NORDIS – NORdic observatory for digital media and information DISorders

An analytical report on online information propagation: A description of the methods applied to characterize the propagation of stories, with case studies

Date: 31-05-2023

Final version

NORDIS

Action No:	2020-EU-IA-0189
Project Acronym:	NORDIS
Project title:	NORdic observatory for digital media and information DISorders
Start date of the project:	01/09/2021
Duration of the project:	24
Project website address:	https://datalab.au.dk/nordis
Authors of the deliverable	Matteo Magnani, Victoria Yantseva, Davide Vega, Uppsala University
Activity number	Activity 5
Task	Task 5.3

Reviewers: Marina Charquero Ballester (Aarhus University, Denmark)

Funding:
The research was funded by EU CEF grant number 2394203.

1 Executive summary

This report describes a new approach to study the diffusion of stories posted on media sites and shared on social media. A typical approach to study this is to look at the features of the individual social media posts including links to those stories, e.g. number of shares or sentiment of individual posts. We propose to organise social media content into coherent sets of posts and their relations (e.g. replies), that we call conversations, and look at the association between stories from different types of sources and features of the conversations where they have been shared. These features include structural measures (length and pace of the conversation), the overall sentiment of the conversation, and user stances. We also look at changes inside individual conversations happening after stories are shared. We test our method on a large number of conversations about immigration from a major Swedish forum. The empirical analysis focuses on stories from the mainstream and alternative media sites, and reveals that conversations where alternative media content is disseminated tend to have a higher share of messages with negative evaluations of the immigration agenda, but also tend to wane more quickly and be shorter; at the same time, our results call into question the effects of exposure to and consumption of alternative media content: content from both mainstream and alternative media is shared by the same sets of actors, and we can only observe limited differences between conversations with no links, with links from mainstream media, and with links from alternative media.

Contents

2 Introduction

This report introduces a method to study online information diffusion, that is then used to investigate the dissemination of immigrant-critical alternative media content and to look at the relation between the sharing of such content, the dynamics of the online conversations where the content is shared, and the users' evaluations of the immigration issue in Sweden. The general question underlying this work is thus: what happens when specific content is shared as part of an online conversation? In particular, we:

- explore the differences in user stances and structure of the conversations depending on the dominant type of content shared, and
- study the evolution of conversations with respect to their pace and the emotional states of the conversation participants.

The question of alternative media content diffusion in online environments, especially the right-leaning and fringe ones, is important for several reasons: because such online environments have been shown to enable offline mobilization or support for protest participation [40, 65]; because they may provide a breeding ground for the spread of not only alternative media content, but also disinformation and conspiracy theories [32, 59]; and, finally, because the consumption of content from right-wing alternative media has been shown to have a profound effect on immigration attitudes [55]. This research complements existing studies on the formation of collective emotions, diffusion of content from alternative media in online networks, and the evolution of conversations online. Appendix A provides an overview of this literature.

The main objective of this report is to extend the existing evidence on the impact of alternative media on the development of collective emotions by looking at the problem in conjunction with the conversation dynamics. One may suggest that the latter inevitably affect the way users perceive and react to the information provided as part of the conversation. In other words, users not only react to alternative news content shared but also take into account the previous content of the conversation, something that has mostly been overlooked by the existing research that has mainly focused on retweet networks or engagement with information from user feeds. Further, we aim to contribute to the existing research by investigating the patterns of circulation of alternative media content *inside* an already segregated and fringe right-leaning environment, which is especially relevant given the proliferation of such communities in the online space. This is important due to the fact that the existing research has mostly provided evidence for the circulation of trusted versus problematic content and resulting user clustering in social networks in general, thereby paying less attention to the environments that may be more susceptible to such information, and where such information may have a stronger effect on users' perceptions. Finally, the direct outcomes of alternative news content sharing beyond general sentiments and network clustering may be difficult to evaluate. However, with this respect, we provide a contribution by using a custom stance classification model, specifically developed for the online data source we use in this study, which makes us able to identify how users' evaluations of the immigration topic change after being exposed to content from alternative media.

Our method draws on the sociological literature on emotional group dynamics, and in particular, Randall Collins' interactional ritual theory (IRC) [11] that provides a framework to understand users' interactions on the forum and link sharing as specific types of social actions to achieve group solidarity and emotional synergy. In particular, it accommodates the idea that conversations in anti-immigrant environments represent not only instrumental but also symbolic and value-oriented behaviour. Moreover, it represents conversations as chains of interactions resulting in emotional and behavioural contagion, and, at the end of the day, ritual participants' emotional (or affective) mobilization, or the production of "emotional energy" in theory's terms [11].

The report is organised in three main parts: the method, two empirical studies performed using our approach based on conversation analysis, and a discussion of the main results and limitations of the method and studies. The method is described under the assumption that the data is extracted from an online forum, but can be adapted to other types of social media sites. We also include background information in two appendices, respectively providing an overview of related literature and the theoretical framework underpinning our work. Data collection and further research with its use have been approved by the Swedish ethics review authority (Etikprövningsmyndigheten).

3 Method

Our method includes the following steps:

1. Extraction of edge lists and reconstruction of conversation networks;
2. Extraction and coding of URLs;
3. Stance classification and sentiment analysis;
4. Statistical and network analysis.

3.1 Conversation reconstruction

Our two primary assumptions for the extraction of conversations are that, on the core level, a) a discussion thread in an online forum consists of one or several conversations, depending on the frequency of interactions; and b) that a conversation can be represented either as a temporal network consisting of interconnected messages or, on an aggregated level, of users (nodes) and connections (edges) between them. However, in contrast to offline conversations that share the same spatial and temporal dimensions, online interactions are asynchronous. One consequence is that forum users do not all participate in the conversations simultaneously, which affects the way conversation networks can be created.

To provide an example, if user A posts a message at time t_1, and user B at time t_2 (within a predefined time interval after A's message), then we assume that user B has read user A's message and has potentially taken its content into account while writing their own message. Thus, we create a link from B's to A's message. Additionally, we create also a link from B's message to

any other message quoted in it (similar to a reply on Twitter), irrespective of the amount of time that has passed, since this indicates a direct reference to the message previously published.A similar approach has already been used in the literature to reconstruct information flows [43].

Following this procedure, we assume that a conversation, on an aggregated level, tends to capture the amount of information from a given thread that a given user will consume. Moreover, our approach to conversations takes into account the sequencing of messages, which helps to account for the conjoint evolution of users' narratives and can be related back to Collin's theory: "The world is a network of conversations, and what people think at any point in it is a product of what has circulated in previous conversations" [14, p.304].

3.2 Classification of stances and sentiments

In this work, we not only aim to identify users' sentiments in general, but also sentiments on and evaluations of a specific topic (immigration and corresponding policies in Sweden). To that end, we use a notion of *stance* and follow the definition of stance-taking as a subjective evaluation or appraisal of a specific target [19, p. 142, 153], "explicitly mentioned or implied within the text" [37]. Although this task is closely related to that of classic sentiment polarity detection, the difference between the two is that sentiment analysis is concerned with identifying the sentiment polarity of a text in general, whereas stance classification requires a specific target (be it an entity, topic, claim etc.). For instance, the following statement "it is frustrating that we still need to fight against implicit ethnic discrimination of newcomers in our country" can be described as having negative sentiment polarity, but at the same time taking a positive or supportive stance with regard to the immigration topic.

