Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

Pathways through Conspiracy: The Evolution of Conspiracy Radicalization through Engagement in Online Conspiracy Discussions

Shruti Phadke[1], Mattia Samory[2], Tanushree Mitra[1]
[1] University of Washington, USA
[2] GESIS, Germany
phadke@uw.edu, mattia.samory@gesis.org , tmitra@uw.edu

Abstract

The disruptive offline mobilization of participants in online conspiracy theory (CT) discussions has highlighted the importance of understanding how online users may form radicalized conspiracy beliefs. While prior work researched the factors leading up to joining online CT discussions and provided theories of how conspiracy beliefs form, we have little understanding of how conspiracy radicalization evolves after users join CT discussion communities. In this paper, we provide the empirical modeling of various radicalization phases in online CT discussion participants. To unpack how conspiracy engagement is related to radicalization, we first characterize the users' journey through CT discussions via conspiracy engagement pathways. Specifically, by studying 36K Reddit users through their 169M contributions, we uncover four distinct pathways of conspiracy engagement: steady high, increasing, decreasing and steady low. We further model three successive stages of radicalization guided by prior theoretical works. Specific sub-populations of users, namely those on steady high and increasing conspiracy engagement pathways, progress successively through various radicalization stages. In contrast, users on the decreasing engagement pathway show distinct behavior: they limit their CT discussions to specialized topics, participate in diverse discussion groups, and show reduced conformity with conspiracy subreddits. By examining users who disengage from online CT discussions, this paper provides promising insights about conspiracy recovery process.

Introduction

Disinformative, panic-inducing online conspiracy theories (CT) are increasingly proving to be threats to productive civic discourse. Violent riots at the U.S. Capitol (Garrett 2021) and COVID-19 vaccine skepticism (Blake 2021) are just two of the many examples of how online CT discussions result in socially harmful situations. Further, by mainstreaming the fringe, social media allow previously disconnected conspiracy theorists to find like-minded individuals, reinforce their beliefs, and mobilize (Marwick and Lewis 2017). Despite the clear implications of online CT engagement, we know little about an individual's journey through CT discussion communities and how they become increasingly engaged (or not) with conspiratorial worldviews. This paper provides just such an understanding.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

What are the pathways of online conspiracy engagement? Using digital traces of 36K Reddit users who participated in r/conspiracy—the biggest CT discussion subreddit, we characterize their Reddit trajectories after they made their first comment in r/conspiracy. Working with their 169M contributions spread across 4K subreddits, we find four distinct types of engagement trends in CT discussion subreddits—steady high, increasing, decreasing and steady low. For instance, users with increasing engagement contribute consistently and predominantly more over time in CT subreddits compared to other subreddits.

How does conspiracy radicalization process evolve for users? To answer, we leverage a theoretical model of internet mediated radicalization (Neo 2016) comprising five phases—Reflection, Exploration, Connection, Resolution, Operation (RECRO). We focus on the first three phases that are most visible in online discussions (Van Raemdonck 2019). They describe the psychological and emotional vulnerabilities (Reflection), development of alternative worldviews (Exploration) and social bonds (Connection) formed during the radicalization process. We model the phases through various linguistic, interaction and activity features. For example, we characterize the Exploration phase by creating a generality scale that models the users' monological conspiracy worldview (Goertzel 1994) signaling adoption of generalized conspiracy thinking.

We find that users with steady high and increasing engagement in CT do show signs of radicalization through increasing use of insider language, and repeated participation in small-group discussions. Conversely, users with decreasing engagement in CT communicate in diverse discussion groups, and never develop lexical conformity with conspiracy communities. Moreover, users with steady high and increasing engagement, increasingly engage in generalist CT discussion subreddits, showing support for the monologicality hypothesis of conspiracy belief evolution—in stark contrast to users on decreasing pathways who limit their contributions to specific CT discussion topics.

Contributions. Through a theory-driven, empirical study of the conspiracy radicalization process, our work lays the

foundation for observing users who may act on their radicalized beliefs, i.e., those who may proceed to the Operation phase. We also offer a generality scale used in characterizing conspiracy worldviews, that captures the generalist or specialist nature of subreddits. Overall, our work characterizes how conspiracy beliefs evolve during an individual's online lifespan since their first contribution to CT discussions. By differentiating between users with high engagement from those who disengage with CT discussions our work has implications for understanding factors in recovery from online CT discussions.

Related Work

Online Conspiracy Engagement

People turn to conspiracy beliefs to fulfill their epistemic, existential, and social needs (Douglas, Sutton, and Cichocka 2017). For example, users who join CT discussion communities during crisis events—characterized by existential threats and uncertainty—engage more and more exclusively with them (Samory and Mitra 2018). Further, social activity, such as facing marginalization from mainstream discussion spaces and bonding with incumbent members, correlates with a higher likelihood of joining CT communities (Phadke, Samory, and Mitra 2021b). Scholars investigating the pathways into conspiracy theory engagement found the evidence of self-selection and shared interests that feed into engagement in CT discussions (Klein, Clutton, and Dunn 2019). Conversely, extreme cognitive dissonance due to conflicting beliefs, correlates with lowered engagement and shorter tenure (Phadke, Samory, and Mitra 2021a). Similarly, encouragement by trusted peers and exposure to evidence-based counter-narratives prompt people to exit their conspiracy beliefs (Xiao, Cheshire, and Bruckman 2021). Adding to this recent body of literature, which uncovers how individuals begin and end their engagement with CT communities, this work connects the unexplored transition between these extremes. Specifically, we investigate how different conspiracy engagement pathways are associated with various phases of internet-mediated radicalization.

