Gaining a better understanding of online polarization by approaching it as a dynamic process

Celina Treuillier, Sylvain Castagnos, Christèle Lagier, Armelle Brun

scientific reports

Check for updates

OPEN

Gaining a better understanding of online polarization by approaching it as a dynamic process

Célina Treuillier[1]✉, Sylvain Castagnos[1], Christèle Lagier[2] & Armelle Brun[1]

Polarization is often a cliché, its conceptualization remains approximate and no consensus has been reached so far. Often simply seen as an inevitable result of the use of social networks, polarization cannot be viewed solely from an ideological or affective perspective. We propose to better understand the issue polarization by approaching it as a sequential process, drawing on a dual expertise in political and data sciences. We compare the polarization process between one mature debate (COVID-19 vaccine) and one emerging debate (Ukraine conflict) at the time of data collection. Both debates are studied on Twitter users, a highly politicized population, and on the French population to provide key elements beyond the traditional US context. This unprecedented analysis confirms that polarization varies over time, through a succession of specific periods, whose existence and duration depend on the maturity of the debate. Importantly, we highlight that polarization is paced by context-related events. Bearing this in mind, we pave the way for a new generation of personalized depolarization strategies, adapted to the context and maturity of debates.

Surveys on polarization proliferate, but no consensus has been reached on the definition of this highly democratic topic[1]. The term *polarization* is employed in favor of very heterogeneous analyses[2]. It is becoming the root of all evil. It is for example used to describe the polarization of juries, or the vote in favor of Brexit, as well as for the decisions made in deliberative assemblies or journalistic commentary in the face of the invasion of the US capitol in January 2021. This predigested approach of polarization obscures the fact that these events are also embedded at the crossroads of social phenomena that political science has long been studying. These include the unequal distribution of political competence[3], resentment[4,5] and distrust of political leaders, the shortcomings of political representation[6], contextual effects[7,8], the weight of primary groups[9,10] or of discussion[11], and politics avoidance[12] in the formation of opinions. These social determinants have an impact on the level of politicization of people and we consider in this work that social networks do not provide a real vision of what is happening among ordinary citizens[13]. In this new informational context, many works attempt to model[14–16], measure[17–20], and identify[21] polarization. These works show that the very format of social networks contributes to exacerbating polarization – that is, in part, artificially co-constructing it – through anonymity and increased selective exposure. Yet these works tend to adopt a purely data-oriented analysis and barely consider the underlying social phenomena, although they are aware of their existence[22]. Finally, by a circular effect, the certainty of the existence of polarization is reinforced by analyses produced on the basis of models that struggle to go beyond a binary identification of polarization (polarized *vs.* non-polarized).

In our view, the main pitfall of these works lies in the fact that polarization is taken for granted in a context of high social and political tension that characterize our society[23]. Two presuppositions, often poorly explained, underlie these approaches. On the one hand, polarization is often reduced to ideological polarization[24] and affective polarization[25] which require knowledge of individuals' partisan labels to be accurately measured. On the other hand, polarization is mainly seen as an effect of social networks[26,27] i.e. where it is measured. Interest in polarization is proportional to the availability of digital data in virtually unlimited quantities. The goal of our work is not to question the impact of social media on polarization[1,13], particularly related to the algorithmic filtering of information that tends to confirm users in their beliefs[28]. We raise the following question: How polarization, and associated behaviors, evolves over time? (RQ1) Combining contributions from political science and data science, we thus propose to focus on issue polarization[29] in a temporal perspective . We assume

[1]Université de Lorraine, CNRS, LORIA, Vandœuvre-lès-Nancy, France. [2]Avignon Université, JPEG UPR 3788, Avignon, France. ✉email: celina.treuillier@loria.fr

that polarization mechanisms (notably social influence and the role played by persuasive arguments[2]) can be cumulative or successive when polarization processes are observed over the medium to long term. In line with the seminal work of social psychology[30], which has contributed to move away from the traditional opposition between the individual and the group, we see polarization as a sequential process rather than a state.

Besides, this work also contributes to overcome the idea that polarization is mainly related to the less informed or misinformed categories of the population. Polarized people are not just the victims of fake news or moral contagion phenomena[31] that crop up in many analyses of the dynamics of opinion[14]. Political science studies have shown that these categories of people are more likely to exit than to take a stand, which is borne out by the high abstention rates in recent elections[32]. On the population at large, people who polarize are first and foremost those who have the ability to have a structured opinion on societal debates, and the more asserted this opinion, the less likely it is to be modified[33].

We therefore propose to get around these biases by working on the population of Twitter users which is particularly interesting, because of its increased awareness about political matters compared to the average citizen[34,35]. We can identify their ideological inclination[36] on the basis of positions taken on certain issues in the public debate. Besides, discussions on Twitter primarily reflect the concerns and topics addressed by mainstream media. In some respects, this social network appears to be tightly correlated with media framing[37,38]. Twitter users are also a population of opinion leaders[33] who are particularly subject to the powerful effects of the environment and selective media exposure.

We focus on the French context, little studied in the literature, where left-right referents still strongly structure the political and media field. Indeed, while this distinction tends to become blurred for part of the population[39], but also under the influence of monopolistic media organization, it still makes sense in a media system inherited from a high level of political homology[40].

Studies on polarization sometimes present contradictory conclusions[1]. However, to the best of our knowledge, most of these works have a global approach, i.e. study polarization as a whole, and few of them address the temporal evolution of associated polarization behaviors. When polarization is seen as a state, it corresponds to the crystallization of opinions, which has very little chance of being modified[41]. Attempts to depolarize mainly rely on the increase of the diversity of the information recommended to users, to make them confronted with a variety of topics and viewpoints. However, these analyses are not conclusive[42–44] as diversity may even further reinforce polarization among users who become resistant to the confrontation of ideas too far from their own[45]. Furthermore, modeling polarization at a given point in time may be completely obsolete sometime later, since it does not consider opinion dynamics. These limits in mind, some works propose to study the evolution of the overall polarization over time[34,37,38].

Although multiple topics and societal issues have been of interest in the literature, we can see that they are all mature topics, i.e. topics that have been discussed for months, even years, about which people have had time to take a stand and express clear opinions[46]. As a consequence, we ask the following question: What are the polarization dynamics when a new controversial debate emerges? (RQ2). This is something which remains understudied. An in-depth analysis of the dynamics of polarization about such new controversial topics will help understand how people polarize in the early stages of an emerging debate. To go further, we also propose an analysis of the relationship between unrelated debates, and study whether users interacting with several debates are polarized along the same lines.

The way of research explored here is to consider this process from a sequential point of view, studying issue polarization on two debates to measure the impact of the maturity of the debate and the influence of context. Concretely, we study temporal polarization behaviors on Twitter about the Ukraine conflict that strongly intensified from February 24, 2022, the date on which the Russian army invaded Ukraine, marking the start of the conflict. The period of study chosen is between January and July 2022. Although this conflict arose from 2014 with the annexation of Crimea, in 2021 this topic was no longer discussed in the French media for several years. We also analyze a more mature debate, about which users should already have taken a position about it. We chose the COVID-19 vaccine debate, with the same period of study. For both debates, we distinguish between users belonging to each community (pro-vaccine *vs.* anti-vaccine, pro-Ukraine *vs.* pro-Russia).

This work is one of the first works that addresses polarization from a sequential perspective, and highlights important elements about the evolution of polarization, the impact of the maturity of the debate, and the influence of the context. We confirm that user polarization not only evolves through time, but also differently according to the maturity of the topic. This evolution is not erratic, and specific periods of polarization are identified. More importantly, patterns of such periods are common to debates. They confirm that polarization is a process in which users tend to gradually and naturally come close to extremes and polarize. The evolution of the polarization process can be disturbed by context-related events (covered by the media and discussed on Twitter), which provoke a reset in the pattern of periods and foster users' interactions with the opposing community. The duration of this depends on the maturity of the debate and the nature of the event. We thus assume that cycles of polarization occur, cadenced by the context such as the news events. This opens up opportunities for depolarization strategies that are not only personalized, but also related to the appearance of specific news events. This goes against current depolarization strategies that simply consider diversity.