For the empirical part of this work, where we aim to identify user stances with regard to the immigration issue, we use a two-step machine learning approach trained specifically on the data from the Flashback forum. The details on model training and selection are discussed in [70]. In short, on the first step, our approach identifies on-topic and off-topic messages, since, as mentioned above, we are only interested in messages discussing immigration. On the second step, particular stances are determined (two classes: negative and non-negative). The reason for choosing a two-class approach is that positive messages are under-represented in the corpus (5% of the training dataset). However, the reason for using a two class-approach is not only the need to compensate for class imbalance but also the fact that, for us, it is more important to distinguish negative rather than moderate expressions about immigration, whether positive or neutral. Stance classification is further supplemented with unsupervised sentiment analysis using VADER dictionary [30].

3.3 Identifying immigrant-critical alternative media

We use the definition of "immigrant-critical alternative media" and use a relational approach of Kristoffer Holt suggesting that "Alternative news media represent a proclaimed and/or (self-)perceived corrective, opposing the overall tendency of public discourse emanating from what is perceived as

the dominant mainstream media in a given system." [29]. Other important traits include alternative news producers, in particular, non-professional actors, such as readers and activists; alternative news content, or the narratives that are perceived as counter-hegemonic and marginalised in the mainstream discourse; and, finally, alternative news organisations, in particular, in low-cost formats, such as blogs and webpages [29]. In contrast, we define *mainstream media* as established news organisations that rely on the work of professional journalists, adhere to ethical and professional journalistic standards and have an editor or an editorial board responsible for the content produced.

For resource labelling, we extract the most used domains mentioned by the users, and we exclude the resources that do not belong to the media category (for instance, websites of private companies, non-profit organisations and government agencies). The rest of the resources are evaluated manually based on a range of criteria following the approach described in [29], in particular a) non-oppositional or corrective stance in relation to what is perceived as mainstream immigration discourse; b) availability in high-cost formats (radio, printed press, TV); c) adherence to the press ethical norms and standards; d) a presence of editor-in-chief and/or editorial board, or similar actors responsible for the published content, e) belonging to a larger organisation or publisher.

Using these five criteria, the resources are assigned one of the three labels: *mainstream/legacy media*, *alternative media* and *other media*. The *mainstream media* label is assigned to the resources that are evaluated positively on the criterion *a)* (as providing non-oppositional framing of the public agenda), plus on at least three other criteria. The procedure for labelling resources as *alternative media* is exactly the opposite: those are evaluated negatively on the criterion *a)* plus on at least three other criteria. The rest of the resources are labelled as *other media*, since the distinction between mainstream and alternative media is rather blurred and represents a continuum, as noted by the previous research [29], rather than a set of distinct media format categories. Finally, it needs to be mentioned that social media groups on popular platforms (e.g. Twitter or Facebook) are excluded from our classification since they accounted only for a minor share of links disseminated on the forum.

3.4 The dynamics of collective stances and emotions and the role of alternative media

As a first step, we compare the characteristics of a) conversations with alternative media versus mainstream media links shared, and b) conversations with any type of links versus no links shared. To that end, we use Mann-Whitney U test that fits well to compare samples where data is non-normally distributed. The conversations are evaluated on a range of criteria, including:

- duration in hours;
- the ratio of messages with negative and non-negative stances;
- average stance probability for each stance class;

- average clustering coefficient or the fraction of possible triangles averaged to the number of nodes in the network;
- the number of nodes and edges, as specified in Section 4.1;
- edge density coefficient, or the fraction of existing ties to the possible hypothetical ties.

Further, we explore conversation entrainment by analyzing the conversation pace and the evolution of user sentiments. In other words, a successful ritual in theory's terms can be measured in two different ways: first, by the frequency of interactions and conversation pace, and, second, by the evolution of sentiments or stances, which, in the case of fringe or right-wing environments, can be denoted by the dominating negative moods and evaluations of the immigration topic. Even more so, a successful conversation may also imply the alignment of users' pace of interactions and subjective expressions, so that each new message arriving as part of the conversation will be dependent on the previous message(-s) posted as part of the same interaction chain. To measure such conversation entrainment, we use two mixed effects models with conversations serving as a random effect, aiming to detect whether there is an autocorrelation between the subsequent messages in each conversation.

In the conversation pace model, we calculate the time intervals between each pair of messages. For example, if message A is published at 15:30, message B at 15:40, and message C at 15:50, then we consider that message C was published at Δt_3 within the interval of 10 minutes. Since we are also interested in checking whether users adapt their interaction pace to that of the conversation in general (or, in other words, if there is a temporal alignment in users' interactions), we also take into account the interval at which the previous message was posted (that is, a lagged time interval Δt_2), which in our example is also ten minutes for message B.

We evaluate three models: a null model with a random conversation effect only, a model that also considers the message posting intervals in the preceding steps and a model that also accounts for the links shared in the previous steps. Thus, we include the message's posting interval Δt as a dependent variable and the previous message's posting interval at Δt_{-1} as an independent variable. As mentioned above, our hypothesis is that users adapt their behaviour to that of conversation, and, if interactions happen at a fast pace, then users will also be trying to adapt to the pace of preceding interactions. Finally, we add link sharing (any links, mainstream media links or alternative media links) at t_{-1} as an independent (dummy) variable. Specifically, we test whether links shared as part of the previous interactions have an impact on users' interaction pace. For example, one could hypothesize that users may need more time to check the link content, and thus the conversation pace will be slower, or, on the opposite, that the distributed content would make users more engaged in the conversation and thus more willing to provide a more prompt reply.

In our sentiment model, we use the results of VADER sentiment analysis to explore whether users' sentiments are impacted by the sentiments of posts previously shared in the conversations. For example, if message A has a VADER score of –0.5, and message B a score of 0.5, then we consider the score of message B $\mathcal{V}_2$ is 0.5, and also take into account the score of a preceding

message A $\mathcal{V}_1$, which is –0.5. This is due to the fact that we are interested in checking users' emotional alignment, or, in other words, whether users adapt their message sentiments to that of the conversation in general, and preceding messages specifically.

In the sentiment model, we also evaluate three models: a null model with a random conversation effect only, a model that also considers the message sentiments in the preceding steps and a model that also accounts for the links shared in the previous steps. Thus, we include the message's sentiment $\mathcal{V}$ as a dependent variable and the preceding message's sentiment $\mathcal{V}_{-1}$ as an independent variable. This model allows evaluating whether there is an impact of preceding message sentiments on the subsequent ones, or, in other words, we test whether there is a co-evolution of user sentiments in the conversations. As in the conversation pace model, we add again a random effect of the media links shared in the preceding messages (any links, mainstream media links or alternative media links) at t_{-1}. In the sentiment model, however, we test whether media links shared in the preceding steps of conversation have an impact on the evolution of user sentiments.