Conspiracy Theorizing and Radicalization

By radicalization, we mean the growing support by individuals or groups for a radical societal change, either through belief or action, that harms the social fabric and functioning (Dalgaard-Nielsen 2010). Researchers have identified three broad ways in which conspiracy theorizing can play a role in advancing radicalization. First, conspiracy theoriz-ing can provide a paranoid interpretation of reality in which there are clear in-groups and out-groups that overall enhance the appeal of extremist narratives (Vermeule and Sunstein 2009). Second, conspiracy ideation and radicalization are believed to contain similar underlying psychological disposition such as feelings of anger, anxiety, paranoia (Butter and Knight 2020). Third, apart from creating clear boundaries between in-group and out-group, conspiracy theorizing can strengthen in-group bonding which is essential for radicalization process (Conway 2012).

Although this literature shows that conspiracy theorizing may lead to radicalization, there is little to no empirical investigation connecting the two phenomena (Butter and Knight 2020). Especially in light of the recent events where online CT discussions led to riots causing national security issues (Garrett 2021), we believe that it is important to study if, and how, online CT discussion participants display markers of radicalization. In this paper, we study the process of conspiracy radicalization using a RECRO model for internet-mediated radicalization (Neo 2016) which we briefly explain next.

Models of Radicalization

Radicalization models describe the progression of the radicalization process where individuals move from socially normative perspectives towards more extremes ones (Neo 2020). However, extant radicalization pathway models and theories either do not consider the role of internet in radicalization or focus mainly on the psychological predispositions of people (Neo 2016). The RECRO model proposed by Neo, however, is a pathway-based theoretical model that views radicalization as an internet-mediated process involving all, individual, epistemic and social factors. Researchers have used RECRO in qualitative analyses of online anti-vaccination discussions finding that social media provides a strong platform for the first three phases (Van Raemdonck 2019)—Reflection, Exploration, Connection. Hence, we study the first three phases of radicalization in the CT engagement which are briefly explained below.

Reflection: The Reflection phase details the vulnerabilities and psychological predispositions that increase one's receptivity towards radicalization (Neo 2016). This is a phase where personality and psychological factors motivate the individuals to open-up, also described as "cognitive opening", to alternate belief systems. Other researchers also agree on the importance of psychological footprints such as anger and heightened emotions in online radicalization (Dalgaard-Nielsen 2010). After the cognitive opening, users begin to form radical worldviews in the Exploration phase.

Exploration: Here, individuals begin to make sense of new information and narratives by forming alternate worldviews in a way that fosters eventual radicalization (Neo 2016). Individuals are primed to form a new, alternate worldview that resonates with their interests and epistemological needs (Neo 2016). Specifically in relation to conspiracy theorizing, researchers propose a "monological belief system," describing it as a stable cognitive style that dictates the perceived functioning of the social world (Goertzel 1994). Monological conspiracy worldviews offer a general set of assumptions, such as cover-ups by powerful people, that are portable across multiple CTs and socio-political phenomena, independently from their specific topic or context (Goertzel 1994). This affords applying CT to any socio-political phenomena, independently from the specific topic or context of an event (Franks et al. 2017a). However, Hence, the monologicality hypothesis paints the picture of a closed-minded CT believer with a strong mobilization potential,

and of a CT ecosystem of broadly applicable, interconnected, mutually supporting ideas. This hypothesis though is contested. Competing research presents a possibility of better educated, open, and socially active CT believers who might restrict their interests to specific conspiracy topics (Franks et al. 2017b). This paper, for the first time, analyzes the generality or specificity of CT belief by modeling how individuals explore the world of online conspiracies after their initial exposure.

Connection: Here, individuals interact to form group bonds with like-minded people (Neo 2016). As opposed to the Reflection phase capturing individual predisposition, the Connection phase describes how bonds with a group of peers advance the radicalization process. Specifically, co-hesion or conformity to one's social group (Crossett and Spitaletta 2010), small-group dynamics (Reedy, Gastil, and Gabbay 2013) and feelings of group affiliation (Dalgaard-Nielsen 2010) are strongly associated with radicalization.

Data

Selecting Reddit users: In this paper, we study the evolution of conspiracy radicalization in social media users after their initial engagement in CT discussions on Reddit. To mark the users' entry into Reddit's CT discussion world, we look at users' first comment into r/conspiracy — Reddit's biggest and most popular conspiracy discussion community. Our goal is to analyze users' long-term engagement in online CT discussions. Hence, we select 42,225 users that contribute at least 20 comments in r/conspiracy over at least one year. Having such activity thresholds ensures that our analysis is not biased by users that have sparse involvement in CT discussions, a practice commonly followed in the social computing research (Kumar et al. 2018; Samory and Mitra 2018). Additionally, we filter the user list based on other robustness checks detailed in Section "Additional Robustness Checks", to avoid selection bias and ensure fair activity coverage across all users. Finally, the filtered dataset includes 36,314 users.

Extracting user timelines: Reddit users differ in their frequency and levels of contribution. To compare evolution of all users on equal footing, we split their Reddit activity in ten equal deciles of contribution volume. For example, if a user x makes their first comment in r/conspiracy on 1st January 2018 and makes n comments on Reddit since then, we split their entire Reddit activity after 1/1/2018 into equal, time ordered batches of n/10 comments (Fig. 1 (b)). This approach ensures that (1) we have ten deciles worth of activity with an equal number of contributions in each decile for every user and (2) that each user is evaluated according to their own pace of engagement in CT discussions. A similar approach for mitigating temporal shifts in user activity has been validated in studying far-right radicalization on Twitter (Vidgen, Yasseri, and Margetts 2021).

Thus, we collect the entire activity for 36,314 users after their first post in r/conspiracy and prior to December 2020, split over 10 deciles of equal contribution volumes. In order to eliminate subreddits that see contributions from users only occasionally, we kept subreddits that had at least 10 contributions from at least 5 different authors, spreading the user activity over 4,756 subreddits and 169M comments. We mine all data using Pushshift (Baumgartner et al. 2020).

