweets does not violate any ethical standards and consists of only publicly available data. The data is complete in the sense that it contains all original tweets that fulfill the search criteria. Retweets of original tweets are not included in the data. Each tweet however contains the retweet statistic, i.e., how many times it has been retweeted. There are a total of 14,353,859 unique tweets (without counting retweets values) and 3,595,205 unique users in the dataset.

Measuring sentiment and opinion. The sentiment of each tweet is computed using VADER (Hutto and Gilbert, 2014) model

which is designed to conduct sentiment classification specifically on short texts like tweets. For each tweet, a score is obtained on the scale -1 (most negative) to $+1$ (most positive). A message with a positive sentiment is usually in favor of the (intended) object in the message. However, the same message may be communicated in a way that contains negative sentiment. Therefore sentiment and opinion contained in a message are two different characteristics of the message. This distinction between sentiment and stance has been clarified by previous studies on stance detection of tweets (Sobhani et al., 2016; Mohammad et al., 2017). From various data used in these previous studies, we use a subset that is used to annotate whether the tweet is in favor or against the statement "climate change is a real concern". This subset contains manual annotation of whether a tweet is in favor or against the statement. We use this data as the training data to predict the opinion expressed in the texts of tweets in our sample.

The tweets were coded as numerical features using the TF-IDF (term frequency-inverse document frequency) representation and the prediction on the dataset for this study was done using the support vector classifier, which had the highest predictive power among other classifiers (logistic regression, decision tree). Each user's tweets were first classified into one of the following categories: in favor of the statement, against the statement, no opinion. (Hence, in the first stage, opinion is assigned to each tweet). Next, each user was classified into one of the above categories based on the category that contains the maximum number of tweets from the user. Since a user on Twitter usually has a belief about climate change (in general), the classification algorithm has managed to predict each users's tweet into a single category only. Only 42 users were classified as having no opinion, and were dropped from the dataset. Most users had tweets in one category only: either in favor, or against the statement. This validates the performance of the classifier at the user level even if the prediction accuracy (at the tweets level) is about 72%.

Measure of homophily. The notion of homophily in communication is the presence of higher interaction among people who hold the same opinion about climate change. For instance, the group of users who believe climate change is not real have a homophily in communication pattern if such users communicate more among themselves than with users who believe climate change is real.

Communication between users can take the forms of retweeting and mentioning (includes replying) on the Twitter platform. Although retweeting, in general, suggests that users share the same ideology, mentioning can occur between users with differing ideologies (tweet wars). To what extent do mentioning patterns in tweets reveal communication among users having similar ideologies? Previous studies (Lietz et al., 2014; Aragón et al., 2013) reveal that homophily of the aggregate network measured using retweeting and mentioning tend to be same, although mentioning is more volatile at the group (sub network with a particular ideology) levels. Hence, based on such studies, we assume that homophily in mentioning patterns reveals the homophily in retweeting (and following) patterns, for which we do not have the data. Communication therefore refers to users interacting via mentioning each other in tweets.

To measure homophily in communication among users with same belief or opinion regarding climate change, we follow the measures used in previous empirical studies on homophily (Currarini et al., 2009; Halberstam and Knight, 2016; Colleoni et al., 2014). Let I be the total number of users in a given month who either mentioned other users or received mentions from other users' tweets. We indicate belief of a user as $b \in \{a, f\}$, where a and f correspond to the cases where the user is against and in

favor, respectively, of the climate change statement. a and f represent two types of belief b. Suppose I_b is the total number of type b users, then $w_b = \frac{I_b}{I}$ is the fraction of type b users. Let v_{ib} be the number of type b users mentioned by user i. Then $s_b = \frac{1}{I_b} \sum_{i \in I_b} v_{i,b}$ is the average number of same type users mentioned by type b users, and $d_b = \frac{1}{I_b} \sum_{i \in I_b} v_{i,-b}$ is the average number of opposite type users mentioned by type b users. Using this, homophily of the group of type b users is defined as $H_b = \frac{s_b}{s_b + d_b}$ and the homophily of the society is defined as

$$H = \sum_{b \in \{a, f\}} w_b H_b.$$

Measure of polarization. A tweet carries an opinion (whether in favor or against climate change being real), denoted as op, with a certain sentiment, denoted as s. (We encode op as 1 if the message in the tweet is against the climate change statement and as -1 if the message is in favor of climate change). Irrespective of the opinion, the sentiment can be positive or negative depending on the way the message is communicated. So we use an emotion-adjusted measure of belief called EAB that combines these two aspects: expressed opinion in the message, and emotional content in the message. For a given tweet with attributes op and s, EAB is defined as $op \cdot |s|$, the product of opinion and absolute value of sentiment.