4 Case Studies

In this section we describe the application of our method on data we collected from a major Swedish forum. The first study looks at the relation between the type of shared stories, in particular, stories shared from alternative or mainstream/legacy media, and features of the conversations. The second study focuses on the relation between the types of shared stories and the *evolution* of conversations, that is, how individual conversations change after a story is shared inside them.

4.1 Data

In the following studies, we explore the dynamics of conversations on Flashback, a Swedish online forum with more than 1.5M registered users as of February 2023, which makes it one of the biggest online platforms in Sweden. While Flashback claims to ensure freedom of speech and users' anonymity, it has also been used for personal attacks and other types of problematic online behaviours [48]. Here we focus on the "Immigration and integration" [Integration och invandring] subforum, where many immigration-related discussions occur. This subforum's case is particularly interesting since previous research has demonstrated that biased and Islamophobic narratives are quite common on the platform [33, 57], which makes it a relevant object of study as an example of a fringe online community.

The data, consisting of 270k messages from the forum spanning the last election cycle in Sweden (2018-2022), were collected in December 2022 using the R package *httr* [69]. In particular, we scraped all discussion threads in the aforementioned subforum that were posted during the last election cycle starting from the 18th of January 2019, when the new Prime minister was chosen, until the 12th of August, 2022, which was the last day for the parties to register their participation in the elections. The choice of this time frame was dictated by the assumption that elections not

only change users' interaction patterns but also the ways they talk about immigration, which is a pressing and highly debatable topic in the Swedish context.

Collected data included the following information: thread titles, usernames of the message authors and quoted users (similar to the reply function on Twitter), as well as message IDs, texts and dates of publications. In total 270k messages from 4692 discussion threads were collected. These threads resulted in 9304 conversations (rolling window of 27 minutes[1]), with an average conversation length of 28.4 and a median length of only 5 messages respectively, suggesting that the vast majority of conversations are rather short. From these data we also extracted 332 unique domains, accounting for 1.5% of all unique domains but 75% of all links shared on the forum.

Table 1 summarizes the main properties of the collected dataset. To preserve users' integrity, the analysis has been performed on the group (aggregated) level, and we avoided pointing out individual users in the manuscript text.

Table 1: *Dataset summary*

Number of posts	276716
Number of threads	4692
Avg. number of posts per thread	59
Number of unique users	14366
Avg. number of posts per user	59

4.2 Study 1: Alternative media in the forum conversations

Regarding our analysis of alternative media in the reconstructed conversations, our results indicate that, quite unexpectedly, links to alternative media (hereafter AM) comprise only a minor part of the links shared on the forum: only 12% of all conversations on the forum has at least one AM link shared, and more than every third one (34%) at least one mainstream media (hereafter MM) link shared. Likewise, of more than 14000 unique users in the dataset, only a small part (8% or 1147 participants) is responsible for the dissemination of alternative media content.

Speaking of particular sources shared, we note that mainstream media are circulated far more actively than alternative ones, which is somewhat striking since one would expect the forum's audience to favour the sources that would problematize the existing immigration discourses. Nevertheless, we find that the shared resources cover the whole ideological spectrum with regard to the immigration agenda: from the more "conventional" narratives articulated by the largest national publishers (such as *SVT* and *Aftonbladet*) and all the way to the extreme right resources that at

[1]We note that more than two-thirds of all messages (third quartile) are published within the rolling window of 27 minutes, therefore, we created a link between any pair of messages in a thread that is published within this time span. We validated the robustness of this rolling window choice by testing thresholds that are two times shorter (13 minutes) and two times longer (54 minutes) to account for the fact that some of our results may depend solely on the choice of the rolling window. Further, we replicate the whole analysis for the conversations created using the three rolling windows and confirm that our results hold irrespective of the chosen threshold (unless specifically mentioned in the next sections).

times resort to hateful and extremist rhetoric when talking about immigration and immigrants in Sweden (for instance, *Nordfront*). In between these two extremes of the spectrum can also be found resources that were assigned to ”other category”, such as, for example *Nyheter Idag*, since they did not fully comply with our understanding of legacy media, but which at times are successful at mimicking the behaviour and publishing strategies of the legacy outlets. Somewhere in the middle of this spectrum can also be found international resources such as *Sputnik News* and *Russia Today* that clearly stand out in their interpretation of political and policy issues.

Table 2: Most frequently shared media sources

Alternative media	**Number of shares**	**Mainstream media**	**Number of shares**
samnytt.se	1131	svt.se	2916
friatider.se	1095	expressen.se	2379
uvell.se	115	aftonbladet.se	1993
petterssonsblogg.se	98	sverigesradio.se	1253
detgodasamhallet.se	88	svd.se	888
nordfront.se	88	dn.se	870
snaphanen.dk	77	gp.se	611
swebbtv.se	44	sydsvenskan.se	310
thereligionofpeace.com	38	svtplay.se	262
breitbart.som	37	hemhyra.se	198

Our assumption is that conversations where alternative media content is shared exhibit higher user activity since such content would spark conversations and provide some ground for users' discussions, as well as reinforce and, potentially, intensify users' already existing negative preconceptions about immigration and immigrants. Since we deal with non-normally distributed data, we use the Mann-Whitney U test to check this hypothesis and compare the means of conversation samples where alternative versus mainstream media content is shared. Since different conversations have varying ratios of links shared, we perform a range of tests using different link ratio thresholds (0.2, 0.4 and 0.6).

When it comes to the characteristics that differentiate conversations with alternative links from those with mainstream ones, we detect substantial differences in user stances on the immigration topic. In particular, conversations where alternative media content is shared indeed tend to have a higher ratio of negative messages. At the same time, quite a counter-intuitive observation is that conversations with alternative media content also tend to be shorter, with a lower number of nodes

and edges between them, but, at the same time, higher edge densities ($p < 0.05$ in both cases irrespective of the chosen threshold for the link ratios).

We also decide to test an alternative hypothesis — for instance, one can suggest that link sharing in general, irrespective of link type, can affect the properties of conversations. Indeed, following the same procedure as the one described above, we find that any link sharing makes the conversations longer, and leads to a higher edge density and a higher number of nodes and edges involved (Mann-Whitney U test, $p < 0.05$ irrespective of the chosen threshold for the link share in the conversations). Thus, one may suggest that content sharing can serve as a factor to spark conversations and users' deliberations, irrespective of the type of the cited source.

Our last observation is also that conversations with any link types tend to have a lower share of negative evaluations of the immigration agenda in comparison with the conversations where no media content is shared ($p < 0.05$ irrespective of the threshold for the link ratios). Such conversations also tend to have a higher share of neutral and positive messages, which, however, holds only for conversations with a sufficiently high share of links in the conversation ($p < 0.05$ for hyperlink ratio over 0.4, irrespective of the rolling window). Finally, we find that there are not only differences in the users' stance polarities but also in their strengths denoted by the probabilities of belonging to a given category (negative or non-negative) calculated by the stance classification model. In particular, in the conversations with any type of links shared, those are lower for the negative messages and higher for the non-negative ones ($p < 0.05$ irrespective of the threshold for the link ratios).