Pathways of Conspiracy Engagement

To understand a user's journey through Reddit after participating in conspiracy discussions, we first model each user's longitudinal development of CT engagement. Fig. 1 (c) outlines our process for extracting pathways of conspiracy engagement. We start by describing the conspiracy similarity scale and our methods for validating the scale.

Creating the Conspiracy Similarity Scale

To understand how conspiratorial are the discussions in each subreddit, we use a scalar conspiracy similarity scale inspired by the methods described by Samory and Mi-tra (2018). Specifically, we map subreddits on a scale from -1 to 1 where 1 represents the highest similarity to r/conspiracy. Previous works find polarized communities in CT and scientific news consumption patterns (Bessi et al. 2015). Moreover, in terms of psychology and norms, there are known biases associated with conspiracy beliefs that would be unacceptable in scientific communities (Kuhn et al. 2021). Hence, to find the CT discussion communities, we contrast the user activity in r/conspiracy with its polar opposite and the largest scientific community — r/science. Using the Reddit activity of users with at least 10 comments in r/conspiracy or r/science, we create a vector representation for each subreddit by calculating pointwise mutual information (PMI) between each pair of subreddits. More simply put, PMI is a co-occurrence based measure (Bouma 2009) that can characterize similarity between two subreddits based on the number of commonly occurring users in them. PMI provides a high dimensional matrix where the pointwise mutual information is provided for each pair of subreddits in the dataset. Hence, to create a lower dimensional embedding for each subreddit, we calculate the singular value decomposition (SVD) matrix on PMI. Finally, to characterize a subreddit's similarity to CT discussions, we calculate the cosine similarity of each SVD vector with the vector for r/conspiracy. The final conspiracy similarity scale has scores for 4,756 subreddits with top most subreddits similar to r/conspiracy listed in Table 1.

Validating the Conspiracy Similarity Scale: To validate the conspiracy similarity scale we refer to the concept of convergent validity that measures the correlation between the conspiracy similarity scale with other subreddit similarity measures based on the same construct. For this comparison we use the only publicly available subreddit embeddings contributed by Kumar et al. (2018). While those subred-dit embeddings are not specifically catered towards finding CT subreddits, the embeddings do provide a general mea-sure of subreddit similarity. Specifically, we calculate Spear-man's rank-order correlation between the 1000 subreddits most similar to r/conspiracy according to our and Ku-mar's rankings. We find a significant ($p < 0.05$) moderate

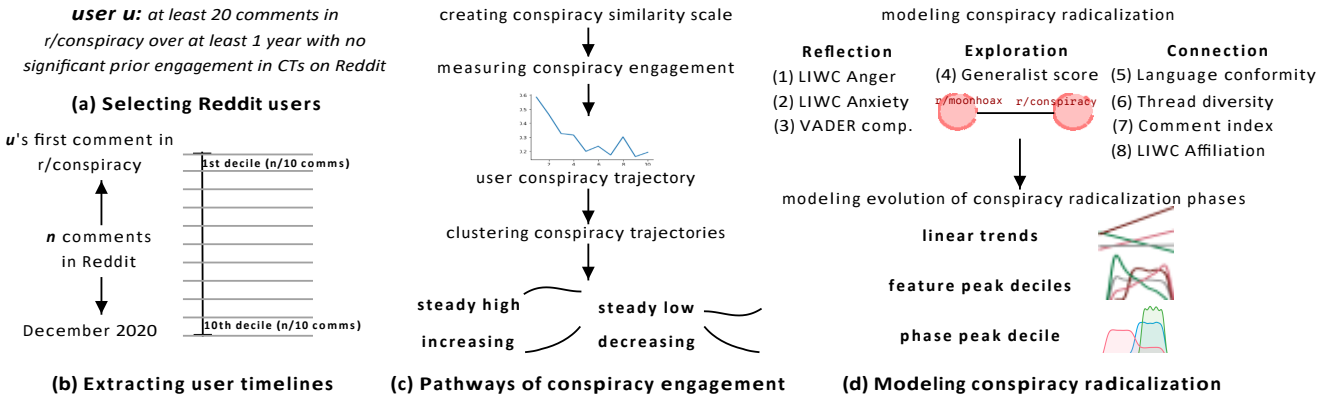

Figure 1: (a) &(b) describe user selection and data preparation processes. We split the user activity into 10 deciles of equal contribution volumes after the user's first comment in r/conspiracy as described in "Data" Section. (c) We then model the conspiracy engagement pathways by unsupervised clustering of user trajectories "Pathways of Conspiracy Engagement" Section. (d) Finally, we model the conspiracy radicalization process by operationalizing the first three phases in the RECRO model (Neo 2016) through 8 features in "Characterizing Conspiracy Radicalization" Section.

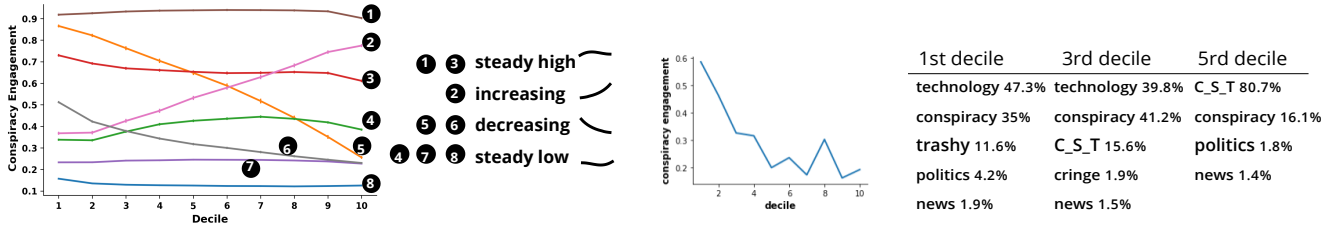

Figure 2: (a) Plot displaying average trajectories across 8 conspiracy engagement clusters. Here we can observe that some trajectories show similar trend, only shifted in amplitude. For example, both 5 & 6 show decreasing engagement. Hence we group together amplitude shifted trajectories as displayed in (b). These are the final conspiracy engagement pathways analyzed throughout this paper. (c) Example user trajectory plots shown to evaluators for validation task 1. (d) Example user-decile activity shown to the evaluators for validation task 2.

conspiracy	ConspiracyMemes
The_Donald	conspiracyundone
C_S_T	ConspiracyII
UFOs	occult
greatawakening	AskThe_Donald

Table 1: Top 10 subreddits on conspiracy similarity scale.

correlation (0.52) between the two. This corroborates that our conspiracy similarity scale successfully measures similarity to r/conspiracy. We further manually analyze the top 100 subreddits on the conspiracy similarity scale and confirm that they do host CT discussions.