A large number of tweets have neutral or close to neutral sentiment, thereby increasing the mass around 0 and decreasing the intensity of bimodality in the distribution of EAB. (The presence of high volume of neutral tweets is not uncommon, for example see Kušen and Strembeck (Kušen and Strembeck, 2018). In this sense, the polarization indicator used during the mathematical analysis in opinion updating model (see Supplementary Information) cannot be carried to the empirical setting directly. We therefore need another measure of polarization that is sensitive to the properties (e.g., kurtosis and skewness) of a distribution and carries the intuition of bi-modality approach (i.e., the statistic should be sensitive to the degree of bi-modality of a distribution). To calculate the degree of polarization involved in the EAB distribution, we use the measure of ideological divergence provided by Lelkes (2016), which characterizes the level of polarization based on bi-modality of the distribution by being sensitive to kurtosis and skewness (Freeman and Dale, 2013; Pfister et al., 2013). Using this measure, the polarization of EAB is defined as

$$\frac{s^2 + 1}{k + 3 \frac{(n-1)^2}{(n-2)(n-3)}},$$

where s and k are the skewness and excess kurtosis of the EAB distribution, and n is the sample size. The values of 1 and 0 correspond to cases when the EAB distribution is perfectly bimodal and perfectly unimodal, respectively. A value greater than 0.56 is categorized as bimodal. Although this threshold may not be reached precisely, approaching this value is considered coming closer to polarization (Lelkes, 2016).

The measure of polarization above is calculated at the levels of tweets and is a stricter measure than when calculated at the level of users. This is because a user might be slightly heterogeneous in his tweets but aggregating all tweets would reveal one exact ideology. For example, when a climate change skeptic user writes a tweet, it is most likely to be against the statement that climate change is real. However a few tweets of his might fall in favor of the statement. In this sense, polarization is easier to obtain at an user level than at the content (tweets) level.

It is easy to see that the polarization measure describes a different phenomenon than the homophily measure. While homophily measure is constructed using users involved in a mention activity in a given month (micro level), polarization is measured using all the EAB of all tweets in the month (macro level). This is in line with Esteban and Ray (1994), according to whom any reasonable measure of polarization must be global in nature.

Results

Negative effect of homophily on polarization. The monthly time series of polarization of beliefs and homophily in communication during 2007–2017 are shown in Fig. 1. (Measures of homophily and polarization are described in Methods section). It appears that the evolutions of the two curves are not independent, and polarization tends to decrease at times when homophily increases. Augmented Dickey-Fuller test and Phillips-Ouliaris test suggest that polarization and homophily are cointegrated, i.e., they are bound in a long term equilibrium relationship such that they mean-revert whenever there are deviations away from the equilibrium. (Supplementary Information contains statistical details of tests of cointegration). Hence it becomes natural to model polarization and homophily jointly in a vector error correction (VEC) model (Granger and Weiss, 1983; Engle and Granger, 1987). VEC models are a special class of models derived from vector autoregression (VAR) models with an additional term, called error correction term, to account for the cointegration. VEC model allows to study the lagged effects of homophily on polarization and also the lagged effects of polarization on homophily. Supplementary Table 4 contains details of the parameter estimates of VEC model: the lagged homophily covariates negatively affect polarization in the long term with high statistical significance and the lagged effects of polarization on homophily are not significant. These estimates however do not form a conclusion about the direction of causality between homophily and polarization. For assessing causality, we perform Granger causality test (Granger, 1969) with modifications, as suggested by Toda and Yamamoto (1995), to adapt to the non-stationary nature of both time series; the results are shown in Table 1. (Further statistical details regarding these tests are available in Supplementary Information). The conclusion that emerges is that only homophily Granger-causes polarization with a negative effect, and the causality in the other direction is absent, i.e., polarization does not Granger cause homophily. This empirical result that homophily negatively affects polarization is counter-intuitive to the discussions made previously about the nature of homophily and polarization.
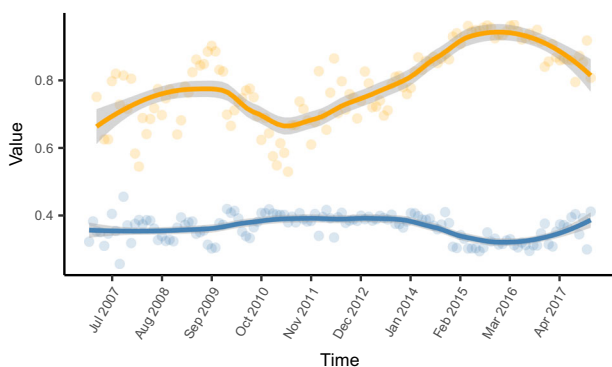
Effect of homophily on polarization depends on information credibility. We study a simple model of emergence of polarization in social networks that highlights the joint effect of credibility of propagating information and homophily in communication of information to affect polarization. We model polarization as bi-modality of the population's distribution of beliefs and it captures the following basic features, as laid down by Esteban and Ray (1994), which are necessary for a distribution to be considered polarized: there must be a high degree of homogeneity within each group (whose agents hold same belief), a high degree of heterogeneity across groups, and a small number of significantly sized groups. If the groups are of insignificant size (e.g., isolated individuals), they do not contribute to polarization. The beliefs can be emotion-adjusted as well, as is done in the empirical analysis (see Methods section), to include the emotional content of the belief; this does not change the nature of predictions of our model.