4.3 Study 2: Alternative media and conversation evolution

In this study, we evaluate whether the preceding conversation state at (t_{-1}) can be used to predict the subsequent conversation state at (t), which denotes conversation entrainment and users' alignment with regard to the conversation pace and shared sentiments. On top of this, we evaluate whether link sharing (alternative media, mainstream media as well as any links in general) affects conversation entrainment, which is, once again, expressed by the conversation pace and user sentiments. In total, 6674 conversation chains were included in the models. The minimum number of observations per conversation was 2 due to the need to include lagged observations in the models.

Most importantly, our results indicate that conversation entrainment is more visible if we consider the pace of interactions rather than message sentiments: in particular, the sentiments of the previously shared messages serve as significant ($p < 0.05$) but weak predictors for the subsequent message's sentiments, suggesting weak dependence between the sentiments of messages in the same interaction chains. More specifically, a model that includes lagged message sentiments as a dependent variable performs only marginally better than the null model with random intercept only. This suggests that users' emotional alignment is less discernible in comparison with conversations' temporal alignment. It also needs to be noted that, in this case, we use only the results of VADER sentiment analysis, which is a general sentiment analysis tool not capable of capturing the differences in users' moods targeting the immigration topic specifically.

In contrast, our model for the conversation pace is more successful in capturing the temporal relationship between the messages, and we find that the pace of preceding interactions can be used to predict the pace of subsequent interactions. Our model with lagged message posting intervals as an independent variable performed significantly better than the null model with random intercepts only (ANOVA test, $chi^2 = 12294$, $p < 0.05$, AIC 92216 versus AIC 104507 respectively). In particular, lagged posting time intervals can serve as a significant predictor for the posting intervals of the subsequent messages (Estimate = 0.221, SE = 0.002, $p < 0.01$). We also note the positive estimate for the predictor, suggesting that conversations' paces get longer with time, which also denotes a decaying pattern of the interaction pace.

Turning back to the question of the role of alternative media content in the conversations, we notice, once again, that link sharing, whether from alternative media or any other type of resources, does not have any sufficient effect on the conversation dynamics neither with regard to the conversation pace nor the dominating sentiments ($p > 0.05$ in both models). In our data, link sharing, irrespective of link type, had a significant impact on the conversation dynamics only when used as a predictor for conversation pace ($p < 0.05$), however, it did not improve model performance (AIC 92193.8, conditional R^2 0.135 for the model with time-lagged post intervals data only versus AIC 92189.3, conditional R^2 0.135 for the model also accounting for previously shared links). Thus, our results indicate that link sharing plays an insignificant role in the evolution of conversation chains.

5 Discussion

5.1 Theoretical foundations of the method

Our method is based on the sociological literature on emotional group dynamics, and in particular, Randall Collins' interactional ritual theory (IRC) [11] that provides a framework to understand users' interactions on the forum and link sharing as specific types of social actions to achieve group solidarity and emotional synergy. In particular, it accommodates the idea that conversations in anti-immigrant environments represent not only instrumental but also symbolic and value-oriented behaviour. Moreover, it represents conversations as chains of interactions resulting in emotional and behavioural contagion, and, at the end of the day, ritual participants' emotional (or affective) mobilization, or the production of "emotional energy" in theory's terms [11].

At the same time, the method is based on previous work proposing a change of perspective in the analysis of social media data, shifting the focus from individual posts to conversations [42, 62, 63].

5.2 Main empirical takeaway

Summing up the results above, our main takeaway from the analysis is that Flashback's audience circulates and consumes different kinds of digital content that covers the whole ideological spectrum with respect to the dominating immigration discourses, from large national mainstream publishers to radical far-right groups openly disseminating biased and racist narratives. This ob-

servation is consistent with the earlier results that provided evidence that even far-right right social media group users consume diverging kinds of content [27]. Nevertheless, it is noteworthy that legacy media content is disseminated in an environment where mainstream narratives are considered invalid, as one would expect the platform's audience to favor the resources that provide what is considered as the alternative and counter-hegemonic view of the immigration topic. Nevertheless, our conclusions confirm the earlier results that users even in similar fringe environments consume cross-cutting content, while the platforms themselves can nevertheless serve as spaces for deliberative talk [45].

Further, the use of mainstream resources in this setting can be explained, for instance, by the mechanisms of selective exposure and confirmation bias, whereby users might want to select and strategically use the content that corresponds to their pre-existing views (demonstrated by e.g. [35]), irrespective of where it was published. Yet another explanation can be that such mainstream narratives are simply denied and are used to exemplify a presumably biased representation of the immigration agenda by the mainstream media. However, since we lack the direct evidence to support this argument, one possible way to extend the existing analysis on mainstream media use in such fringe environments is to analyze what kinds of argumentation and reasoning (for instance, approval or contradiction) are used by the forum users to comment on the content originating from the mainstream media.

Nevertheless, despite the fact that we do not notice any users' preferences for specific types of media content, in accordance with our expectations, we find that conversations where alternative media content is disseminated differ from those with the content from the legacy media in that they tend to have a higher share of messages with negative evaluations of the immigration agenda. However, this does not necessarily imply that such content *causes* more negative expressions. On the contrary, one may suggest that alternative media sharing may correlate with users' preconceptions: in other words, those who have negative pre-existing attitudes to the immigration policies in Sweden may be more likely to circulate alternative media content in their messages, but this does not necessarily mean that those attitudes are adapted by other participants. Another counter-intuitive observation is that alternative media content has in a way a de-activating function with respect to the users' engagement on the platform: although users' expressions in such conversations tend to be more negative, the conversations with alt-right content tend to wane quicker and be shorter. This may partly be because they might help to quickly reach a consensus, or because they are distributed by the users who have peripheral roles on the subforum since we also find that only a small minority of users is responsible for the dissemination of the links leading to alternative media content. Thus, a possible avenue for future research is to explore user centrality and influence with regard to the consumption and dissemination of specific types of content.

The existing research has also suggested that online platforms may enable users' radicalization [50, 47] and provide discursive opportunities for right-wing violence [65], however, particular low-level mechanisms through which online platforms enable users' radicalization remain underinvestigated. In this study, we evaluated conversation dynamics, as well as users' emotional and temporal alignment in particular, as possible factors that may pave the way for users' more extreme views with regard to immigration policies. However, we have not found any evidence that

conversations on Flashback´s subforum generate more extreme or negative expressions about the topic. On the other hand, this observation partly speaks in favour of the earlier evidence that the emotional tone of online discussions can be best described as stable rather than fluctuating towards particular extremes [23], which highlights the need to explore alternative mechanisms through which the users become indoctrinated into more negative narratives.

Our observations also call into question the effects of exposure to and consumption of alternative media content, and, in our empirical case, we notice no such effect with respect to the development of conversations, which also has some implications for the debate about the effect of alternative media content on its consumers and the effect of participation in the online fridge political groups. Specifically, we have focused on the distribution of different kinds of content, including mainstream or alternative media, as a particular conversation driver. We found that sharing any type of links has a limited effect only on users' temporal alignment in the conversations, which is the only case when this effect is significant. Thus, turning back to the explanation offered by Collins' IRC theory, link sharing can primarily be represented as an element of the collective symbol system [51], which is enabled by Flashback as an online platform, on par with, for example, specific jargon used to talk about immigrants (see e.g. [1]).