Results

An aggregate analysis of polarization

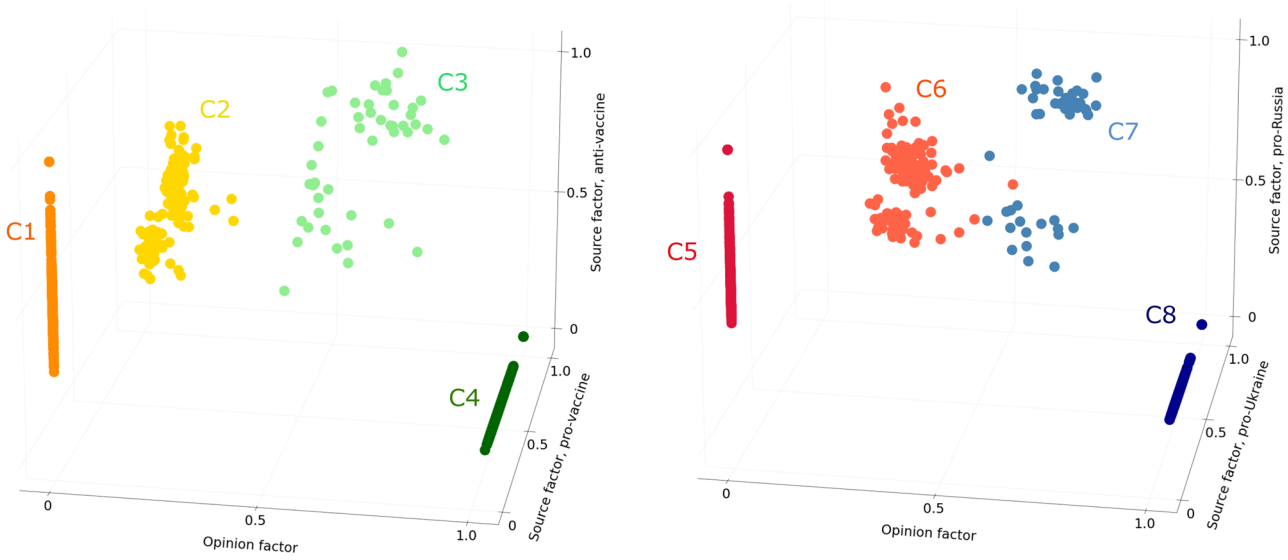
We first adopt an aggregate analysis of polarization and compare issue polarization within and between debates: the COVID-19 vaccine and the Ukraine conflict debates. The aggregate analysis relies on an automatic clustering of users, based on their retweet activity (interactions). We consider that clusters formed are made up of users adopting similar polarization behaviors. We exploit three factors. One opinion factor, computed from users'

interactions with each of the confronting communities, represents the diversity of opinions (in the Ukraine conflict debate for example, 1 represents an extreme polarization in the pro-Ukraine community and 0 an extreme polarization in the pro-Russia community, while 0.5 represents a fair distribution between both communities). Two source factors, computed from users' interactions on sources of information, represent to which extent a user accesses a large set of sources in either community (with 1 representing an access to a unique source and 0 the access to all sources of information in the community). Identified clusters thus capture the polarization of users according to both their degree of belonging to one or other of the communities (opinion) and the diversity of sources they interact within each of these communities (sources).

For each debate, the optimal number of clusters is 4. The clustering is of high quality as Silhouette indexes are equal to 0.85 and 0.87 (the higher the better) and Davies–Bouldin indexes equal to 0.35 and 0.32 (the lower the better) for the COVID-19 vaccine and Ukraine conflict debates respectively. Identified clusters are presented in a three-dimensional space, where each dimension corresponds to one factor (opinion factor and the two source factors) (see Fig. 1).

In each debate findings are similar. We identify two major clusters corresponding to users who interact with one community only (C1 and C4, C5 and C8), whatever is the diversity of sources they access in this community. We refer to these users as polarized users. The two other clusters (C2 and C3, C6 and C7) are made up of users interacting with both communities, but who still have a preference for one side of the debate, with which they are more active. In their favored community, users interact with multiple sources. In the other community, they interact with a varying number of sources, ranging from a unique to all sources. These users thus confront opposing viewpoints. We refer to them as intermediate users. They represent a significant part of the population: 19% for the COVID-19 vaccine debate and 21% for the Ukraine conflict debate. Such a significant representation of intermediate users is an interesting finding of our study, especially given the polarizing nature of the debates. It has been shown that the emotions triggered by a polarizing debate can lead to further polarization[47]. We could therefore expected the number of intermediate users to be low, but that is not the case in this study (20% intermediate users on average). We also note that these proportions of intermediate users are similar for both debates. We might have thought that the number of intermediate users involved in the emerging debate would have been higher.

To summarize, this aggregate analysis, based on an individual and multi-factorial representation of polarization, contributes to differentiating between four distinct polarization classes. It therefore goes beyond the simple distinction between polarized and non-polarized users from the literature[19,48]. In particular, it contributes to highlighting the existence of clusters of intermediate users, with a moderate level of polarization. Although the two selected debates are not related and have different maturity levels at the time of data collection, the clusters formed are similar in number and interpretation, and are made up of a similar proportion of users. In light of this aggregate analysis, one could conclude that the strength of issue polarization does not seem to depend on the maturity of the debate. Nevertheless, we wonder how these clusters evolve over time (RQ1), both in terms of number and nature, and if this evolution differs according to the maturity of the debate.



(a) COVID-19 vaccine debate.

(b) Ukraine conflict debate.

Figure 1. Clusters resulting from the aggregate analysis. Figure (**a**) presents the clusters identified among users interacting about the COVID-19 vaccine debate (n = 1000). The proportions of users in each cluster are as follows: C1 = 36%, C2 = 14%, C3 = 5%, and C4 = 45%. Figure (**b**) presents the clusters identified among users interacting about the Ukraine conflict debate (n = 1000). The proportions of users in each cluster are as follows: C5 = 34%, C6 = 16%, C7 = 5%, and C8 = 45%.

A time-aware analysis of polarization

Recall that in this work we approach polarization as an evolving process, by which it may be possible to identify if the clusters previously identified evolve over time. The study relies on the analysis of clusters of users automatically identified within timeframes. Each timeframe is 4 weeks long and consecutive timeframes have a 2-week overlap. By looking at the number of identified clusters in each timeframe, we can observe that it actually varies through time (see Fig. 2). This shows that user polarization actually evolves, whatever is the maturity of the debate. While variations were conceivable for the Ukraine conflict debate which was recent at the time of data collection, such variations are particularly intriguing for the COVID-19 debate, which had been discussed for a long time.

In addition, this evolution differs between debates. For the long-lasting COVID-19 vaccine debate, the number of clusters varies between 2 and 4 clusters. During the three first timeframes, ranging from January 1 to February 26, 4 clusters are identified. These clusters (C9, C10, C11, and C12 in Fig. 3a) are consistent with those identified in the aggregate analysis of polarization (C1, C2, C3, and C4 in Fig. 1a). Indeed, two clusters (C8 and C12) are made up of polarized users, interacting in a unique community, while two clusters (C10 and C11) are made up of intermediate users, interacting in both communities, with unbalanced interactions with one community. From the 4th timeframe till the end of the dataset timespan, only 2 clusters are discriminated (C13 and C14 in Fig. 3b). During these timeframes, all users are identified as polarized.

For the Ukraine conflict emerging debate, the number of clusters identified is much more variable, ranging between 2 and 9 clusters. Before the conflict was officially declared (first and second timeframes), the number of clusters is high: 8 and 9 clusters are identified. From the third timeframe, the number of clusters greatly reduces, with only three clusters differentiated (C15, C16, and C17 in Fig. 4a). As in the case of the aggregate analysis, two clusters of polarized users are identified, but the difference lies in the clustering of intermediate users who are gathered into a unique cluster (C16). Right away, from the beginning of March, four clusters are identified and last from March to June (C18, C19, C20, and C21 in Fig. 4b). During these consecutive timeframes, clusters are similar to those identified during the aggregate analysis (C5, C6, C7 and C8 in Fig. 1b), with two clusters of polarized users, and two clusters of intermediate users. Finally, during the last two timeframes, only two clusters are discriminated (C22 and C23 in Fig. 4c). As for the last timeframes of the COVID-19 vaccine debate, all users are identified as polarized users.

We can conclude that analyzing polarization as an evolving process contributes to highlight that Twitter users' polarization evolves over time, as well as the clusters they form, both in terms of number and nature. While a population of stable polarized users is maintained, presumably comprising the most politicized individuals, one or more intermediate sets of individuals are subject to change. Besides, this time-aware analysis contributes to differentiate between debates, which were identified as similar with the aggregate analysis. While some similarities can be noticed between the two debates, their overall evolution is different. These differences could be associated with their maturity. This answers the first research question (RQ1) and strengthens the relevance of considering issue polarization as a sequential process, not only as a state.

A period-based analysis of polarization

To gain a better understanding of the polarization process, we now focus on the structure of the previously identified changes over time. We start by defining a period as a sequence of consecutive timeframes with consistent clusters. Concretely, two sets of clusters are considered as consistent if their interpretation is the same.