Measuring Conspiracy Engagement

Conspiracy engagement measure: The conspiracy engagement measure captures what proportion of the user's Reddit activity is dedicated to CT discussion subreddits in each decile. We calculate the weighted average of contributions in each subreddit weighted by that subreddit's score on the conspiracy similarity scale. More formally, for a user u, with N contributions in a decile i, their conspiracy engagement score C_u^i will be calculated as: $C_u^i = \sum_{j=1}^{J} n_j s_j / N$, where n_j is number of contributions in the j_{th} subreddit and s_j is the subreddit's conspiracy similarity score. C_u^i is bounded between 0 to 1 with higher scores indicating higher proportion of engagement in the CT discussion subreddits.

Clustering Conspiracy Trajectories

We represent every user's conspiracy engagement trajectory as a time series of C_u^i over ten deciles. We perform unsupervised clustering of the user trajectories to find common temporal patterns, or pathways, of conspiracy engagement—a method commonly used to characterize longitudinal behavior (Genolini et al. 2016). We next discuss our choice of clustering algorithm, distance measure, and number of clusters.

In a large-scale benchmark spanning over 128 synthetic, natural and pre-processed datasets, Javed, Lee, and Rizzo (2020) found that there is no particular algorithm that outperforms others in time series clustering. However, the dy-

namic time warping (DTW) distance measure outperformed alternatives for a nontrivial number of datasets. DTW is especially suited for time series clustering for its robustness to time shifts between different trajectories (Javed, Lee, and Rizzo 2020). We thus combine K-Means clustering with a DTW distance measure. We make this choice to complement the computational complexity of DTW with fast, adaptable and convergent cluster assignments produced by K-Means. We tune the number of clusters by training K-Means models for $k \in \{2, \ldots, 15\}$, and by identifying the elbow point in the silhouette coefficients. We select $k = 8$ as a trade-off between number of clusters and cluster quality.

Results: Conspiracy Engagement Pathways

Fig. 2 (a) shows the results from our cluster analysis. We observe distinctive increasing and decreasing conspiracy engagement patterns across the 8 clusters. While DTW is not sensitive to time (x-axis) shifts in the trajectories, it is still sensitive to magnitude (y-axis) shifts. For example, clusters 5 and 6 in Fig. 2 (a) both show decreasing trend but are shifted in magnitude. Hence, we further group the magnitude-shifted clusters based on common patterns. Finally, we end up with four distinct patterns of engagement representing groups of trajectories—steady high, increasing, decreasing and steady low—as described in Fig. 2 (b).

Validating conspiracy engagement pathways: We invited 6 evaluators proficient in statistics and data analysis to manually assess the quality of conspiracy engagement pathway assignments in two annotation tasks.

1. User trajectory labeling: We asked the evaluators to label a user conspiracy trajectory plot—for example, the trajectory displayed in Fig. 2 (c)—as either steady high, increasing, decreasing or steady low. As instructions, we additionally provided C_u thresholds for each pathway and demonstrated sample trajectories from each pathway as a guideline for annotations.
2. User-decile activity labeling: We showed user contributions over subreddits in every other (1st, 3rd, 5th, 7th, 9th) decile (Fig. 2 (d)) and asked the evaluators to label the user activity by one of the four conspiracy pathways.

We randomly selected 12 users from each conspiracy pathways separately for each of the tasks. Each trajectory was labeled by two evaluators. We consider a true positive assessment for a trajectory only when both evaluators agree on the conspiracy pathway label. Evaluators labeled trajectories in task 1 with an accuracy of 78%, while 83% validation samples received perfect agreement. Task 2 resulted in accuracy of 84% accuracy with a perfect agreement of 96%. The validation performance across both tasks suggests that the computational conspiracy pathway assignments are cohesive and can be inferred through both, user trajectory plot (task 1) and the user's raw activity data (task 2).

Characterizing Conspiracy Radicalization

Next, we model the conspiracy radicalization process for users along the four conspiracy engagement pathways. We leverage the radicalization framework outlined in the RECRO model and operationalize the first three phases. Fig. 1 (d) displays the summary of all features.

Characterizing Reflection Phase

The Reflection phase captures the psychological predispositions of users towards adopting radicalization narratives online (Neo 2016). Predisposition towards radicalization can be visible through the psycho-linguistic footprints left by the users online (Dalgaard-Nielsen 2010). Specifically, researchers found that language reflecting anger, anxiety and heightened emotions was used by online radicalized groups (Dalgaard-Nielsen 2010).

Hence, to measure the language related to anger and anxiety, we use anger and anxiety lexicons, respectively, from Linguistic Inquiry and Word Count (LIWC) dictionary (Tausczik and Pennebaker 2010). LIWC encodes words capturing affective, emotional and cognitive processing expressions and is often used for psycho-linguistic analysis of online texts. To measure emotionality, we calculate the average compound VADER sentiment scores (Hutto and Gilbert 2014). In total, we calculate 3 linguistic features to characterize the Reflection phase.