In our model, we consider a social network with each agent receiving information about a topic (e.g., the reality of climate change) from the same number, k, of speakers (for simplicity). Speakers of an agent are other agents in the network from whom he receives information. An unobserved fundamental θ, with values in the real line, describes the accuracy of the beliefs: an agent with a higher value is more distant from the truth. At a given time t, we assume there are two types of agents—θ_l, θ_r— informed and misinformed, and the social network has a homophily coefficient of $\frac{h}{k}$ with respect to this property: h speakers of an agent of type θ_i, $i \in \{l, r\}$, are of the same θ_i type. We model the prior beliefs of informed (θ_l) and misinformed (θ_r) agents as distributions of θ over the real line to allow for heterogeneous prior beliefs at time t. We assume that the prior beliefs of informed and misinformed agents derive from distributions $\mathcal{N}(\theta_0, \delta_l^{-1})$ and $\mathcal{N}(\theta_0 + \xi, \delta_r^{-1})$, respectively, with a strictly positive value for ξ, in order to ensure that the misinformed type are, on average, farther away from truth compared to the informed type. The belief of the entire population of the social network at time t, a density-weighted average of the above distributions, is considered to be polarized if it has two modes (peaks in the distribution).

We assume each agent communicates a particular realization of his belief using a message (e.g., via a tweet) during the period between t and a future time $t + 1$, and all agents update their prior beliefs at $t + 1$ after incorporating beliefs about the fundamental expressed in their speakers' messages. Mathematically, a message from a speaker is modeled as an independent noisy signal about the speaker's realized belief, with the degree of the noise (or uncertainty) being conditional on the type of speaker. Since credibility of a communicated message is the precision of belief expressed in the message, in our model, true and fake information about the reality, arising from informed and misinformed speakers, respectively, propagate with different credibilities in the social network. (We denote the credibilities or precisions of true and fake information as β_l and β_r, respectively). We assume that the messages from informed agents carry a positive credibility. (In our context of the reality about climate change, this is a reasonable assumption). It is noteworthy

Fig. 1 Polarization of beliefs, homophily in communication. The figure shows monthly evolutions of polarization of beliefs about climate change (blue) and homophily in communication patterns on Twitter (orange) during 2007–2017

Table 1 Granger-causality tests

Null hypothesis	Test statistic
Homophily does not Granger-cause polarization	12.3***
Polarization does not Granger cause homophily	4.0