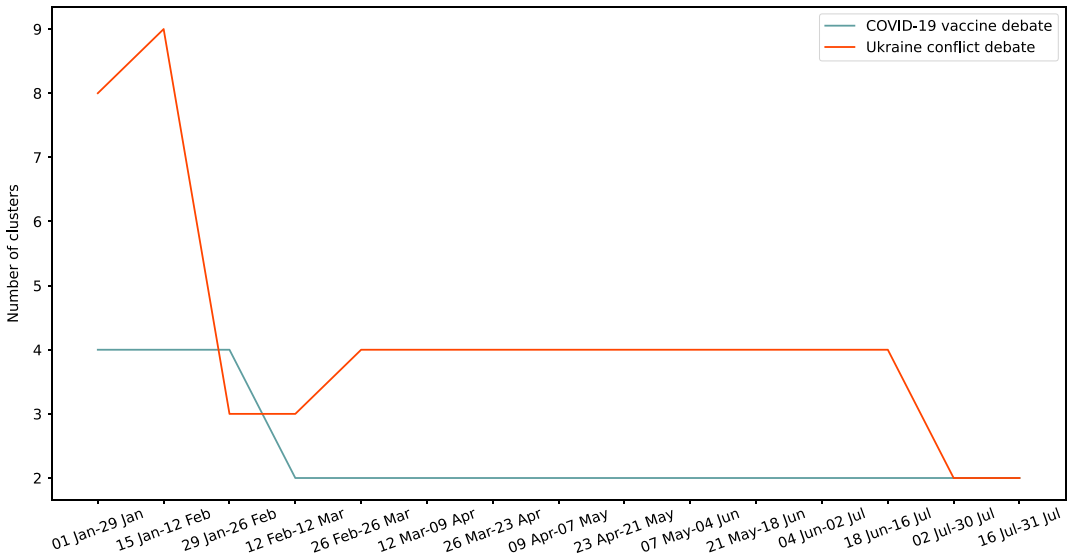
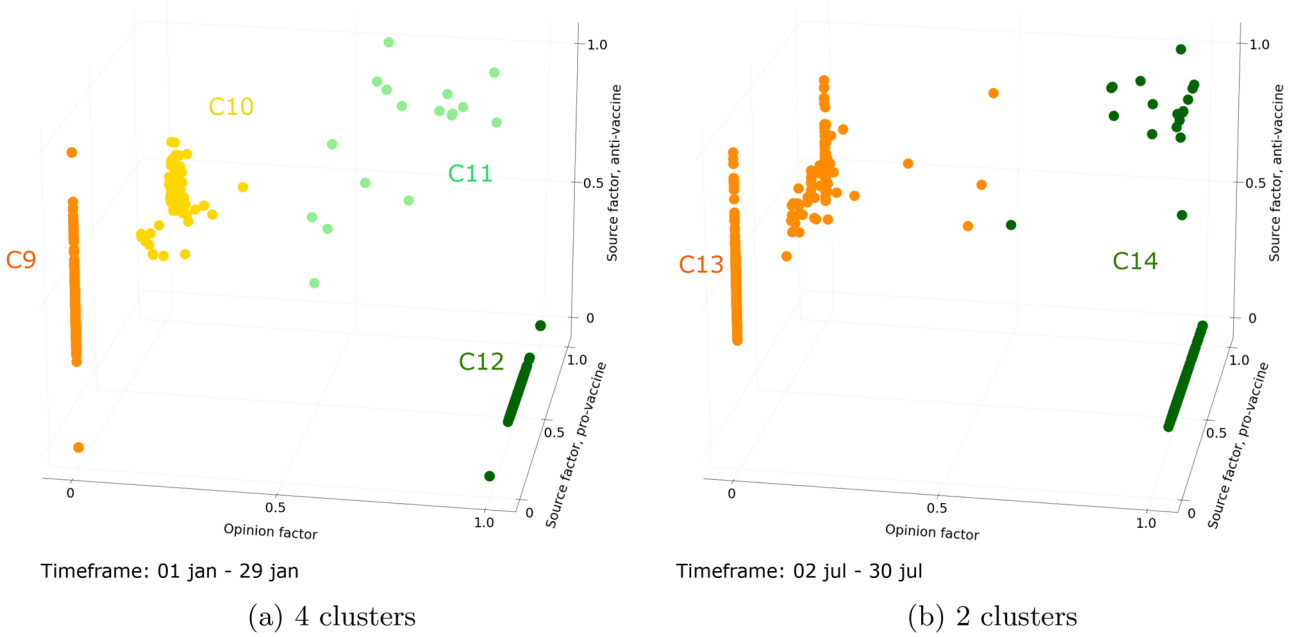


Figure 2. Evolution of the number of identified clusters. The figure shows the evolution of the number of identified clusters over each timeframe.



(a) 4 clusters (b) 2 clusters

Figure 3. Clusters of users interacting about the COVID-19 vaccine debate (n = 685 users active on at least 80% of timeframes). Figure (**a**) presents the clusters identified among users during a 4-clusters timeframe, extending from January 1 to January 29, 2022. The proportions of users in each cluster are as follows: C9 = 53%, C10 = 13%, C11 = 2%, and C12 = 32%. Figure (**b**) presents the clusters identified among users during a 2-clusters timeframe, extending from July 2 to July 31, 2022. The proportions of users in each cluster are as follows: C13 = 66%, C13 = 34%.

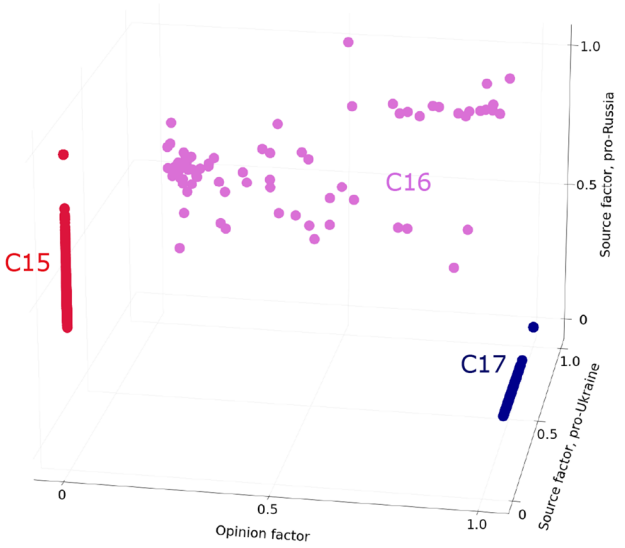
In both debates, periods are actually identified. For example, in the COVID-19 vaccine debate, a first period made up of 4 clusters persists over 3 timeframes. A second period, made up of 2 clusters, remains during the following 5 months. In the Ukraine conflict debate, a greater number of periods is identified: a first 1.5 month long period is made up of numerous clusters, followed by a short 3-cluster period, and then by two periods of 4 and 2 clusters.

In order to understand the transition from one period to another, we propose to look closer at the evolution of the distribution of users between periods. In particular, we focus on intermediate users, who are those who vary the most over time. For the COVID-19 debate, we see that intermediate users from the first period (C10 and C11 in Fig. 3a), have moved closer to polarized users, ultimately forming only two clusters of polarized users in the last period (C13 and C14 in Fig. 3b). For the Ukraine conflict debate, we see that intermediate users who have no preferred community in the second period (C16 in Fig. 4a) then split into two clusters of intermediate users (C19 and C20 in Fig. 4b), finally getting closer to polarized users and thus being identified as polarized during the last timeframes (C22 and C23 in Fig. 4c).

To go further, we analyze how the users evolves within the longest periods identified for each debate. Looking first at the evolution of users during the 5 months long period for the COVID-19 vaccine debate (Fig. 5), users continue to polarize even more over time, especially those who were identified as intermediate during the previous period. The interactions between opposing communities are thus drastically reduced, and this divergence gradually increases during this period. About the Ukraine conflict debate, looking at the evolution of the distribution of users during the longest period (see Fig. 6), we confirm that intermediate users are gradually moving closer to polarized users. This reflects a drop in interactions in the community that is not their preferred community. Once the imbalance between is too important, and intermediate users are too close to polarized users, this convergence period ends and forms only two clusters of polarized users.

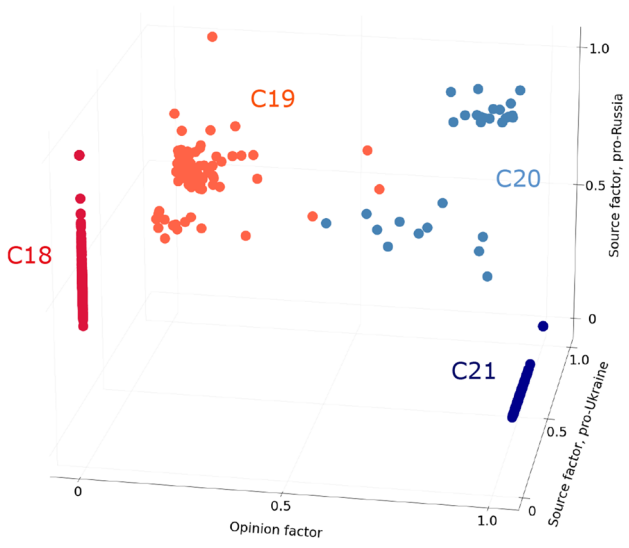
Here we characterize the set of previously identified periods:

- **Unstructured periods** are periods with numerous clusters, among which no specific behaviors can be identified. In the Ukraine conflict debate, it corresponds to the first period. This period has no instance in the more settled COVID-19 vaccine debate.
- **Balanced periods** are periods made up of three clusters: two clusters of users adopting clear-cut positions and one cluster of users maintaining balanced interactions in the two opposing communities. This period is the second one identified for the Ukraine conflict debate, while it has no instance in the period analyzed for the COVID-19 vaccine debate.
- **Convergence periods** are periods where four clusters are differentiated: two clusters of polarized users, and two clusters of intermediate users. The latter are getting closer to polarized users over the period. Such a period is identified in both debates.
- **Polarized periods** are periods composed of only two clusters of polarized users, belonging to either community. Such a period is identified in both debates.