Further viewing our results through the prism of Collins' theoretical approach, we have been able to detect what Collins describes in terms of participants' entrainment, which, for him, is a necessary condition for a successful ritual. While we have evaluated two different ways to approximate such conversation entrainment, namely temporal alignment denoted by conversation pace and emotional alignment denoted by the dominating message sentiments, we find that temporal alignment seems to be a more straightforward way to uncover ritual entrainment. In other words, our interpretation is that, in online rituals, synchronizing conversation rhymes is much easier than generating common moods, while the former nevertheless serves as an important condition for the ritual to occur. This observation is supported by Collins' own remarks regarding the limits to which online rituals can generate emotional energy. Moreover, our results to some extent correspond to those of previous work that primarily focused on conversation length as a successful predictor of ritual's success [18], which, once again, speaks in favour of users' temporal alignment as one of the important indicators for the development of online interactions.

Our last important point is that far from all conversations can be regarded as successful rituals: while many of them are simply too short, others become neither temporally nor emotionally aligned and, thus, cannot be regarded as successful. Thus, future work may focus on the conditions that differ between those different types of conversations on online communication platforms.

5.3 Limitations

One of the constraints with regard to the chosen methodological approach is of course that only a subset of the most frequently shared links was labelled, which might have left out some alternative media sources that are much smaller in terms of readership in comparison with established media. The same goes also for social media groups and accounts that also generate online content, however, we note that their prevalence is much more limited in comparison with other media

resources. Another limitation relates to resource labelling, which is in many ways a problematic task, since, as mentioned earlier, mainstream – alternative represents a continuum rather than a distinct set of categories, which makes it difficult to draw the boundary between different kinds of resources, and especially given that some alternative resources try to mimic the behaviour of their mainstream counterparts. Manual labelling also leaves some space for annotator subjectivity, especially with regard to the identification of oppositional or corrective stances. Finally, another constraint of our analysis lies in the fact that it is difficult to disentangle different conversations that are part of the same thread, and identifying the timeframe for conversation entrainment can be performed in different ways. Our solution to account for this limitation was to test alternative time limits.

References

[1] Mathilda Åkerlund. Dog whistling far-right code words: the case of 'culture enricher' on the Swedish web. *Information, Communication & Society*, 25(12):1808–1825, 2022.

[2] Kim Andersen, Adam Shehata, and Dennis Andersson. Alternative news orientation and trust in mainstream media: A longitudinal audience perspective. *Digital Journalism*, 0(0):1–20, 2021.

[3] Pablo Aragón, Vicenç Gómez, and Andreaks Kaltenbrunner. To thread or not to thread: The impact of conversation threading on online discussion. In *Eleventh international AAAI conference on web and social media*, 2017.

[4] Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 13–22, 2013.

[5] Arunkumar Bagavathi, Pedram Bashiri, Shannon Reid, Matthew Phillips, and Siddharth Krishnan. Examining untempered social media: Analyzing cascades of polarized conversations. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '19, page 625–632, New York, NY, USA, 2020. Association for Computing Machinery.

[6] Jiajun Bao, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. Conversations gone alright: Quantifying and predicting prosocial outcomes in online conversations. In *Proceedings of the Web Conference 2021*, pages 1134–1145, 2021.

[7] Jonah Berger and Katherine L. Milkman. What makes online content viral? *Journal of Marketing Research*, 49(2):192–205, 2012.

[8] Josemar Alves Caetano, Gabriel Magno, Marcos Gonçalves, Jussara Almeida, Humberto T. Marques-Neto, and Virgílio Almeida. Characterizing attention cascades in WhatsApp groups. In *Proceedings of the 10th ACM Conference on Web Science*. ACM, jun 2019.

[9] Meng Cai, Han Luo, Xiao Meng, Ying Cui, and Wei Wang. Influence of information attributes on information dissemination in public health emergencies. *Humanities and Social Sciences Communications*, 9(1):1–22, 2022.

[10] Jun Chen, Chaokun Wang, Heran Lin, Weiping Wang, Zhipeng Cai, and Jianmin Wang. Learning the structures of online asynchronous conversations. In *Database Systems for Advanced Applications: 22nd International Conference, DASFAA 2017, Suzhou, China, March 27-30, 2017, Proceedings, Part I 22*, pages 19–34. Springer, 2017.

[11] Randall Collins. Emotional energy as the common denominator of rational action. *Rationality and Society*, 5(2):203–230, 1993.

[12] Randall Collins. *Interaction Ritual Chains*, chapter The Mutual focus / emotional-entrainment model, pages 47–101. Princeton University Press, STU - Student edition, 2004.

[13] Randall Collins. *Interaction Ritual Chains*. Princeton Studies in Cultural Sociology ; 62. Princeton University Press, Princeton, NJ, course book edition, 2014.

[14] Randall Collins. Interaction ritual chains and collective effervescence. *Collective emotions*, pages 299–311, 2014.

[15] Randall Collins. Social distancing as a critical test of the micro-sociology of solidarity. *American Journal of Cultural Sociology*, 8(3):477–497, 2020.

[16] Randall Collins and Robert Hanneman. Modelling the interaction ritual theory of solidarity. *The problem of solidarity: Theories and models*, pages 213–237, 1998.

[17] Munmun De Choudhury, Hari Sundaram, Ajita John, and Dorée Duncan Seligmann. What makes conversations interesting? Themes, participants and consequences of conversations in online social media. In *Proceedings of the 18th international conference on World wide web*, pages 331–340, 2009.

[18] Paul DiMaggio, Clark Bernier, Charles Heckscher, and David Mimno. Interaction ritual threads: Does IRC theory apply online? In *Ritual, emotion, violence*, pages 81–124. Routledge, 2018.

[19] John W Du Bois. The stance triangle. *Stancetaking in discourse: Subjectivity, evaluation, interaction*, 164(3):139–182, 2007.

[20] Rui Fan, KE Xu, and Jichang Zhao. Higher contagion and weaker ties mean anger spreads faster than joy in social media. *arXiv preprint arXiv:1608.03656*, 2016.

[21] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential than joy: Sentiment correlation in Weibo. *PloS one*, 9(10):e110184, 2014.

[22] Emilio Ferrara and Zeyao Yang. Measuring emotional contagion in social media. *PloS one*, 10(11):e0142390, 2015.

[23] Antonios Garas, David Garcia, Marcin Skowron, and Frank Schweitzer. Emotional persistence in online chatting communities. *Scientific Reports*, 2(1):1–8, 2012.

[24] David Garcia, Arvid Kappas, Dennis Küster, and Frank Schweitzer. The dynamics of emotions in online interaction. *Royal Society Open Science*, 3(8):160059, 2016.

[25] Paolo Gerbaudo, Ciro Clemente De Falco, Giulia Giorgi, Silvia Keeling, Antonia Murolo, and Federica Nunziata. Angry posts mobilize: Emotional communication and online mobilization in the facebook pages of western european right-wing populist leaders. *Social Media + Society*, 9(1):20563051231163327, 2023.

[26] Jürgen Habermas. Political communication in media society: Does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. *Communication Theory*, 16(4):411–426, 2006.