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. Based on the test statistics, only the first null hypothesis is rejected
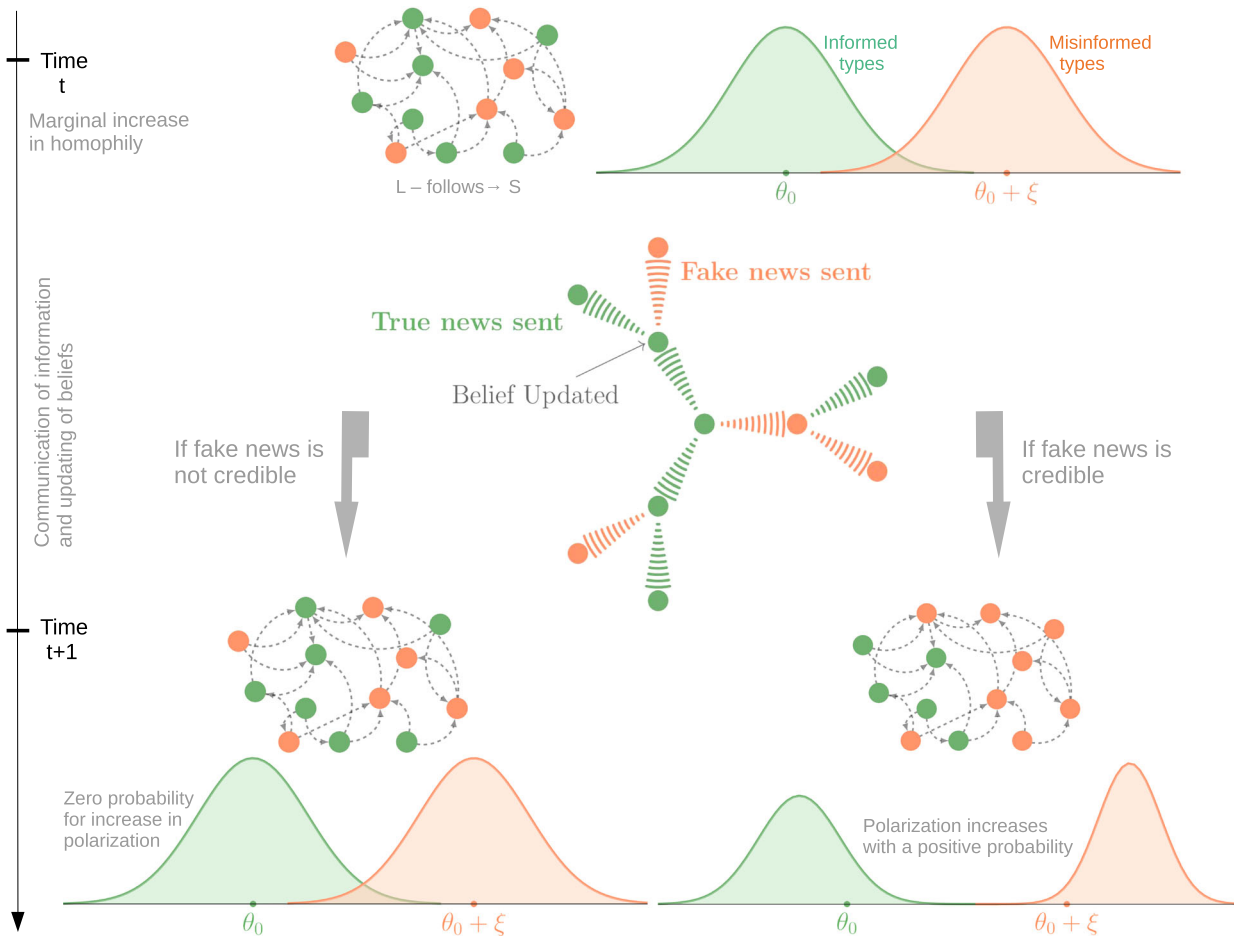
Fig. 2 Belief updating model. The figure illustrates increase in polarization at time $t+1$ due to increase in homophily (based on beliefs, e.g., reality of climate change) at time t. There is zero probability for such an increase if fake information that propagates in the social network has zero credibility

to mention that for a speaker of a given type θ_i, $i \in \{l, r\}$, β_i is conceptually different from δ_i: while δ_i characterizes the probability with which a particular random realization will be selected as the belief for the speaker, β_i characterizes how precisely the realized belief is communicated in a message. In a Bayesian update of beliefs, how a communicated message affects the beliefs of listeners of the message depends on the belief expressed in the communicated message, the precision (or credibility) of the message, prior beliefs held by individual listeners and the precision of such prior beliefs of listeners.