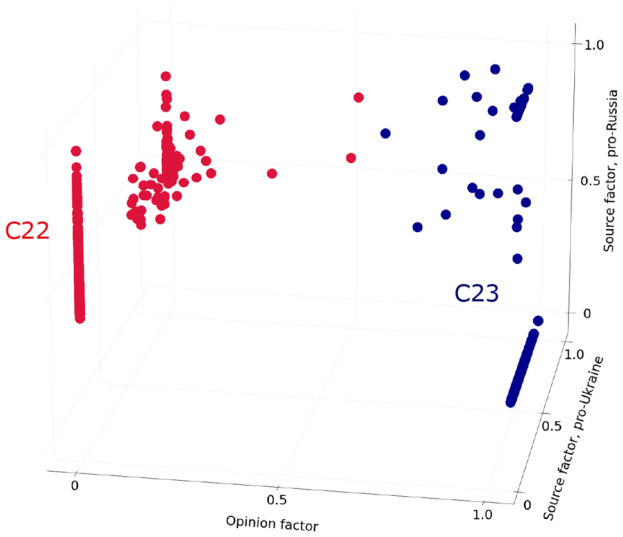
Timeframe: 12 feb - 12 mar

(a) 3 clusters



Timeframe: 12 mar - 9 apr

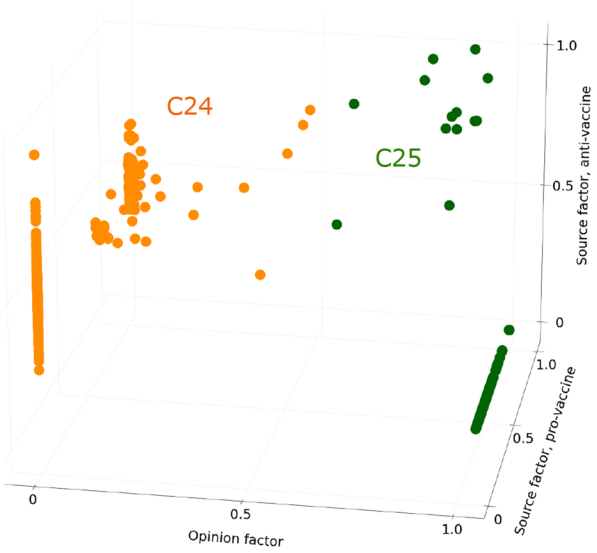
(b) 4 clusters



Timeframe: 02 jul - 30 jul

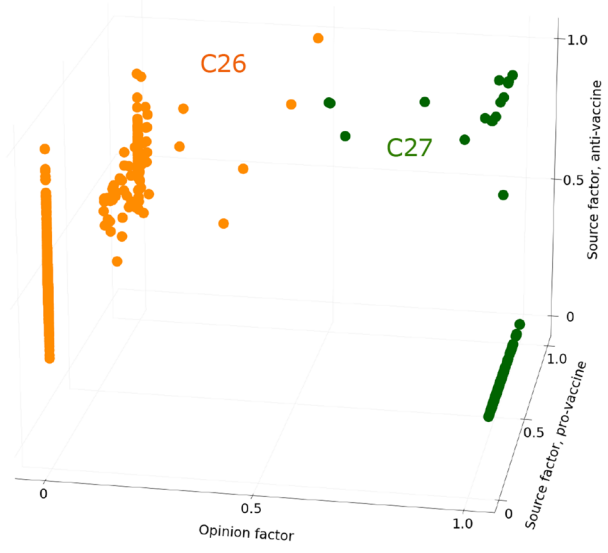
(c) 2 clusters

Figure 4. Clusters of users interacting about the Ukraine conflict debate (n = 784 users active on at least 80% of timeframes). Figure (**a**) presents the clusters identified among users during a 3-cluster timeframe, extending from February 12 to March 12, 2022. The proportions of users in each cluster are as follows: C15 = 50%, C16 = 10%, C17 = 40%. Figure (**b**) presents the clusters identified among users during a 4-cluster timeframe, extending from March 12 to April 9, 2022. The proportions of users in each cluster are as follows: C18 = 42%, C19 = 14%, C20 = 4%, C21 = 40%. Figure (**c**) presents the clusters identified among users during a 2-clusters timeframe, extending from July 2 to July 30, 2022. The proportions of users in each cluster are as follows: C22 = 56%, C23 = 44%.



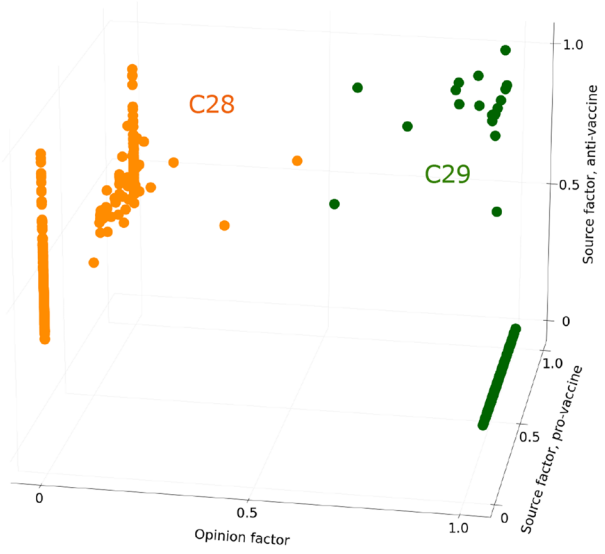
Timeframe: 12 feb - 12 mar

(a) Initial position.



Timeframe: 23 apr - 21 may

(b) Intermediary position.

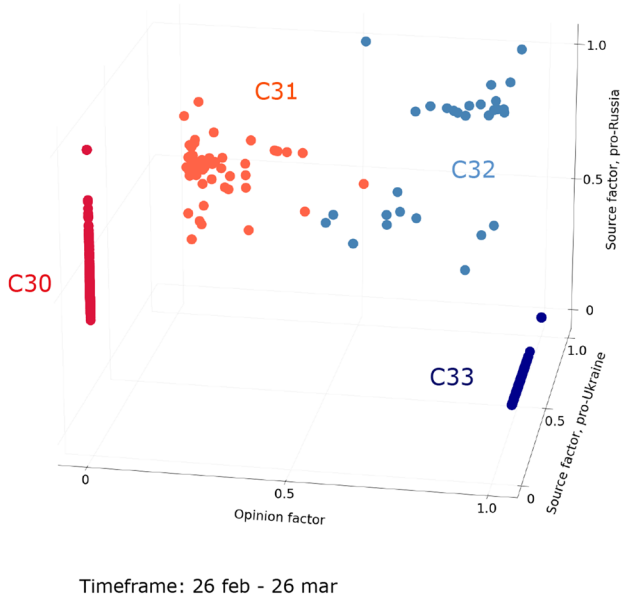


Timeframe: 02 jul - 30 jul

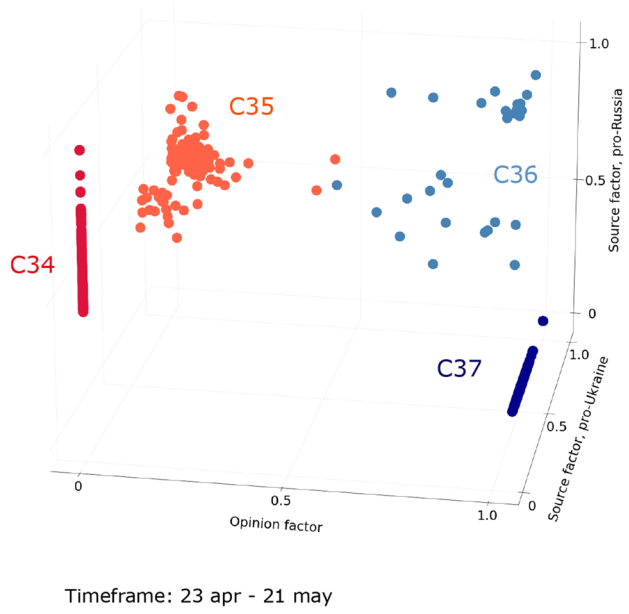
(c) Final position.

Figure 5. Evolution of users interacting with the COVID-19 vaccine debate during the longest period (n = 685 users active on at least 80% of timeframes). Figure (**a**) shows the initial position of users during a timeframe extending from February 12 to March 12, 2022. Figure (**b**) shows the intermediary position of users during a timeframe extending from April 23 to May 21, 2022. Figure (**c**) shows the final position of users during a timeframe extending from July 2 to July 30, 2022.