Characterizing Exploration Phase

The Exploration phase describes a period in which users develop alternate worldviews that advance the radicalization process. Specifically in conspiracy theorizing, scholars have debated whether conspiracy theory belief evolves into a monological worldview—a tendency to analyze all events through the lens of conspiracy theorizing (Goertzel 1994). Previous researchers have concluded that online discussions could be useful in understanding the users' conspiracy worldview (Wood and Douglas 2015). To understand how online users explore the world of Reddit CTs, here we characterize conspiracy worldviews by calculating conspiracy generalist or specialist engagement. Specifically, we create a generality scale, that scores a subreddit based on the generality of topic discussions.

Creating generality scale One of the prominent perspectives in research on conspiracy world-views focuses on the "monological belief system" (Goertzel 1994). The monological perspective describes conspiracy belief as closed in itself in which, each conspiracy belief reinforces another. That is, belief in one CT is correlated with belief in other conspiracy theories (Swami et al. 2011). Such generalized conspiracy thinking stemming out of monological worldview could contribute towards more extreme belief in conspiracies. In subreddit generality scale, we want to capture the extent of generality (or specificity) of topics discussed in any subreddit. For example r/conspiracy hosts more general CT discussions compared to r/moonhoax which focuses on a specific moon landing conspiracy. Reddit houses thousands of such general and special discussion subreddits that allow users to participate in a topic with different levels of specialization. How can we computationally determine how generalist a subreddit is?

Previous researchers have proposed a generalist-specialist ranking for subreddits based on how generalist or special-

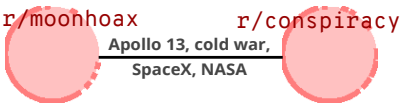

Figure 3: Figure showing example edge connections made in subreddit entity network. We connect two subreddits with an edge if they share same entities (e.g, Apollo 13, cold war) in top scoring submissions. The aggregate edge weight between two subreddits is the sum of inverse frequencies of shared entities across the entire corpus.

ist is the subreddit's user base (Waller and Anderson 2019). However, this scale does not consider the actual topic of discussion through the content in subreddit posts. In this paper, we instead build a subreddit-entity network that simultaneously captures the content discussed in a subreddit and how general that content is across Reddit.

Subreddit-entity network: Intuitively, more general subreddits will host content that is less exclusive across Reddit. For example, r/conspiracy hosts political conspiracies on topics also discussed in r/politics and medical conspiracies on topics in r/science. However, subreddits such as r/moonhoax host specific conspiracy topics, that are less likely to be popular across rest of the Reddit. We leverage this intuition and build "subreddit-entity network".

Subreddit-entity network is a graph in which subreddits are the nodes connected based on common content. To find the content representative of a subreddit (Horne, Adali, and Sikdar 2017), we analyze the top 200 submissions made in every subreddit and extract named entities (names of people, places, organizations etc.) from the submission text. We create an edge between two subreddits if top posts in both subreddits mention the same entity (see Fig. 3). To further improve the quality of edges, we consider the inverse term frequency of the shared entity across the entire corpus. The total edge weight between two subreddits is the sum of inverse term frequency of the entities shared between the subreddits.

In the subreddit-entity network, intuitively, more generalist subreddits will share more entities with other subreddits. In other words, generalist subreddits might be in a densely connected neighborhood connecting various subgraphs. Accordingly, generalist subreddits may be influential in the subreddit-entity network. Hence, to assess the degree of generality of a subreddit, we calculate the eigenvector centrality for all subreddits which is a measure of influence, where a node is considered to be influential if it is connected to other influential nodes. We consider the eigenvector centrality as the subreddit generality score where subreddits with higher eigenvector centrality hosting more general discussions.

Validating subreddit generality scale: Do eigencentrality values actually represent how generalist or specialist the subreddits are? We validate this in two different steps.
Step 1: We created the subreddit generality scale to assess the generality of discussions in CT subreddits. Hence, using the conspiracy similarity scale, we compile a list of 171 CT discussion subreddits and contrast the generality of their themes with their ranking in the subreddit generality scale. We manually examine the relative ranking of 171 CT discussion subreddits and validate that the subreddit generality scale places generalist CT subreddits on higher end and specialist subreddits on the lower end (example Table 2).
Step 2: For non CT related subreddits, we create 400 pairs where the first subreddit in the pair is more generalist compared to the second (r/Guitar → r/AcousticGuitar). Subreddits included in the 400 pairs range over a diverse list of 45 topics referred from r/ListOfSubreddits. We calculate the number of pairs for which our scale scores first subreddit higher than the second finding that 81% of the pairs are ranked correctly. This indicates that our scale is able to capture the overall generalist-specialist themes in subreddits.

Most general	Most specialist
r/C_S_T	r/SacredGeometry
r/HighStrangeness	r/flatearth
r/conspiracyundone	r/theworldisflat
r/conspiracy	r/AntarcticAnomalies
r/conspiracytheories	r/chemtrails

Table 2: Most general and most specialist conspiracy discussion subreddits on the subreddit generality scale.

r/chemtrails	r/AncientAliens	r/McDonalds
contrail	sitchin	mcdouble
chemtrail	anunnaki	mcchicken
geoengineering	nibiru	frappe
stratospheric	panspermia	mcflurry
haarp	gobekli	mcnugget

Table 3: Example of top 5 words (out of 1000) in SAGE lexicons with first two columns showing CT discussion subreddits. We see that our lexicon captures vocabulary that is distinctively specific to the subreddit. For example, 'haarp" in r/chemtrails refer to H.A.A.R.P project by U.S. Air Force that is theorized to be a weather control weapon. Similarly, anunnaki in r/AncientAliens is believed to be a race of ancient aliens. Note that outside of the CT groups, these words have little meaning, making them especially relevant for measuring group language conformity.