We now present the findings of some analyses conducted using the model discussed above. (Mathematical proofs of the results presented below are available in Supplementary Information). We find that in a random communication network with agents being sufficiently uncertain about the prior beliefs they hold, polarization can not arise at time $t+1$, after agents have incorporated their speakers' beliefs. This means that, irrespective of the precisions with which true and fake information diffuse in the social network, they do not play a role in the emergence of polarization without the support of non-random topology of the social network and agents' prior beliefs. We also find that if marginal increase in homophily at time t does not increase the probability of polarization at $t+1$, it can happen only in the case when the fake signals that propagate in the social network do not have a minimal level of credibility and carry zero precision. A visual illustration of this result is shown in Fig. 2. We believe this provides a potential reason for our previous observation that, in the climate change discussions on Twitter, polarization of beliefs

does not increase when homophily in communication patterns increases.

Discussion

The online social network Twitter has remained an important media for rapid spread of opinions. We studied the opinions expressed in Twitter during 2007–2017 regarding the reality of climate change. Below, we briefly discuss our main findings and their limitations, and suggest directions for future research.

The analysis over a long time period provides insight about the direction of potential causality between homophily and polarization which would otherwise not have been possible in cross-sectional observations. In social networks, polarization of beliefs (existence of large groups of people with opposing beliefs) and homophily in communication (communication among people having same beliefs) tend to be highly correlated. One potential mechanism for such a correlation is that increase in homophily can reinforce individual beliefs leading to the creation of echo chambers and hence increase polarization. Another mechanism for such a correlation is that increase in polarization increases segregation of a society into different beliefs thereby acting as a natural source to increase the probability of like-minded communication (homophily). In the case of Twittersphere of climate change conversations, we performed Granger causality tests on the evolutions of homophily and polarization, and found that only homophily Granger-causes polarization and not vice-versa, i.e., increasing polarization does not lead to increase in

like-minded communication. It is important to note that Granger causation of homophily on polarization does not fully establish its causality. Granger causality (as explained in Supplementary Information) is a concept about precedence of the cause (homophily) before the effect (polarization). It says that the evolution of homophily significantly helps to predict evolution of polarization in future. Existence of Granger causality therefore does not exclude the possibility of an unobserved variable to drive the evolution of both homophily and polarization. Although the presence of such an unobserved variable is highly unlikely within the framework of our model and the variables used for empirical analysis, we encourage future research to build upon the above results to improve the causal nature of the effect of homophily on polarization—this would require analyzing scenarios where homophily can be argued to be exogenous so that we are better ensured about effects not being driven by unobserved variables. (Such causal relationship may be investigated for topics other than climate change as well).

We also found that the effect of homophily on polarization is negative, i.e., increase in homophily in communication leads to decrease in polarization of beliefs in future. This is counter-intuitive because we would expect an increase in polarization when homophily increases. Increasing homophily leads to situations where people are exposed to a narrow set of beliefs (Acemoglu et al., 2011, 2010) that conforms to their existing beliefs. When such homophily in communication happens in two large sections of the society with differing beliefs, it enhances polarization (Esteban and Ray, 1994).

Polarization of beliefs can be affected not only by homophily in communication among people, but also by the credibility of information that propagates on Twitter. In this study, we investigated whether credibility of information source plays a role to increase polarization. We studied the 'credibility' factor because this has received less attention in the literature and is a very intuitive factor. It is intuitive since information from a source which is not credible is naturally least likely to affect or change the belief of an individual. Credibility is a dimension of information independent of the veracity of information. For example, let us consider a certain information (either true or fake) that comes from two different sources at the same time: one source is a Twitter account of a person who is not very well known, and another is the Twitter account of BBC News (as instance) which has a much larger credibility in the society. In this case the news sent by BBC News has a higher probability to influence the beliefs of people at a large scale. More so, if the news happens to be (intentionally or unintentionally) fake, it can possibly affect the beliefs of a certain section of the society in a wrong direction (since the news is fake) thereby increasing polarization in the society because a new section of the society emerges with beliefs much different from what the entire population believed previously.

We modeled these two determinants of polarization discussed above, homophily in communication and credibility of information, jointly in a belief updating model where agents in a networked society receive true and fake information from their neighbors. We modeled the credibility of each type of information using the precision or certainty of the information. The model predicts that marginal increase in homophily always leads to increase in polarization expect for the only case when fake information has no credibility. In the case when fake information is not credible, the model predicts a negative effect of homophily on polarization. (This description is illustrated in Fig. 2). Since we observed a negative effect of homophily on polarization in the empirical analysis of climate change discussions on the Twitter platform, we conclude that tweets expressing anti-climate change beliefs are largely not credible to the broader society.