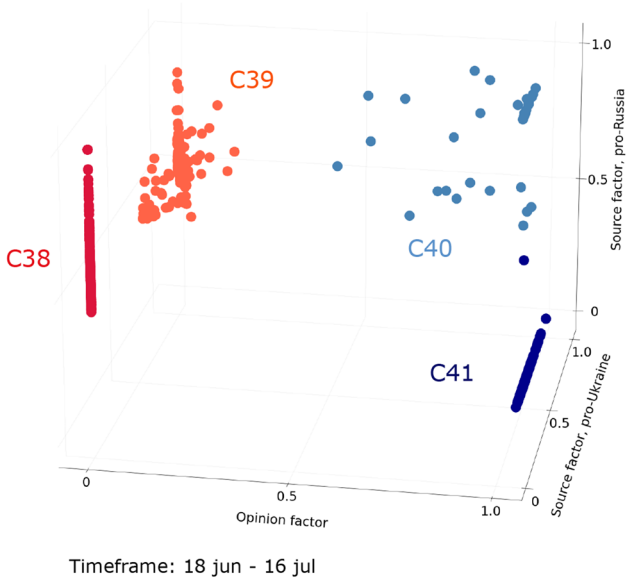
If one steps back, the difference in periods between the two debates can be explained by the maturity of the debates which probably influences the way users interact with each other. In the Ukraine conflict debate, the unstructured and balanced periods are identified, which translates the clear distinction between polarization ahead of the emergence of the debate (unstructured period), and polarization from the start of the debate (balanced period). These two periods are not identified for the COVID-19 vaccine debate, which was discussed for a long time at the time of the data collection. Balanced users thus do not exist anymore in the COVID-19 vaccine debate as most of the users already have a preferred side of the debate.



(a) Initial position.



(b) Intermediary position.

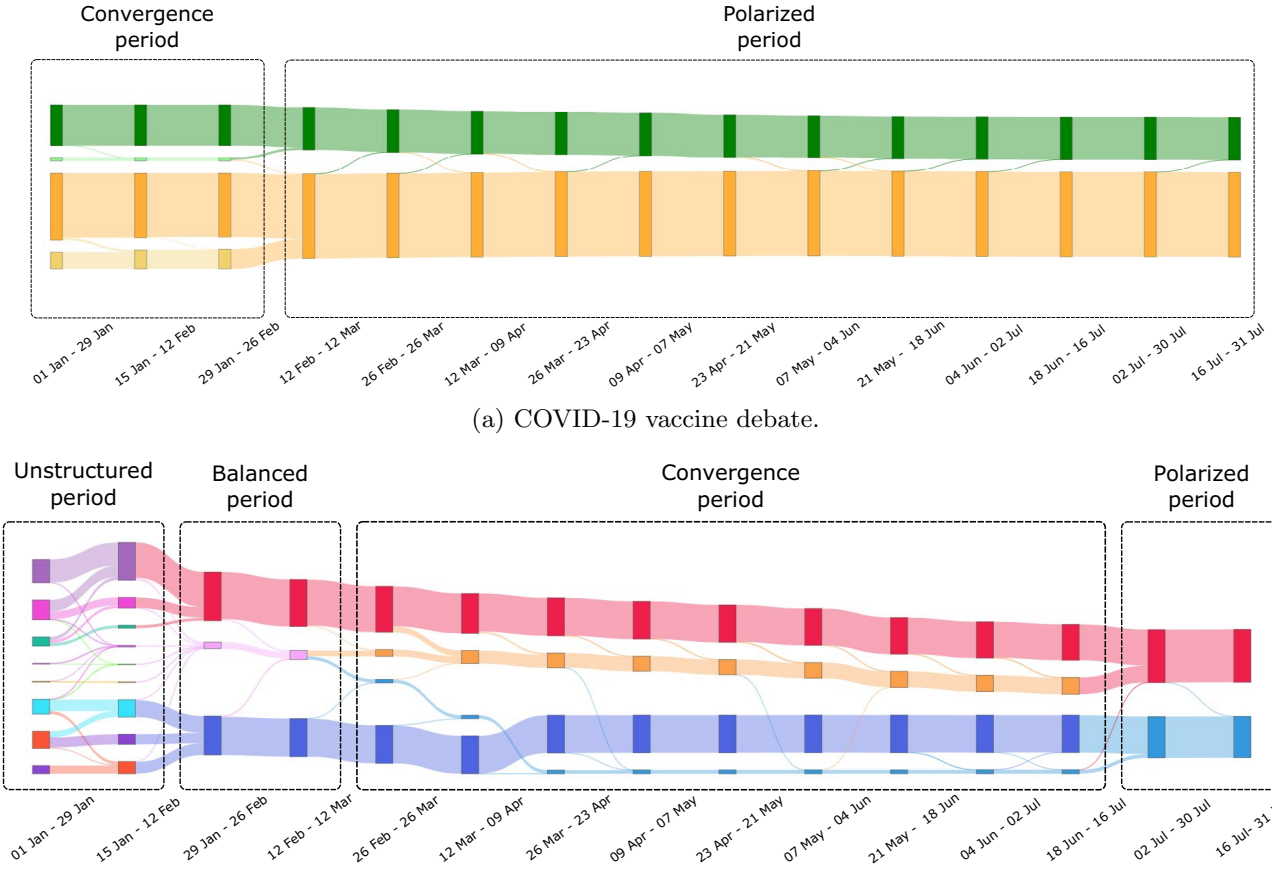


(c) Final position.

Figure 6. Evolution of users interacting with the Ukraine conflict debate during the longest period (n = 784 users active on at least 80% of timeframes). Figure (**a**) shows the initial position of users during a timeframe extending from February 26 to March 26, 2022. Figure (**b**) shows the intermediary position of users during a timeframe extending from April 23 to May 21, 2022. Figure (**c**) shows the final position of users during a timeframe extending from June 18 to July 16, 2022.

Nevertheless, a specific pattern of periods is shared by both debates: the convergence period is systematically followed by the polarized period. This pattern reflects a gradual decline in the interest of intermediate users in the community that is not their major community, and users systematically end up polarized. Intermediate users therefore do not appear to be stable over time. However, the duration of the periods in this pattern varies between debates. Here again, the maturity of the debate can explain the difference. The sequence of periods and associated clusters are presented in Fig. 7.

To summarize, this period-based analysis contributes to highlight clear similarities between the two debates. Yet, we have highlighted the existence of specific periods, some of them forming a pattern common to both debates. This answers the second research question (RQ2) and shows that user polarization follows a predefined



(a) COVID-19 vaccine debate.

(b) Ukraine conflict debate.

Figure 7. Sankey diagrams of the evolution of identified clusters during different periods. Figure (**a**) presents the evolution of the number of clusters during convergence and polarized periods for the COVID-19 vaccine debate, and figure (**b**) presents the evolution of the number of clusters during the unstructured, disturbance, convergence and polarized periods for the Ukraine conflict debate.

pattern, independent of the debate and its maturity. To go further, an analysis of the contexts underlying the temporal evolution of clusters could help to better characterize polarization dynamics.

A context-aware analysis of polarization

User behavior obviously takes place within a broader context. We hypothesize that the differences previously outlined between debates are in part the result of external or debate-related events and we discuss this hypothesis here.

We would like to first get back to the unstructured period identified for the Ukraine conflict debate only. Even if the tensions between Russia and Ukraine were discussed before February 2022, the media framing of this topic was secondary at the time, and did not promote opinion crystallization. It is thus probable that people were not much aware of the debate. As soon as the war was declared, this unstructured period ended, followed by the balanced period. The balanced position of intermediate users in this period is transient, and results in the emergence of imbalanced interactions between the two communities in an associated extended convergence period. This long period can be explained by the fact that news media have devoted considerable space to covering this topic in Spring 2022. This stimulates discussion with close friends and family, keeps users aware of the debate, but it does not necessarily change the ideas already shared[9].

About the COVID-19 vaccine debate, the lack of unstructured and balanced periods does not imply that they did not occur, they may simply have occurred prior to the period of study. For the specific case of the unstructured period, it is probable that it did not appear some time before January, 2022. Indeed, the discussions about the vaccine debate reemerged with the pandemic, and the debate was already active. The identification of a convergence period is quite surprising, as it was expected that the debate would be stable at the beginning of 2022. Looking at the French events in these times, it turns out that a measure was taken some time before the starting date of the dataset. In fact, the vaccination campaign was opened to children aged 5 to 11 on December 21, 2021. This was covered by the media and thus sparked discussions on Twitter, making some users more aware of discussions from their opposite community. The question of the possible existence of a balanced period at the end of December remains. In all cases, the convergence period has started at the latest on December 21, it is thus at most 1.5 months long. It makes a significant difference with the duration of this period in the Ukraine conflict debate (5 months). From a general perspective, we can conclude that in each debate the emergence of

intermediate users (balanced or convergence periods) is triggered by a disruptive event. On one side, the invasion of Ukraine by Russia, which is a sudden event in an immature debate. It has led to mass mobilization, had a major impact on society, and generated massive reactions. On the other side, a governmental decision linked to the COVID-19 vaccine, which is not sudden (an announcement has been made some days before its implementation) and occurs in a well-established debate. These events are supported by a modified media framing, and an increase in Twitter users' activity on the related debate. However, the impact of both events differs: in the emerging debate, it first leads to a balanced period, where a significant proportion of users show no preference for either community, followed by a convergence period lasting several months. In the mature debate it encourages some users, probably polarized, to seek information in the opposing community during a short period of convergence. In our view, the maturity of the debate and the suddenness of the event are the two main reasons for this difference. It confirms that users with a constructed opinion on a societal debate can remain aware of the evolution of the debate, by getting interested in the opposing community. However, the more crystallized this opinion, the faster they go back to their community.