Generalist engagement scores With the generality scale we can now measure whether users increasingly engage in more generalist or more specialist CT discussion subreddits. Using the same computation as the conspiracy engagement score, we calculate the generalist engagement score in every decile as the weighted average of a user's contributions in subreddits weighted by the subreddits' generality score. Higher generalist score would indicate high engagement in subreddits with general discussions.

Characterizing Connection Phase

In the Connection phase, individuals form social connections to support and reinforce their alternate worldviews, facilitating the relational bond between an individual and

a wider radical movement (Neo 2016). To characterize the this phase, we capture language conformity and group connections established by users with conspiracy communities.

Language conformity: Cohesion or conformity with one's social group is a fundamental requirement in the process of radicalization (Crossett and Spitaletta 2010). Especially in CT discussions, groups conform by establishing shared interpretations of reality around them (Butter and Knight 2020). This process of interpreting reality, or meaning-making, often manifests into the insider language used by conspiracy groups (Leone 2017). Hence, we measure language conformity by assessing how much of the subreddit's characteristic language does the user use.

To measure a user's language conformity, we consider each subreddit the user contributes in as her "social group" and calculate the overlap between the language used by the user and that subreddit's characteristic language. To understand the characteristic language for each subreddit, we utilize Sparse Additive Generative models (SAGE) (Eisenstein, Ahmed, and Xing 2011) that uses a regularized log-odds ratio to compare word distributions across various text corpora. We compare the word distributions in the text corpus of each subreddit with that of all other subreddits using SAGE. As a result, for each subreddit we obtain a lexicon of 1000 words that distinctively represent the language used in that subreddit. Table 3 displays example lexicon words for various subreddits showing how SAGE is able to effectively capture the language specific to the discussions in each subreddit. Finally, we calculate a user's language conformity in a subreddit s as the intersection of words used by the user in subreddit s with the SAGE lexicon of the subreddit, normalized by total word count used by the user in s.

Group connections: Interactions within small groups of like minded people can help in creating unambiguous shared narrative of events that is instrumental to the process of radicalization (Reedy, Gastil, and Gabbay 2013). Discussions within small groups can also limit the number of diverse opinions and information users might get exposed to, thus contributing to what researchers call as "crippled epistemeology" (Vermeule and Sunstein 2009). Hence, we characterize users's small group interactions by analyzing the diversity of audience and repeated contributions in comment threads populated by users.

First, we measure the diversity of audience in threads, or thread diversity, by comparing number of unique contributors to the total number of comments in a thread. While calculating the thread diversity ratio, we remove all contributions made by the subject user so as to not bias the ratio calculation by the user's contributions. Low thread diversity would indicate that users engage in comment threads where limited number of other users contribute large number of comments. Second, to measure a user's involvement in grouped discussions, we calculate comment rank—the number of times a user repeatedly contributes in the same thread. Finally, we also measure group connections by analyzing how users express affiliation to their groups using LIWC's affiliation lexicon.

Analyzing Evolution of Radicalization Phases

1. Linear regression fits for trends: For each user on each conspiracy engagement pathway, we have ten values of all features corresponding to every decile. To understand how features evolve over time, we fit a linear regression line with deciles as the independent variable and the feature value as the dependent variable. The magnitude of the fit coefficient (β) tells us the degree of increase and decrease over time and the p-value indicates whether the fit is significant. We display all trends lines and coefficients inside the 171 CT subreddits in Fig. 4 along with the coefficients for trends outside of conspiracy subreddits as well. Trend coefficients and significance outside CT show whether the trends we observe inside are specific to conspiracy discussion.

2. Users' conspiracy tenure with peak feature values: We characterize phases of radicalization that, in theory, take effect one after another. Hence we next analyze how soon after CT joining, users attain peak values for features. For every user, we pick the decile with highest feature value and plot the density distribution of peak decile for all users grouped by conspiracy pathway. Density peak in early deciles would mean that users attain highest value for that feature immediately after initial participation in conspiracy discussions. All density plots are based on users' activity inside CT subreddits.

3. Users' conspiracy tenure with peak phase features: While the previous analysis displays how individual features peak across decile, here we group the feature belonging to same radicalization phases and plot similar peak density plots for each pathway (Fig. 5). For example, Fig. 5 (a) indicates that for steady high pathway, Reflection features peak immediately after initial conspiracy participation whereas Exploration and Connection peak later in the CT journey. Visualizing this phase progression can inform whether users develop RECRO phases successively in time.

Conspiracy Radicalization: Results

How do users on display markers of reflection phase? We characterized the Reflection phase using anger (Fig. 4 (a)), anxiety ((Fig. 4 (b)) and emotionality (Fig. 4 (c)), expressed inside and outside of CT subreddits. Overall we find that use of language related to anger and anxiety decreases over time for users on all pathways. There are no significant trends for emotionality inside or outside of CT, except for users on the increasing pathway who show increasing emotionality inside CT over time ($\beta = 4e^{-3}$). Earlier peaks in Reflection features may indicate what Neo (2016) describes as "cognitive opening" where individuals turn to internet to express their grievances and vulnerabilities. Do all users advance to subsequent phases of radicalization after the cognitive opening? To find out, we next analyze the results of the Exploration phase.