Our results disentangle the *presence* of fake information on social media from its potential *negative effect* on the society. The spread of fake information (either as misinformation or disinformation), as has been prevalent during recent years in online social media (Vosoughi et al., 2018; Del Vicario et al., 2016), has the potential to pose harm to the society by polarizing the beliefs of people (Vosoughi et al., 2018; Del Vicario et al., 2016; Bessi et al., 2016); for instance, it can influence political election outcomes (Shin et al., 2017). However, does it mean that fake information always has a negative effect on the society? Based on our results of this study, we can say that it is not always the case. In the case of reality about climate change, although there are many climate-skeptic messages, we showed that such messages are largely not credible and hence polarization (a negative consequence for society) does not increase when users on the platform communicate among themselves. Hence, in general, we believe that the presence of fake information (on various topics including climate change) in social media and the web is not a conclusive evidence of its negative effect on the society at large. This also reaffirms the fact that the negative effects of echo chambers in social networks, which are known to arise due to high levels of homophily in communication, might be overstated (Warner, 2010; Bail et al., 2018; Flaxman et al., 2016; Dubois and Blank, 2018).

We now discuss some assumptions about agents in our model, and how alternative assumptions about human behavior (in real world) may also explain how increase in homophily can lead to decrease in polarization in future. Agents in the model are rational (Madsen et al., 2018) and purely rely on previously held beliefs and beliefs expressed by neighbors in their immediate social networks to arrive at new beliefs. Such assumption of Bayesian updating of beliefs dictates that when homophily increases (i.e., agents' exposure to non-conforming beliefs decreases, or in other words, cross-attitudinal interactions decreases), the ability of misinformed agents to move closer to truth decreases. Human behavior, in general, is highly heterogeneous and there may exist different ways in which people update beliefs in real world. Let us consider a different situation where the utility of agents depends not only on the belief of own type but also on the (opposite) belief held by the other type. In particular consider a situation where people update beliefs in the following manner—when people are exposed to cross ideologies, instead of gaining higher utility by incorporating and averaging it with own ideologies, they gain higher utility by the fact that they know something (what they think is supposed to be known) and the opposite party is misguided and carries the wrong belief. In such a case, if homophily increases, cross-attitudinal interactions decrease and the misinformed agents are less convinced about their belief because their exposure to informed type (opposite type) agents decreases. This leads to dilution of ones's held beliefs and creates a decrease in polarization. Although such a behavioral assumption can lead to an alternative way to explain negative effect of homophily on polarization, we believe that such assumption may not scale to a wider population. (Especially in the case of reality about climate change, we believe that informed people would also want ill-informed sections of the society to know about the reality as they do). In this context, we believe that by imposing a Bayesian updating criteria we have studied the expected outcomes in a baseline model, and is (ideally) representative of the behavior of a larger population. We consider the model we analyzed in this study to be parsimonious enough to be able to explain the negative effect of homophily on polarization by highlighting the role of information credibility. Future research, depending on the nature of investigation, may find our analysis as a useful guide.

Individual rumors are being debunked in due time by fact checking measures (Zhao et al., 2015; Ciampaglia et al., 2015), however, public perception about some issues like human induced (anthropogenic) climate change have remained controversial for a long time (Tvinnereim et al., 2017). Such controversy appears to persist even after several investigations made by the scientific community. Various factors like exposure to different kind of information on the media (Iyengar and Massey, 2019; Williams, 2016), politicization of climate change (McCright and Dunlap, 2011), and exposure to fake information on social media platforms are potential reasons that can contribute to polarization of beliefs about the reality of climate change. In this study, we showed that social media messages on platform like Twitter is not a potential concern because although there are many sources propagating fake beliefs regarding climate change, the collective credibility of such sources is negligible. Future research may contribute to improve our understanding about how public perceives the reality about climate change. It is better that the society is not polarized on beliefs about the reality of climate change, so that timely environmental policies can be implemented with least public resistance. Since credibility of major fake information sources about climate change can polarize beliefs on a large scale, it is important to assess credibility of such sources and their potential negative effects on society. (In this context, polarization is one such negative or undesired effect).