This context-aware analysis has shown that, whatever the maturity of the topic, some events may impact user polarization, the duration of which depends on the maturity of the debate and the type of event. We can thus suppose that debates, even when opinions are crystallized, may cyclically have intermediate users appearing, triggered by external events.

A cross-debate analysis of polarization

To go beyond considering debates independently of each other, we propose to study the relationship between both debates. Indeed, people who take a clear stance on one issue often belong to a predictable and specific community on another issue[49]. This is an important challenge linked to the polarization phenomenon, as people polarize themselves on the same lines on several debates, without necessarily being properly informed about underlying factors. We thus study the behavior of the subset of 170 users interacting with both debates. Firstly, performing an aggregate analysis, the community to which users belong on one of the debates correlates with the community to which they belong on the other debate (Cramer's V = 0.96). Indeed, the 133 users interacting with the anti-vaccine all interact with pro-Russia content. Besides, among the 37 users interacting with pro-vaccine sources, 36 interact with pro-Ukraine content, and one single user rather interacts with pro-Russia sources. This clearly highlights that, even on seemingly unrelated debates, the adopted position on a debate can be driven by the position on another debate. However, focusing on the strength of polarization of users, a user who is highly polarized on one debate may be intermediate on the other one, and vice versa (Cramer's V = 0.60). For example, in the set of 170 users interacting with both communities, among the 75 users being polarized in the anti-vaccine community (cluster C1 on Fig. 1a), 28% are intermediate on the emerging Ukraine debate (cluster C6 on Fig. 1b). Conversely, among the 58 intermediate users who are closer to the anti-vaccine debate (cluster C2 on Fig. 1c), 50% are also intermediate about the Ukraine conflict, and 50% are completely polarized on this debate. Similar conclusions are drawn for users interacting with pro-vaccine and pro-Ukraine sources. We can thus conclude that the strength of polarization on one individual debate does not necessarily impact the strength of polarization on other debates, but rather influences towards a specific community. From a temporal perspective, the dynamics of polarization can vary from one debate to another, i.e. users can observe variations on one debate, probably due to the context, but not on another. This cross-debate analysis thus allows us to highlight that the position adopted on one debate is strongly correlated with that of a different unrelated debate. Users thus take a stance on emerging debates along pre-existing lines. However, the strength of the associated polarization behavior may vary between debates.

Discussion

The analyses conducted in this work aimed to deconstruct the notion of polarization and refine the the understanding of issue polarization, in particular by analyzing its temporal and sequential evolution depending on the maturity of the debate . Within a population of Twitter users that tends to be more politicized than the average, the identification of several groups of users with specific polarization behavior, provides the potential for the distinction of different crystallized states of polarization. This crystallization is not only expressed by very decided opinions on specific subjects[50], but by a tendency to reduce the information spectrum to an echo chamber or to aggressively refuse all contradiction[45]. It is known to have a major impact on the reliability of opinion surveys, and seems to go through different phases before stabilizing. The temporal analysis contributes to distinguish between "stable" and "unstable" polarized users. We assume that the former are the most ideologically polarized. In this sense, we use the term "politicized". The latter are individuals likely to be polarized on a specific issue. But this may be the result of a transient state linked either to the maturity of the debate, or to the introduction of an event that forces users to reposition themselves. Our work allows an unprecedented identification of periods that contributes to deconstruct polarization: polarization is not a simple state, it is an evolving process. In fact, this work also contributes to better measure the way context-related events can modify the framing of a debate and provoke a systematic dispersion of a subset of Twitter users (intermediate users, affectively polarized) before they gradually polarize, whatever the nature of the debate (mature or emerging) (RQ1). We do not rule out the possibility that this repositioning can be due to the specificity of the Twitter population, i.e. a population particularly sensitive to the effects of media framing. However, our findings extend those from Waller and Anderson[51], that highlight an alignment between online polarization and external events in the US context, to the French-context.

The nature and time of broadcast of the information provided in the debate is therefore crucial, and can have a different effect on users: undifferentiated when the debate is recent (a single intermediate cluster), more targeted when the debate is long-standing (two intermediate clusters) (RQ2). The more sudden, unexpected and unpredictable changes in frameworks, the more destabilized users, who take time to (re)position themselves,

even when the debate is mature (COVID). Our conclusions are in line with Goffmanian "secondary frames" embodied here by media frames. The latter can occasionally overlap with the "primary frames" that enable individuals "in a given situation, to give meaning to this or that aspect"[52], and this can lead to periods of hesitation, confrontation, and repositioning, during which users vacillate between the consensus variants of polarization and compromise[53,54]. Polarization becomes possible again if the elements of the debate become embedded in the primary frameworks forged by the group.

The study of the relationship between the two debates, which seemingly have nothing in common, highlights that when a new debate emerges, users naturally polarize themselves following the same lines as a longer-established debate. However, considering the dynamics and the intensity of polarization, no correlation has been observed. The stance of users on a debate can be predicted based on their behavior on other debates, but not the strength of their conviction.

We wonder how we can take advantage of this unprecedented understanding and characterization of the polarization process to reduce animosity between opposing communities, while keeping users properly informed. Let us recall that our aim is not to make polarization disappear as it contributes to maintaining a healthy democratic debate, but rather to reduce its potentially harmful effects. Furthermore, from an ethical point of view, we cannot manipulate the opinions of social media users, but we can contribute to keep them informed as widely as possible, so that they can make informed decisions in a democratic context. We therefore advocate the implementation of ethical strategies that respect users' beliefs. The literature generally addresses this by diversifying the set of news recommended. However, it has been shown that this simple strategy does not have the same impact across all users and can even reinforce polarization[45,55]. The findings from this work are an opportunity to rethink diversification through multiple strategies. Concretely, the convergence and balanced periods, in which intermediate users appear, are probably the periods during which personalized and diverse recommendations can be proposed by adopting recommendation strategies that differ between intermediate and polarized users. We can for example imagine confronting polarized users with other debates, i.e. providing content diversity, so as to draw their attention to something different and avoid further polarization. However, when confronting users to a wider range of contents or debates it is important to bear in mind that the stance of a user on a specific issue can predict the position of the user on another issue. In addition, recommendations could be adapted to the temporality of the debate, i.e. acting at the most appropriate moment to maintain intermediates with diversified sources of information, thus helping to foster a healthy debate and reduce animosity. Finally, the appearance of a new debate is an opportunity to promote diversity of opinions so that users can quickly understand all the positions expressed on the subject.

Looking ahead, we expect to compare the debates studied with debates offering a clearer political reading. Indeed, both debates - the COVID vaccine and the Ukraine conflict - have been subject to considerable disruption. For the former, the spread of conspiracy theories, and for the latter, strong anti-Americanism, have largely blurred the traditional political cleavages between left and right. It would therefore be interesting to carry out the analysis again by choosing more politically divisive topics (taxation, interventionism versus liberalism, etc.) to better identify periods of repositioning and convergence. It also appears that this temporal analysis will benefit from the study on periods of greater crystallization of opinions, such as election campaigns, which are much more sensitive to the effects of media framing, and over a longer time.

Methods

Data

We used the Twitter API (*v2*), with academic research access to collect data. Our methodology relies on the concept of elite users[56] that represent users who are relevant to the subject matter. We assume that elite users' tweets are in line with their beliefs about the selected debate. Inspired by the methodology of Primario et al.[56], we fix conditions to select legitimate elite users: they need to (1) have a significant number of followers; (2) personally manage their Twitter account; (3) are known by the general audience, through media or government interventions; and (4) are qualified by education and/or profession to address the subject of matter.

Elite users are an effective entry point for collecting data about a specific topic because their opinions are publicly known[56]. Nevertheless, our objective is to analyze the interaction behaviors of standard users, interacting with elite users' tweets. These standard users thus do not meet the 4 criteria cited above (non-elite), and are comparable to everyday users of social media, not publicly known. It is thus necessary to have a dataset that is balanced in terms of opinion carriers, but also representative of standard users' behaviors on social media about a specific debate and during a specific time span. To build such a quality dataset, we carried out several steps, run after having chosen the debates, identified a relevant set of elite users, and defined a collection time span. These steps are as follows: (1) Collect all tweets published by elite users during the predefined period; (2) Filter tweets about the topic of interest; (3) Collect information about a random subset of interacting standard users for each collected tweet; (4) Identify the most active standard users among those selected in Step 3; (5) Collect all interactions of selected standard users on collected elite users' tweets during the defined period; (6) Among all collected interactions, filter those that are related to the tweets collected in step (1).