[27] André Haller and Kristoffer Holt. Paradoxical populism: How PEGIDA relates to mainstream and alternative media. *Information, Communication & Society*, 22(12):1665–1680, 2019.

[28] Dan Hiaeshutter-Rice and Brian Weeks. Understanding audience engagement with mainstream and alternative news posts on Facebook. *Digital Journalism*, 9(5):519–548, 2021.

[29] Kristoffer Holt, Tine Ustad Figenschou, and Lena Frischlich. Key dimensions of alternative news media. *Digital Journalism*, 7(7):860–869, 2019.

[30] C.J. Hutto and Eric Gilbert. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, pages 216–225, May 2015.

[31] Joshua Introne, Luca Iandoli, Julia DeCook, Irem Gokce Yildirim, and Shaima Elzeini. The collaborative construction and evolution of pseudo-knowledge in online conversations. In *Proceedings of the 8th International Conference on Social Media amp; Society*, SMSociety17, New York, NY, USA, 2017. Association for Computing Machinery.

[32] Bohan Jiang, Mansooreh Karami, Lu Cheng, Tyler Black, and Huan Liu. Mechanisms and attributes of echo chambers in social media, 2021.

[33] Lisa Kaati, Katie Cohen, Björn Pelzer, Nazar Akrami, Eric Andersson, and Felix Knutas. En studie i fördom: Om rasistiska stereotyper i digitala miljöer, 2022. Totalförsvarets forskningsinstitut.

[34] Theodore D Kemper. *Status, power and ritual interaction: A relational reading of Durkheim, Goffman and Collins*. Routledge, 2016.

[35] Silvia Knobloch-Westerwick, Cornelia Mothes, and Nick Polavin. Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, 47(1):104–124, 2020.

[36] Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[37] Dilek Küçük and Fazli Can. Stance detection: A survey. *ACM Comput. Surv.*, 53(1), feb 2020.

[38] Anders Olof Larsson. Right-wingers on the rise online: Insights from the 2018 Swedish elections. *New Media & Society*, 22(12):2108–2127, 2020.

[39] Hugo Leal. Networked disinformation and the lifecycle of online conspiracy theories. In *Routledge handbook of conspiracy theories*, pages 497–511. Routledge, 2020.

[40] Dennis K. K. Leung and Francis L. F. Lee. Cultivating an active online counterpublic: Examining usage and political impact of Internet alternative media. *The International Journal of Press/Politics*, 19(3):340–359, 2014.

[41] Juan Pablo Luna, Sergio Toro, and Sebastián Valenzuela. Amplifying counter-public spheres on social media: News sharing of alternative versus traditional media after the 2019 Chilean uprising. *Social Media + Society*, 8(1):20563051221077308, 2022.

[42] M. Magnani, D. Montesi, and L. Rossi. Conversation retrieval for microblogging sites. *Information Retrieval*, 15(3-4), 2012.

[43] Matteo Magnani, Danilo Montesi, and Luca Rossi. Friendfeed breaking news: Death of a public figure. In *Proceedings - SocialCom 2010: 2nd IEEE International Conference on Social Computing, PASSAT 2010: 2nd IEEE International Conference on Privacy, Security, Risk and Trust*, 2010.

[44] Alexandre Maros, Jussara M. Almeida, and Marisa Vasconcelos. A study of misinformation in audio messages shared in WhatsApp groups. In Jonathan Bright, Anastasia Giachanou, Viktoria Spaiser, Francesca Spezzano, Anna George, and Alexandra Pavliuc, editors, *Disinformation in Open Online Media*, pages 85–100, Cham, 2021. Springer International Publishing.

[45] Clara Juarez Miro and Benjamin Toff. How right-wing populists engage with cross-cutting news on online message boards: The case of forocoches and vox in spain. *The International Journal of Press/Politics*, 0(0):19401612211072696, 0.

[46] Ida Momennejad, Ajua Duker, and Alin Coman. Bridge ties bind collective memories. *Nature communications*, 10(1):1–8, 2019.

[47] Özen Odag, Anne Leiser, and Klaus Boehnke. Reviewing the role of the Internet in radicalization processes. *Journal for deradicalization*, (21):261–300, 2019.

[48] Karin Östman and Richard Aschberg. Flashback – ett laglöst land. *Financial Times*, 9 Feb 2015. Available at: https://www.aftonbladet.se/nyheter/a/ddG3Rq/flashback--ett-laglost-land (Accessed: April 27th, 2023).

[49] Jeannette Paschen. Investigating the emotional appeal of fake news using artificial intelligence and human contributions. *Journal of Product & Brand Management*, 2019.

[50] Thea Riebe, Katja Pätsch, Marc-André Kaufhold, and Christian Reuter. From conspiracies to insults: A case study of radicalisation in social media discourse. *Mensch und Computer 2018-Workshopband*, 2018.

[51] Jörg Rössel and Randall Collins. Conflict theory and interaction rituals: The microfoundations of conflict theory. *Handbook of sociological theory*, pages 509–531, 2001.

[52] Linn A.C. Sandberg and Karoline Andrea Ihlebæk. Start sharing the news. *Statsvetenskaplig tidskrift*, 121(3), 2019.

[53] Heidi Schulze. Who uses right-wing alternative online media? An exploration of audience characteristics. *Politics and Governance*, 8(3):6–18, 2020.

[54] Kate Starbird, Ahmer Arif, Tom Wilson, Katherine Van Koevering, Katya Yefimova, and Daniel Scarnecchia. Ecosystem or echo-system? Exploring content sharing across alternative media domains. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.

[55] Nora Theorin and Jesper Strömbäck. Some media matter more than others: Investigating media effects on attitudes toward and perceptions of immigration in Sweden. *International Migration Review*, 54(4):1238–1264, 2020.

[56] Anton Törnberg and Anita Nissen. Mobilizing against islam on social media: Hyperlink networking among European far-right extra-parliamentary Facebook groups. *Information, Communication & Society*, pages 1–19, 2022.

[57] Anton Törnberg and Petter Törnberg. Combining CDA and topic modeling: Analyzing discursive connections between Islamophobia and anti-feminism on an online forum. *Discourse & Society*, 27(4):401–422, 2016.

[58] Salla Tuomola and Karin Wahl-Jorgensen. Emotion mobilisation through the imagery of people in finnish-language right-wing alternative media. *Digital Journalism*, 11(1):61–79, 2023.

[59] Petter Törnberg. Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLOS ONE*, 13(9):1–21, 09 2018.

[60] Teun A Van Dijk. Stories and racism. *Narrative and Social Sontrol: Critical Perspectives*, 21:121–142, 1993.

[61] Sander Van Haperen, Justus Uitermark, and Alex Van der Zeeuw. Mediated interaction rituals: A geography of everyday life and contention in Black Lives Matter. *Mobilization: An International Quarterly*, 25(3):295–313, 2020.

[62] Davide Vega and Matteo Magnani. Foundations of Temporal Text Networks. *Applied Network Science*, 3(1):25, 2018. _eprint: 1803.02592.

[63] Davide Vega and Matteo Magnani. Metrics for Temporal Text Networks. pages 147–160. Springer, Cham, 2019.