Data availability

All data used can be publicly accessed from Twitter's Advanced Search function. Another way would be to use Twitter's API and query the tweets by their IDs; for this we make available the list of IDs of all tweets used for this study in the Dataverse repository: https://doi.org/10.7910/DVN/LNNPVD. The code to reproduce the results of the empirical analysis is also available in the above repository.

References

Acemoglu D, Dahleh MA, Lobel I, Ozdaglar A (2011) Bayesian learning in social networks. Rev Econ Stud 78(4):1201–1236

Acemoglu D, Ozdaglar A, ParandehGheibi A (2010) Spread of (mis)information in social networks. Games Econ Behav 70:194–227

Adamic LA, Glance N (2005) The political blogosphere and the 2004 U.S. Election: divided they blog. In: Proceedings of the 3rd international workshop on Link discovery. ACM

Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236

Aragón P, Kappler KE, Kaltenbrunner A, Laniado D, Volkovich Y (2013) Communication dynamics in Twitter during political campaigns: the case of the 2011 Spanish national election. Policy Internet 5(2):183–206

Bail CA, Argyle LP, Brown TW, Bumpus JP, Chen H, Hunzaker MBF, Lee J, Mann M, Merhout F, Volfovsky A (2018) Exposure to opposing views on social media can increase political polarization. Proc Natl Acad Sci 115 (37):9216–9221

Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. Science 348(6239):1130–1132

Baldassarri D, Gelman A (2008) Partisans without constraint: political polarization and trends in American public opinion. Am J Sociol 114(2):408–446

Barberá P (2015) Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. Polit Anal 23(1):76–91

Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2016) Homophily and polarization in the age of misinformation. Eur Phys J Spec Top 225(10):2047–2059

Bessi A, Zollo F, Vicario MD, Puliga M, Scala A, Caldarelli G, Uzzi B, Quattrociocchi W (2016) Users polarization on Facebook and Youtube. PLoS ONE 11(8):e0159641

Center PR (2014) Political polarization in the American public. Ann Rev Polit Sci

Ciampaglia GL, Shiralkar P, Rocha LM, Bollen J, Menczer F, Flammini A (2015) Computational fact checking from knowledge networks. PLoS ONE 10(6): e0128193

Colleoni E, Rozza A, Arvidsson A (2014) Echo chamber or public sphere? predicting political orientation and measuring political homophily in twitter using big data. J Commun 64(2):317–332

Conover MD, Ratkiewicz J, Francisco M, Goncalves B, Flammini A, Menczer F (2011) Political polarization on Twitter. In: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media. AAAI Press

Currarini S, Jackson MO, Pin P (2009) An economic model of friendship: homophily, minorities, and segregation. Econometrica 77(4):1003–1045

Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. Proc Natl Acad Sci 113(3):554–559

DiMaggio P, Evans J, Bryson B (1996) Have American's social attitudes become more polarized? Am J Sociol 102(3):690–755

Dubois E, Blank G (2018) The echo chamber is overstated: the moderating effect of political interest and diverse media. Inf Commun Soc 21(5):729–745

Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation, and testing. Econometrica 55(2):251–276

Esteban J-M, Ray D (1994) On the measurement of polarization. Econometrica 62 (4):819–851

Flaxman S, Goel S, Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. Public Opin Q 80(51):298–320

Freeman JB, Dale R (2013) Assessing bimodality to detect the presence of a dual cognitive process. Behav Res Methods 45(1):83–97

Garrett RK (2009) Echo chambers online?: politically motivated selective exposure among internet news users. J Comput-Mediated Commun 14(2):265–285

Golub B, Jackson MO (2012) How homophily affects the speed of learning and best-response dynamics. Q J Econ 127(3):1287–1338

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438

Granger C, Weiss A (1983) Time series analysis of error-correcting models. In: Karlin S, Amemiya T, Goodman LA (eds) Studies in econometrics, time series, and multivariate statistics. Academic Press, p 255–278

Greer JD (2003) Evaluating the credibility of online information: a test of source and advertising influence. Mass Commun Soc 6(1):11–28

Halberstam Y, Knight B (2016) Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. J Public Econ 143:73–88

Hamilton LC (2011) Education, politics and opinions about climate change evidence for interaction effects. Clim Change 104(2):231–242

Heesacker M, Petty RE, Cacioppo JT (1983) Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. J Personal 51(4):653–666

Hovland CI, Weiss W (1951) The influence of source credibility on communication effectiveness*. Public Opin Q 15(4):635–650