We collected data about two topics: the COVID-19 vaccine and the Ukraine conflict. Following the procedure detailed above, we manually identified 20 French-speaking elite users having a legitimate voice in the vaccine debate (10 pro-vaccine and 10 anti-vaccine), and 20 other French-speaking users expressing themselves about the Ukraine conflict (10 pro-Ukraine and 10 pro-Russia). Their opinion is known because they have clearly expressed it publicly, and the community to which they relate is therefore unambiguous. To preserve their confidentiality and meet Twitter policy, we do not share the names or usernames of the selected accounts. We collected all elite users' tweets over a 7-month time span, extending from January 1, 2022, to July 31, 2022.

Based on relevant debate-related French hashtags, either for the COVID-19 vaccine or the Ukraine conflict debate, and a random tweet corpus[57], we trained a two-class classifier based on BertTweetFR[58]. This classifier allowed us to keep only elite users' tweets dealing with the selected debates. Here, we focus on retweets, which are signs of approval and thus give information about what users agree with[17].

Following this methodology, we collected information about 100 randomly selected retweeters for each collected tweet, which we hope to be representative of all users. Among the selected retweeters, we focused on the 1,000 most active ones in each debate (500 pro-vaccine/500 anti-vaccine, 500 pro-Ukraine/500 pro-Russia). All in all, the collected dataset contains 299,879 retweets about the COVID-19 vaccine debate (16,791 retweets in the pro-vaccine community, 283,088 in the anti-vaccine community), and 152,802 about the Ukraine conflict debate (41,631 retweets in the pro-Ukraine community, 111,171 in the pro-Russia community), made by a set of 1,000 standard users on 20 elite users' tweets for each debate.

Evaluation of polarization on social media

We rely on three factors to measure polarization. First, we study the opinions that are shared by standard users, according to the communities within which they interact. As we study two bi-community debates, we secondly study the diversity of sources with which they interact in one community, and the diversity of sources with which they interact in the confronting community. We can represent each factor as a probability distribution, specific to each user, and then compute normalized entropy H_N:

$$H_N(X) = \frac{-\sum_x P(x)log(P(x))}{log(n)} \tag{1}$$

where X is a discrete random variable that takes n possible values, and $P(x)$ is the probability of entity x. To ensure that higher computed values are associated with high polarization, we use $H' = 1 - H_N$.

To make the opinion factor gives an indication about the community a user belongs, the opinion factor is oriented[48]. To this end, the sign of the normalized entropy computed based on the probability distribution of interactions within both considered communities is set according to the community within which a user interacts more. We fix positive values for polarization in the pro-vaccine or pro-Ukraine communities, and negative values for polarization in the anti-vaccine or pro-Russia communities. Finally, we then apply the following transformation so that the values lie within [0, 1]:

$$H' = \frac{H^{\pm} + 1}{2} \tag{2}$$

This way, a H' value close to 0 represents a user either closer to the anti-vaccine or pro-Russia community depending on the considered debate, while $H' = 1$ indicates a user closer to the pro-vaccine or pro-Ukraine community. Besides, as the source factors are specific to each community, they are not oriented, and range in [0, 1]. A source factor equal to 0 indicates that the user interacts only with one source in the community. The more she interacts with diverse sources in that community, the higher the source factor.

For the aggregate analysis of polarization, these factors are computed for each user during the entire timespan, between January 2022 and July 2022. For the time-aware analyses of polarization, factors are computed over specific timeframes.

Definition of timeframes for the time-aware analysis

To study the evolution of polarization behaviors, we defined 4-week sliding timeframes, with a 2-week overlap. Over the 7 months of data collection, 15 timeframes are thus formed between January 1 and July 31, 2022. Polarization factors, about opinions and sources, can thus be measured on every successive timeframe.

Of course, some users are not active on all defined timeframes. We have defined a threshold of 20% inactive periods, and users who are inactive more than 20% of the timeframes are removed from the dataset in the time-aware analysis. Among the initial 1,000 users studied in the aggregate analysis, we thus kept 685 users for the COVID-19 vaccine debate, and 784 users for the Ukraine conflict debate.

Identification of polarization behavioral classes

To discriminate between polarization behavioral classes, we use the k-means[59] algorithm. The number of clusters k is optimized by maximizing 2 traditional metrics: Davies–Bouldin Index[60] (the lower the better) and Silhouette Index[61] (the higher the better). The clustering algorithm is applied to the three factors of polarization computed for each user (opinion factor and both sources factors).

However, as we study polarizing debates, some users may remain tightly bunched for some of the factors. This limits the differentiation between them. In order to improve clustering performance, we apply a polynomial transformation having a sigmoid-like pattern to the three factors, as follows:

$$f(x) = \frac{x^a}{x^a + (1-x)^a} \tag{3}$$

The parameter a allows to control the stiffness of the curve and to better discriminate either extreme values or intermediate values. Besides, we assumed that each factor can have different weightings in the evaluation of polarization. This means that factors can have different weights in the computation of the Euclidean distance on which the k-means algorithm relies. Thus, both the parameter a of Eq. (3) and the weights of the three factors were optimized through the clustering process. We selected values that allowed higher clustering performances.

The optimal value of a is $a = 0.5$ for the COVID-19 vaccine debate, while $a = 0.33$ for the Ukraine conflict debate. For both debates, the source factor has a weight of 60%, while the 40% remaining is distributed among the two source factors.

We use the same algorithm and parameters to discriminate between behavioral classes over each timeframe for the time-aware analysis.

The identified clusters are also studied for the cross-debate analysis of polarization. To this end, the set of 170 users interacting with both debates is selected. We compare the polarization behavioral classes of these users between the two debates, carrying out an aggregate and temporal analysis. To evaluate the correlation between categorical variables, corresponding to clusters or community of belonging of users, we use the Cramer's V value[62].

Data availability

Minimal datasets required to replicate the methods presented in this paper are available on a GitHub repository located at: https://github.com/Celina-07/polarization_social_media. Due to the terms of data license agreement signed with Twitter, Inc., all data are not publicly available. Nevertheless, all relevant data are available upon reasonable request to the corresponding author, Célina Treuillier, at the following email address: celina.treuillier@loria.fr.

Code availability

The source code of the study is available at the following public GitHub repository: https://github.com/Celina-07/polarization_social_media. All the analysis were performed using Python[63] v3.9.7, and following libraries: matplotlib[64] v3.5.3, numpy[65] v1.22.4, pandas[66] v1.5.3, plotly v5.4.0,pydlc[67] v0.2., scikit-learn[68] v1.0.2, scipy[69] v1.7.3, statsmodels[70] v0.13.5, tqdm[71] v4.62.3.

Received: 21 September 2023; Accepted: 4 April 2024
Published online: 15 April 2024