How do users explore Reddit's conspiracy world? To characterize the Exploration phase, we investigate how users develop conspiracy worldview through generalist engagement (Fig. 4 (d)). Higher generalist engagement would indicate that users engage in general CT discussions, thus devel-

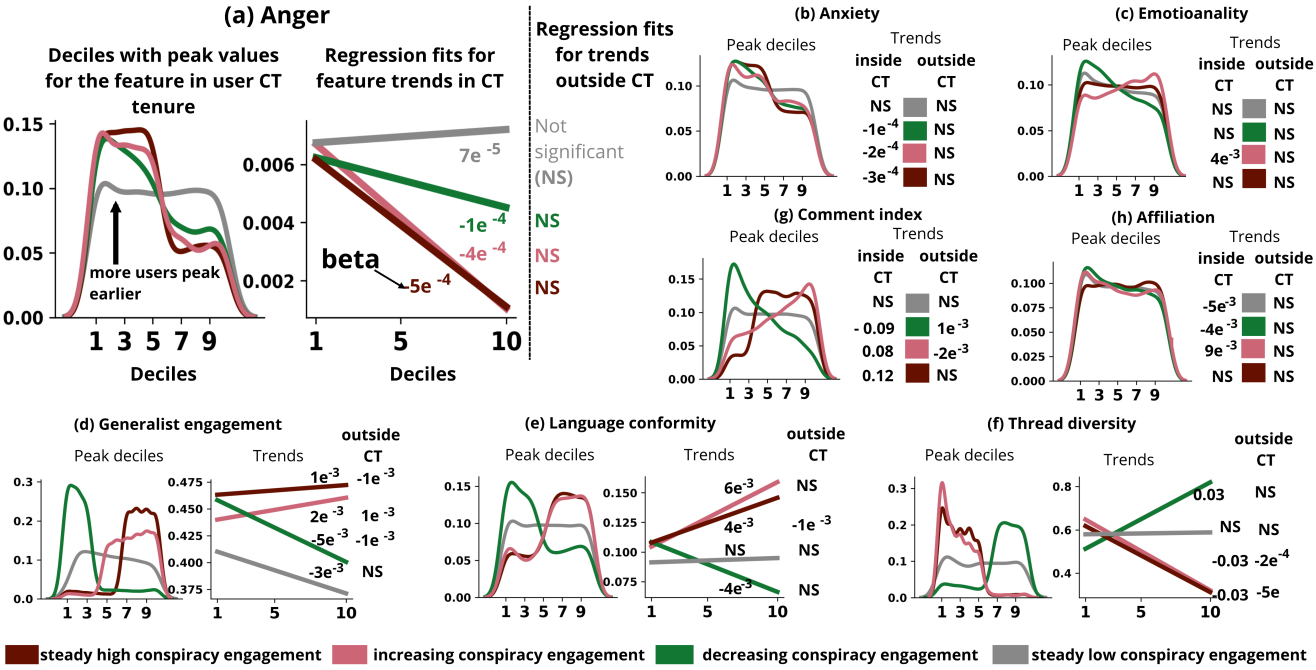

Figure 4: Figure presenting the peak deciles distributions and linear regression trends grouped by the conspiracy engagement pathways. In (a) we present an enlarged view of a typical result showing deciles with peak values and trends inside and outside of CT subreddits. For all features, peak decile distributions represent the density plots for deciles at which the users attain the highest feature value. For example, users on all pathways show highest values for anger in earlier deciles (subfigure a). Trend in each subplot represents the linear regression fit for various pathways over deciles. For every line we denote the β coefficient if the trend is significant. Non significant trends ($p > 0.05$) are denoted with NS. We show trend coefficients for feature calculated both, inside and outside CT subreddits. Due to space limitations we show the actual trend lines only for 4 features.

oping monological worldview. We find that users on steady high ($\beta = 1e^{-3}$), and increasing pathway ($\beta = 2e^{-3}$) increasingly participate in the generalist CT subreddits. Interestingly, steady high users show reduced generalist engagement outside of CT subreddits ($\beta = -1e^{-3}$). Conversely, users on decreasing pathway increasingly contribute in specialist CT discussions ($\beta = -5e^{-3}$) and have highest generalist engagement only in the earlier deciles. We ran an additional robustness check to ensure that this result is not an artifact of correlation between conspiracy similarity scale and generality scale. Overall, we find that users with steady high and increasing CT engagement may also adhere to monological conspiracy worldview by increasingly participating in general conspiracy discussion subreddits.

How do users make connections inside conspiracy communities? We measure group bonding through language conformity (Fig. 4 (e)), audience diversity in threads (Fig. 4 (f)), repeated comments (Fig. 4 (g)) in threads, and affiliation related language (Fig. 4 (h)). Overall, we find that users on steady high ($\beta = 4e^{-3}$) and increasing ($\beta = 6e^{-3}$) pathways develop high language conformity with CT subreddits. However, users on steady high pathways show reduced language conformity outside of CT subreddits ($\beta = -1e^{-3}$). Interestingly, users on decreasing pathway ($\beta = -4e^{-3}$), despite exhibiting early high engagement, never develop as high language conformity with CT subreddits in comparison to the other cohorts. Hence, early lexical conformity could be one of the important precursor of sustained CT engagement. Users with steady high engagement also participate repeatedly (β=0.12) in smaller discussion groups with less audience diversity (β = -0.03). Users on increasing pathways also show similar trends. These results suggest that users on steady high and increasing pathways repeatedly show engagement in discussions with less diverse user base inside CT subreddits.

How does conspiracy radicalization evolve? Fig. 5 shows the deciles in which users attain peak feature values in different phases. We observe that in steady high (Fig. 5 (a)) and increasing (Fig. 5(b)) pathways, users show higher feature values in the Reflection phase right after starting CT participation and develop high Exploration and Connection feature values in later deciles. This may suggest that the internet-mediated conspiracy radicalization does evolve through different phases over time. Interestingly, users on decreasing pathway show high feature values for all phases only early on, while for users on steady low pathway, there is no discernible peak for these phases. We discuss the implications of these results in the discussion section.