Hutto C, Gilbert E (2014) VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media. AAAI Press

Iyengar S, Massey DS (2019) Scientific communication in a post-truth society. Proc Natl Acad Sci 116(16):7656–7661

Iyengar S, Sood G, Lelkes Y (2012) Affect, not ideology: a social identity perspective on polarization. Public Opin Q 76(3):405–431

Iyengar S, Westwood SJ (2015) Fear and loathing across party lines: new evidence on group polarization. Am J Polit Sci 59(3):690–707

Jackson MO, López-Pintado D (2013) Diffusion and contagion in networks with heterogeneous agents and homophily. Netw Sci 1(1):49–67

Jasny L, Waggle J, Fisher DR (2015) An empirical examination of echo chambers in US climate policy networks. Nat Clim Change 5:782–786

Kiousis S (2001) Public trust or mistrust? Perceptions of media credibility in the information age. Mass Commun Soc 4(4):381–403

Kušen E, Strembeck M (2018) Politics, sentiments, and misinformation: an analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. Online Soc Netw Media 5:37–50

Lelkes Y (2016) Mass polarization: manifestations and measurements. Public Opin Q 80(S1):392–410

Lietz H, Wagner C, Bleier A, Strohmaier M (2014) When politicians talk: Assessing online conversational practices of political parties on Twitter. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. AAAI Press

Madsen JK, Bailey RM, Pilditch TD (2018) Large networks of rational agents form persistent echo chambers. Sci Rep 8:12391

Markham D (2006) The dimensions of source credibility of television newscasters. J Commun 18(1):57–64

McCright AM, Dunlap RE (2010) Anti-reflexivity. Theory, Cult Soc 27(2-3):100–133

McCright AM, Dunlap RE (2011) The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. Sociological Q 52(2):155–194

McCright AM, Dunlap RE, Xiao C (2014) The impacts of temperature anomalies and political orientation on perceived winter warming. Nat Clim Change 4:1077–1081

Mohammad SM, Sobhani P, Kiritchenko S (2017) Stance and sentiment in tweets. ACM Trans Internet Technol 17(3):26:1–26:23

Moore FC, Obradovich N, Lehner F, Baylis P (2019) Rapidly declining remarkability of temperature anomalies may obscure public perception of climate change. Proc Natl Acad Sci 116(11):4905–4910

Pfister R, Schwarz KA, Janczyk M, Dale R, Freeman JB (2013) Good things peak in pairs: a note on the bimodality coefficient. Front Psychol 4:700

Pornpitakpan C (2004) The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Soc Psychol 34(2):243–281

Prior M (2013) Media and political polarization. Annu Rev Polit Sci 16(1):101–127

Shin J, Jian L, Driscoll K, Bar F (2017) Political rumoring on Twitter during the 2012 US presidential election: Rumor diffusion and correction. New Media Soc 19(8):1214–1235

Sobhani P, Mohammad SM, Kiritchenko S (2016) Detecting stance in tweets and analyzing its interaction with sentiment. In: Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics. Association for Computational Linguistics, p 159–169

Taylor CE, Mantzaris AV, Garibay I (2018) Exploring how homophily and accessibility can facilitate polarization in social networks. Information 9:325

Toda HY, Yamamoto T (1995) Statistical inference in vector autoregressions possibly integrated processes. J Econ 66:225–250

Tvinnereim E, Liu X, Jamelske EM (2017) Public perceptions of air pollution and climate change: different manifestations, similar causes, and concerns. Climatic Change 140(3–4):399–412

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Warner BR (2010) Segmenting the electorate: the effects of exposure to political extremism online. Commun Stud 61(4):430–444

Westerman D, Spence PR, Van Der Heide B (2014) Social media as information source: recency of updates and credibility of information*. J Computer-Mediated Commun 19(2):171–183

Williams AE (2016) Media evolution and public understanding of climate science. Politics Life Sci 30(2):20–30

Williams HT, McMurray JR, Kurz T, Lambert FH (2015) Network analysis reveals open forums and echo chambers in social media discussions of climate change. Glob Environ Change 32:126–138

Zhao Z, Resnick P, Mei Q (2015) Enquiring minds: early detection of rumors in social media from enquiry posts. In: Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, p 1395–1405

Acknowledgements

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1057/s41599-019-0344-4.

Correspondence and requests for materials should be addressed to A.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.