References

1. Kubin, E. & von Sikorski, C. The role of (social) media in political polarization: A systematic review. *Ann. Int. Commun. Assoc.* **45**(3), 188–206 (2021).
2. Sunstein, C. R. The law of group polarization. *University of Chicago Law School, John M. Olin Law & Economics Working Paper*, (91) (1999).
3. Carpini, M. X. D. & Keeter, S. *What Americans Know About Politics and Why It Matters* (Yale University Press, 1996).
4. Cramer, K. J. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker* (University of Chicago Press, 2016).
5. Hochschild, A. R. *Strangers in Their Own Land: Anger and Mourning on the American Right* (The New Press, 2018).
6. Saward, M. The representative claim. *Contemp. Polit. Theory* **5**, 297–318 (2006).
7. Huckfeldt, R. Politics in context: Assimilation and conflict in urban neighborhoods (1986).
8. Huckfeldt, R. R. & Sprague, J. *Citizens, Politics and Social Communication: Information and Influence in An Election Campaign* (Cambridge University Press, 1995).
9. Lazarsfeld, P. F., Berelson, B. & Gaudet, H. *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign* (Columbia University Press, 1968).
10. Zuckerman, A. S. *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior* (Temple University Press, 2005).
11. Pattie, C. J. & Johnston, R. J. Conversation, disagreement and political participation. *Polit. Behav.* **31**, 261–285 (2009).
12. Eliasoph, N. *Avoiding Politics: How Americans Produce Apathy in Everyday Life* (Cambridge University Press, 1998).
13. Prior, M. Media and political polarization. *Annu. Rev. Polit. Sci.* **16**, 101–127 (2013).
14. Sîrbu, A., Loreto, V., Servedio, V. D. & Tria, F. Opinion dynamics: Models, extensions and external effects. In *Participatory Sensing, Opinions and Collective Awareness*, 363–401 (Springer, 2017).
15. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**(4), 048301 (2020).
16. Chen, T., Li, Q., Yang, J., Cong, G. & Li, G. Modeling of the public opinion polarization process with the considerations of individual heterogeneity and dynamic conformity. *Mathematics* **7**(10), 917 (2019).
17. Conover, M. *et al.* Political polarization on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* **5**, 89–96 (2011).
18. Guerra, P., Meira, W. Jr., Cardie, C. & Kleinberg, R. A measure of polarization on social media networks based on community boundaries. *Proceedings of the International AAAI Conference on Web and Social Media* **7**, 215–224 (2013).
19. Becatti, C., Caldarelli, G., Lambiotte, R. & Saracco, F. Extracting significant signal of news consumption from social networks: The case of Twitter in Italian political elections. *Palgrave Commun.* **5**(1), 1–16 (2019).
20. Cicchini, T., Del Pozo, S. M., Tagliazucchi, E. & Balenzuela, P. News sharing on twitter reveals emergent fragmentation of media agenda and persistent polarization. *EPJ Data Sci.* **11**(1), 48 (2022).
21. Garimella, K., Smith, T., Weiss, R. & West, R. Political polarization in online news consumption. *Proceedings of the International AAAI Conference on Web and Social Media* **15**, 152–162 (2021).
22. Cinus, F., Minici, M., Monti, C. & Bonchi, F. The effect of people recommenders on echo chambers and polarization. *Proceedings of the International AAAI Conference on Web and Social Media* **16**, 90–101 (2022).
23. Geschke, D., Lorenz, J. & Holtz, P. The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *Br. J. Soc. Psychol.* **58**(1), 129–149 (2019).
24. Iyengar, S., Sood, G. & Lelkes, Y. Affect, not ideology: A social identity perspective on polarization. *Public Opin. Q.* **76**(3), 405–431 (2012).
25. Wagner, M. Affective polarization in multiparty systems. *Electoral Stud.* **69**, 102199 (2021).
26. Jost, J. T., Baldassarri, D. S. & Druckman, J. N. Cognitive-motivational mechanisms of political polarization in social-communicative contexts. *Nat. Rev. Psychol.* **1**(10), 560–576 (2022).
27. Valensise, C. M., Cinelli, M. & Quattrociocchi, W. The drivers of online polarization: Fitting models to data. *Inf. Sci.* **642**, 119152 (2023).
28. Pariser, E. *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think* (Penguin, 2011).

29. Castle, J. J. & Stepp, K. K. Partisanship, religion, and issue polarization in the united states: A reassessment. *Polit. Behav.*, 1–25 (2021).
30. Moscovici, S. & Zavalloni, M. The group as a polarizer of attitudes. *J. Pers. Soc. Psychol.* **12**(2), 125 (1969).
31. Goldstein, J. "Moral contagion": A professional ideology of medicine and psychiatry in eighteenth-and nineteenth-century France. *Prof. French State* **1**, 700–1900 (1984).
32. Braconnier, C., Coulmont, B. & Dormagen, J.-Y. The heavy variables are still alive and kicking. *Revue Française de Science Politique* **67**(6), 1023–1040 (2017).
33. Katz, E., Lazarsfeld, P. F. & Roper, E. *Personal Influence: The Part Played by People in the Flow of Mass Communications* (Routledge, 2017).
34. Boyadjian, J., Neihouser, M., Skoric, M., Parycek, P. & Sachs, M. Why and how to create a panel of twitter users. In *CeDEM Asia 2014: Conference for E-Democracy an Open Government*, 247–252. Donau-Universität Krems Krems (2014).
35. Walker, M. & Matsa, K. E. News consumption across social media in 2021 (2021).
36. Barberá, P. Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, vol. 34 (2020).
37. Russell Neuman, W., Guggenheim, L., Mo Jang, S. .a. & Bae, S. .Y. . The dynamics of public attention: Agenda-setting theory meets big data. *J. Commun.* **64**(2), 193–214 (2014).
38. McCombs, M. E. & Shaw, D. L. The agenda-setting function of mass media. *Public Opin. Q.* **36**(2), 176–187 (1972).
39. Marchand-Lagier, C. & Weill, P.-E. "How" silent citizens" perceive Europe? (2011).
40. Hallin, D. C. & Mancini, P. *Comparing Media Systems: Three Models of Media and Politics* (Cambridge University Press, 2004).
41. Norris, P. Preaching to the converted? pluralism, participation and party websites. *Party Polit.* **9**(1), 21–45 (2003).
42. Helberger, N., Karppinen, K. & D'Acunto, L. Exposure diversity as a design principle for recommender systems. *Inf. Commun. Soc.* **21**(2), 191–207 (2018).
43. Heitz, L. *et al.* Benefits of diverse news recommendations for democracy: A user study. *Digit. Journal.* **10**, 1710–1730 (2022).
44. Joris, G. *et al. News Diversity and Recommendation Systems: Setting the Interdisciplinary Scene* 90–105 (Springer, 2020).
45. Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**(37), 9216–9221 (2018).
46. Garimella, K., Morales, G. D. F., Gionis, A. & Mathioudakis, M. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the Web Conference* (2018).
47. Nguyen, C. G. *et al.* The impact of emotions on polarization. Anger polarizes attitudes towards vaccine mandates and increases affective polarization. *Res. Politics* **9**(3), 20531680221116572 (2022).
48. Treuillier, C., Castagnos, S. & Brun, A. A multi-factorial analysis of polarization on social media. In *UMAP'23* (Limassol, 2023).
49. Druckman, J. N., Peterson, E. & Slothuus, R. How elite partisan polarization affects public opinion formation. *Am. Polit. Sci. Rev.* **107**(1), 57–79 (2013).
50. Krueger, T., Szwabiński, J. & Weron, T. Conformity, anticonformity and polarization of opinions: Insights from a mathematical model of opinion dynamics. *Entropy* **19**(7), 371 (2017).
51. Waller, I. & Anderson, A. Quantifying social organization and political polarization in online platforms. *Nature* **600**(7888), 264–268 (2021).
52. Goffman, E. *Frame Analysis: An Essay on the Organization of Experience* (Harvard University Press, 1974).
53. Wehman, P., Goldstein, M. A. & Williams, J. R. Effects of different leadership styles on individual risk-taking in groups. *Hum. Relat.* **30**(3), 249–259 (1977).
54. Jesuino, J. C. Influence of leadership processes on group polarization. *Eur. J. Soc. Psychol.* **16**(4), 413–423 (1986).
55. Treuillier, C., Castagnos, S., Dufraisse, E. & Brun, A. Being diverse is not enough: Rethinking diversity evaluation to meet challenges of news recommender systems. In *Fairness in User Modeling, Adaptation and Personalization (FairUMAP 2022)* (2022).
56. Primario, S., Borrelli, D., Iandoli, L., Zollo, G. & Lipizzi, C. Measuring polarization in twitter enabled in online political conversation: The case of 2016 US presidential election. In *2017 IEEE International Conference on Information Reuse and Integration (IRI)* 607–613 (IEEE, 2017).
57. Turenne, N. The rumour spectrum. *PLoS One* **13**(1), e0189080 (2018).
58. Guo, Y., Rennard, V., Xypolopoulos, C. & Vazirgiannis, M. BERTweetFR : Domain adaptation of pre-trained language models for French tweets. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)* (Online) 445–450 (Association for Computational Linguistics, 2021).
59. Likas, A., Vlassis, N. & Verbeek, J. J. The global k-means clustering algorithm. *Pattern Recognit.* **36**, 451–461 (2003).
60. Davies, D. L. & Bouldin, D. W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **2**, 224–227 (1979).
61. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
62. Cramér, H. *Mathematical Methods of Statistics* Vol. 26 (Princeton University Press, 1999).
63. Van Rossum, G. *et al. Python Reference Manual* Vol. 111 (Centrum voor Wiskunde en Informatica, 1995).
64. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**(3), 90–95 (2007).
65. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
66. McKinney, W. *et al.* pandas: A foundational python library for data analysis and statistics. *Python High Perform. Sci. Comput.* **14**(9), 1–9 (2011).
67. Moritz, D. & Fisher, D. Visualizing a million time series with the density line chart. arXiv preprint arXiv:1808.06019 (2018).
68. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
69. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. & SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
70. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference* (2010).
71. da Costa-Luis, C. O. TQDM: A fast, extensible progress meter for python and CLI. *J. Open Source Softw.* **4**(37), 1277 (2019).

Acknowledgements

This research was supported by French ANR BOOM project (Modeling and Opening Opinion Bubbles) (ANR-20-CE23-0024).

Author contributions

C.T., S.C., and A.B. designed the research, C.T. collected the data, prepared the data for analysis, performed the computational analysis, C.T., S.C., and A.B. analyzed the data, C.T., C.L., and A.B. wrote the first draft of the paper, C.T., S.C, C.L., and A.B. were involved in the writing process and provided critical revisions.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to C.T.

Reprints and permissions information is available at www.nature.com/reprints.

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