[64] Christian Von Scheve and Mikko Salmella. *Collective emotions: Perspectives from psychology, philosophy, and sociology*. OUP Oxford, 2014.

[65] Mattias Wahlström and Anton Törnberg. Social media mechanisms for right-wing political violence in the 21st century: Discursive opportunities, group dynamics, and co-ordination. *Terrorism and Political Violence*, 33(4):766–787, 2021.

[66] Chunyan Wang, Mao Ye, and Bernardo A Huberman. From user comments to on-line conversations. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 244–252, 2012.

[67] Yuping Wang, Savvas Zannettou, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, and Gianluca Stringhini. A multi-platform analysis of political news discussion and sharing on web communities. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1481–1492, 2021.

[68] Jie Wei and Ling Zhang. Analysis of information dissemination based on emotional and the evolution life cycle of public opinion. In *2019 International Conference on Robots Intelligent System (ICRIS)*, pages 265–268, 2019.

[69] Hadley Wickham. *httr: Tools for Working with URLs and HTTP*, 2023. https://httr.r-lib.org/, https://github.com/r-lib/httr.

[70] Victoria Yantseva and Kostiantyn Kucher. Stance classification of social media texts for under-resourced scenarios in social sciences. *Data*, 7(11):159, 2022.

[71] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtelris, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. The web centipede: understanding how web communities influence each other through the lens of mainstream and alternative news sources. In *Proceedings of the 2017 internet measurement conference*, pages 405–417, 2017.

A Related work

A.1 Alternative media content on social media platforms

The existing research demonstrated that platforms play quite a significant role in news sharing and distribution of alt-right content [53]. In the Swedish context, alternative media readership was found to stand out from a sample of other Northern and Central European countries [53], which makes it hard to underestimate the role of this information source and the potential effects it may have on the readers' attitudes and political orientations. On top of this, alternative media readership was associated with distrust in the mainstream media, however, it also needs to be mentioned that it was found to supplement rather than completely replace traditional news outlets [2]. On social media and, in particular, on Twitter, alternative media presence was described in terms of "echo-system" where the same alternative media content was disseminated by an ideologically diverse set of actors [54]. On Facebook, links to Swedish alternative media were found to constitute almost one-third of all URLs shared during the 2018 election period, with link engagement levels almost on par with that of traditional media [52]. Even more so, the existing research demonstrated that Swedish right-wing actors were more successful in engaging the public during the 2018 election in comparison with their established counterparts, in particular, by means of resorting to offensive language and negative emotionality [38]. A similar study in the US setting also found that ideologically extreme news sources enjoyed the highest engagement levels on Facebook, despite that well-known mainstream publishers had larger audiences, since the latter did not necessarily imply high engagement with their news content [28].

When it comes to the use of alternative versus mainstream media in right-wing online environments, the existing research pointed out that, quite surprisingly, the use of and references to mainstream media sources are just as popular as those to alternative ones [27, 56], suggesting that right-wing audiences consume different types of content despite the dominating narrative of mistrust into the established media. Thus, the research suggests, despite the circulation of illiberal or radical opinions inside fringe and echo chamber-like environments, such platforms still allow for the dissemination of cross-cutting content and can nevertheless serve as deliberative spaces [45].

With regard to the dynamics and patterns of alt-right content sharing, a multi-platform comparison of alternative versus mainstream media news spread demonstrated that some of the fringe communities across the platforms generate a disproportionately high volume of alternative links sharing events that also seem to influence the emergence of the same content on other platforms [71]. Another study by Wang et al. that considered trustworthy and untrustworthy news sources instead of mainstream versus alternative ones arrived at a similar result and suggested that small communities can become extremely effective in pushing news stories to other communities [67]. Another relevant contribution to the existing evidence was made by Luna et al. who highlighted the differences in the dissemination of mainstream versus alternative media on Facebook: while alternative news content sharing was steadily increasing with time after the publication, mainstream news content sharing, in contrast, followed a burst and decay pattern [41]. Despite the existing evidence, little is known about the impact of alternative media content on the dynamics of discus-

sions in fringe communities and whether such content reinforces or even produces more negative attitudes and narratives about the discussion topics. Some of the existing exceptions is a study by Introne et al. who have provided a description of how pseudo-knowledge comes into being as a result of collaborative effort in the forum discussions [31].

A.2 Collective emotions and emotional contagion

Quite logically, negative and emotional content was shown to catch more users' attention and disseminate more intensively [9, 68]. Likewise, the content diffused by alternative media was shown to be more negative and emotionally charged in comparison with the mainstream media content [44] and was shown to be more likely to be expressing such emotions as anger and disgust [49, 58]. On Facebook, immigration- and security-related posts by right-wing actors were found to be especially likely to cause "anger" reactions and further sharing of those posts [25]. However, Berger and Milkman [7] demonstrated also that the relationship between emotion and virality is more complex than that: in particular, some emotions, such as sadness, might actually have a reverse, deactivating, influence on content virality. Rather, the emotions that evoke high arousal (no matter if they are positive or negative) are the main driving force for the diffusion of any type of content in online networks.

Some of the existing studies, mostly based on the results of experiments and agent-based modelling, also suggested that emotional content leads to higher arousal [24], while the valence of posts published on Facebook was found to depend on the valence of posts previously seen by a given user [36]. Another observational study of Twitter users and their followees' posts arrived at a similar conclusion, thereby giving evidence to the presence of emotional contagion [22]. Some works suggested also that negative emotions, such as anger, expressed in the initial social media content, are more contagious than positive ones, such as joy [21, 20]. Yet another study demonstrated that, although the emotional tone of users' messages is usually adapted to that of the chat, chat's emotion tone can be characterized as persistent, rather than fluctuating towards particular negative or positive emotions [23]. What distinguishes our approach from the earlier effort is that we focus not only on the process of emotional contagion online in general (e.g. through news feeds and following/befriending other users) but rather on the group dynamics emerging as a result of users' participation in online *conversations* that may lead to different dynamics and consequences of exposure to different types of emotions in news content.

A.3 Conversation dynamics online

With regard to the structure and dynamics of the conversations, the existing research focused on several aspects such as the roles of individuals [46, 39], factors that make conversations interesting [17], conversations' length and user participation prediction [4], or even the outcome of conversations in terms of demonstrated prosocial behaviours [6]. From a more technical side, taking into account the structure of conversations was also found to improve model accuracy in the dispute escalation prediction tasks. On top of this, many of the existing works are based on the

assumption that threads themselves represent distinct conversations, for instance, conversations' thread-based structure was found to enable higher reciprocity between users [3].

Further, work by Bagavathi et al. attempted to study conversations on Gab as an example of cascades and identified several types of cascades typical on the platform [5]. A similar approach to model conversation dynamics as a cascading process has been proposed by [66]. Yet another approach to conversation reconstruction also took into account the linguistic features of the messages to determine the reply-to relationship between them [10]. The study most relevant to our empirical case, even though it does not explicitly focus on conversations or alternative news content, is the study by Caetano et al. [8] who explored structural, temporal and user engagement properties of "attention cascades" with falsehoods in non-political versus political WhatsApp groups. In particular, they found that the latter generate deeper and wider cascades [8].

Finally, it is also worth mentioning the effort to adapt Collins' micro-sociological approach to account for mediated forms of interaction and online communication, even though it was initially created with only offline interactions in mind. For instance, DiMaggio et al. applied Collins' IRC theory to explore users' behaviours online. This study used also used a thread-based approach to user's online interactions and explored the predictors of thread length, with latter serving as a way to evaluate the interactions' success [18]. Another interesting example is the work of van Harpenen et al. who applied methods of image recognition to develop a typology of mediated interaction rituals using a dataset of Instagram pictures related to the Black Lives Matter movement [61]. They proposed that such visual content served as a tool to connect with the movement despite geographical and physical separation, which also speaks in favour of the theory's relevance for mediated types of communication.

B Theory: online conversations and collective emotions

This work uses Randall Collins' micro-sociological approach and, in particular, International Ritual Theory (IRC), a framework to study collective emotions that come into life as a result of social interactions [64]. In accordance with the IRC approach, conversations can be represented as one type of social interactions being able to produce emotional synergy and group solidarity [34, p.197] in cases when the necessary conditions are fulfilled, in particular, group assembly and barriers to outsiders, as well as common mood and focus of attention (for more details on the key principles of the IRC theory, see e.g. [13]). Collins' take on conversations is in many ways coherent with our understanding of conversations as an evolving chain process dependent on the previous interactions as part of the same chain, which can be perfectly summarized by the following excerpt: "At any particular moment, people are speaking certain words or thinking certain thoughts; the thoughts that go through one's head are internalized from previous talk with other people; more innovative thoughts are assembled out of the ingredients of verbal ideas already internalized. The world is a network of conversations, and what people think at any point in it is a product of what has circulated in previous conversations. There is a crucial emotional component: ideas are better remembered, and make more sense, if they were associated with emotion when they were previously talked about" [64, pp. 303–304] .

Contrary to the Habermasian statement about the deliberative and rational character of information exchange [26, p. 415], one can suggest that such conversations do not necessarily need to follow any rational or instrumental value – in other words, these discussions can be described as "...emotional, symbolic or value-oriented behaviour" [11, p.205], which distinguishes rituals, including the mediated ones, from purely instrumental types of actions. In particular, users' main motive for the participation in conversations can lie in the production of particular emotions and "emotional solidarity with the group" [11, p.215], rather than in reaching a consensus on a controversial topic or exchanging rational arguments for or against the discussed issue.

Following the same line of reasoning, an act of sharing content with other conversation participants can be described as both a symbolic and instrumental action. On the one hand, content sharing can be seen as a practice to initiate higher user involvement in conversation and thus higher mobilization and solidarity, and, possibly, invoke specific types of emotional reactions in those who consume such content. It can also be used to justify particular attitudes or support particular opinions expressed by the users. Indeed, it has been shown that covert racist or biased attitudes are often justified with the help of argumentative and rational reasoning [60, p. 35]. On the other hand, content sharing can also serve as a specific collective and ritual symbol [11, p.212], just as, for example, as specific jargon and language to talk about the immigrants [1]. Such symbols may be geared towards achieving higher group solidarity and the sense of group belonging enabled by the platform architecture.

When it comes to the right-wing, populist or reactive movements specifically, Collins suggests that those are more prone to emotional mobilisation: "...reactive movements... are easier to mobilize, and generally more emotionally aroused that positive movements seeking a better future..." [14, p.305]. This, Kemper adds, is even more relevant in cases when there exists an out-group (in this case, those labelled as "immigrants"), so that the focus of common attention, which is a necessary trait of any ritual, can be directed at those considered enemies [34, p.177].

The question of whether Collins' sociology of emotions is applicable to online contexts has lately been subject to debate. Collins himself has dismissed the idea that online environments can general emotional solidarity due to the participants' timely and spatial misalignment [15]. Yet, some of the recent studies have successfully demonstrated the theory's applicability to online contexts and found evidence for the emergence of collective emotions and group solidarity. A particularly interesting example is a study by DiMaggio et al. who analysed forum discussion threads and provided evidence that some of the theory's propositions hold even for online contexts and may be illustrative of some traits of human communication in general [18].

However, in contrast to offline and in-person conversations, mediated conversations such as those on Flashback differ in a few ways. The most obvious one is, of course, temporal and spatial misalignment. The conversation participants are not only physically disconnected from each other, but they also participate in conversations asynchronously, with silent intervals from several seconds to several days. Despite such a misalignment, the ordering of interactions and reply-to functionality allow users to infer the previous content of conversations without losing too much context. On top of this, despite temporal misalignment, one can nevertheless suggest that online conversations have their own pace, which is an important component in generating emotional

synergy in Collins' approach [16]. Indeed, as it follows from the Results section, the vast majority of utterance exchanges in our dataset happen within a quite narrow time span of minutes. If the pause between messages is too large, then the conversation pace gets violated - that is why our approach, in contrast to the earlier works that equate threads and conversations, disentangles various conversations that take place as part of the same thread. For instance, if the thread topic catches users' attention, they might not only participate in the conversation by posting an utterance, but also stay online and follow up on how the conversation unfolds, which creates entrainment mentioned by Collins, and, potentially, generates emotional energy.

Secondly, the format of online conversations is underpinned by the platform affordances or, in other words, the logic and functionality of the platform. If offline rituals can take numerous forms, online interaction rituals, in their basic form, presume the written form and permanency of the messages (if those, of course, are not based on disappearing or voice messages). Another aspect related to the platform affordances is their functionality - for instance, Facebook allows liking others' messages, while Flashback does not - in other words, this constraints users' behaviours or, rather, allows users to act only in specific ways, which shapes the ways online conversations may develop. One such affordance that is quite common to almost all platforms is exactly content sharing – or the possibility for the users to circulate and access additional digital content, such as external websites, photos and videos, which can provide additional context or support the speaker's utterance. Thus, exploring different platform affordances and their effect on online interactions can serve as another direction for further research and theory building.

It also needs to be mentioned that online conversations are also different from other types of online communication, such as for instance, retweeting others' messages or commenting on other users' posts or content. In the case of retweets, the retweeting user does not add any substantive content on top of the content already provided by the retweeted message. In the case of user comments, then, the comment author does not need to take the previous comments into account - in other words, s/he can post the comment irrespective of other users' expressions. On the contrary, some other forms of online communication, such as, for example, instant messaging services such as WhatsApp or e-mail chains, can also be described as online conversations, since they presume that the replying user needs to be aware of the previous content for the conversation to make sense. On the other hand, not every online conversation can be described as an online ritual, since it might not necessarily meet the requirements of common mood and sense of group belonging [14].

Finally, one can suggest that what Collins himself calls as the production of "emotional energy" as a result of a successful interaction ritual can be operationalized in at least two different ways using the idea of *entrainment*, or users' engagement with and involvement in the ritual [12]. The conversation entrainment can be expressed, for instance, in terms of conversation length - successful conversation will, of course, be longer, or in terms of its pace, or rhythmic coordination [16] - the shorter and more consistent the interval between the posts, the more engaged with the conversation the users are. The second way is through measuring emotional mobilisation, or the evolution and intensification of shared emotions generated in the conversation [12].