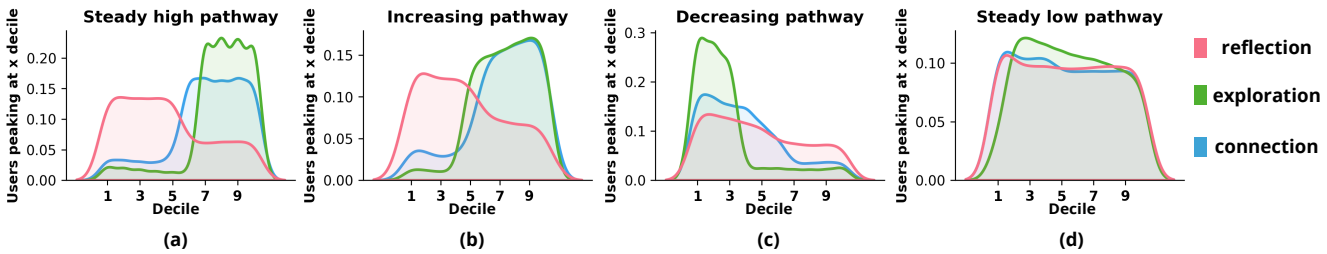

Figure 5: Figure outlining peak deciles for features in various radicalization phases for (a) steady high (b) increasing (c) decreasing and (d) steady low pathways. Peak in earlier deciles means that users attain highest values for that phase, early on.

Discussion and Limitations

In this paper, we present the first, large-scale modeling of long term engagement and radicalization in the online conspiracy communities using a longitudinal digital trace data of 36K Reddit users. Below we discuss how our findings may impact the understanding of online CT participation and motivate further research.

Monologicality as a varying process: While not all conspiracy believers adopt a general conspiratorial worldview as the primary sensemaking device (Franks et al. 2017b), we find that two groups of users (those on steady high and increasing engagement trajectories) do contribute prominently in general conspiracy subreddits that host all types and topics of conspiracies. Are users on steady high and increasing pathways predisposed to monological thinking? The original theoretical proposition by Goertzel (1994) describes monologicality as a stable cognitive style, trait or disposition. However, looking specifically at users on increasing pathway, the discussion spaces they engage in become more generalist over time. This suggests that monological conspiracy worldviews can develop over time. In fact, our quantitative results align with qualitative observation of Franks et al. (2017b) depicting monologicality as a variable endpoint of various social processes rather than a cognitive predisposition. In particular, counter to the popular rabbit-hole metaphor, individuals who show signs of radicalization do not seem to narrow their interests down to fringe theories. Instead, such individuals adopt venues of generalist CT discussions together with their idiosyncratic lingo. This observation, on the one hand, purports a parallelism between monological worldviews and radicalization. On the other, it begs the question of what types of online discussion environments harbor the potential for mobilization of radicalized individuals: topically and socially fringe spaces that may host extreme ideas, or comparatively mainstream spaces that afford perception biases of false consensus?

Resolution and Operational Phases: We offer a systematic characterization for how radicalization in online CT discussions progresses, finding that users on steady high and increasing pathways do progress through these first three RECRO phases during their conspiracy tenure. The theoretical model suggests that users may then enter a resolution and operational phase (Neo 2016). The resolution phase describes a period in which individuals gain momentum to convert their radical beliefs into action. While not all users that internalize radical beliefs actually act on them, those who do, act from a biased perspective solidified during the earlier stages (Neo 2016). Simply put, users who go through all three, Reflection, Exploration and Connection phases over time, change not just their beliefs but also their behaviors. Finally, the operational phase indicates a period in which individuals get mentally or physically prepared to commit acts that advance their radical objectives in the real world. They may influence others or actively look for openings to form plans of physical actions (Neo 2016).

Recovery from online conspiracies: Perhaps one of the most unexplored, yet highly impactful research directions is to understand why and how users disengage from online conspiracy discussions. In this paper we provide quantitative evidence that a significant group of users do gradually decrease their participation in online conspiracy discussions while displaying measurably distinctive behaviors compared to steady high and increasing trajectories. For example, users on decreasing pathway do not develop lexical conformity with conspiracy subreddits, engage in discussions containing diverse contributors and significantly reduce affiliation-related language in conspiracy subreddits. These observations provide valuable insights for understanding the process of recovery from conspiracy engagement. Our results provide ground to investigate conspiracy believers that do not adopt a monological conspiracy worldview separately from those who do. By understanding the difference between these two types of online conspiracy beliefs, our findings can help refocus research efforts on communities that have higher chances of advancing conspiracy radicalization. Conversely, the methods detailed in this paper can help in identifying users on decreasing pathways who might be more receptive of cross-cutting narratives, and in transforming insights gained by studying them into design interventions, to counter the spread of disinformation and conspiracy theory radicalization.

Limitations and Future Work

Our work has some limitations that should be acknowledged. First, we characterize the conspiracy engagement trajectories by using the contribution volumes of users, a measure commonly used as a strong latent proxy for user engagement (Hamilton et al. 2017). While this affords studying the complete evolution of contributions in the CT com-

munities, analyzing contribution volume alone can limit the interpretability of quality of contributions. For example, it is possible that some users may be contributing troll posts while keeping the same contribution volume as others. A more nuanced measure of CT engagement could involve analyzing text and context of the user contributions. Second, this work offers empirical insights on how users escalate through the formative phases of radicalization. Yet, it would be crucial to unpack when and how this potential is turned into action in the Resolution and Operational phases. Our work provides a framework for experimental designs in this direction. Next, our characterization of CT engagement trajectories relies on subreddit contributions. We found that five subreddits higher up in the conspiracy similarity scale were banned before 2020, which could have potentially affected the CT disengagement of some users. Currently, the literature examining the effect of subreddit bans on user engagement poses mixed results claiming that the bans are (Chandrasekharan et al. 2017; Thomas et al. 2021) or aren't (Habib et al. 2019) effective in specific cases they study. While the banned conspiratorial subreddits in our dataset made up for less than 0.2% of the dataset volume, a more individualized investigation of the authors who prominently contributed in banned communities could reveal whether the subreddit bans actually affected the users' disengagement. Moreover, this paper observes the radicalization process from aggregated user activity. Qualitative analyses of CT narratives and of how those change across radicalization phases, should complement our work and provide a fuller understanding of the phenomenon. Also, this paper offers findings that are correlational in nature. Causal models could reveal more nuanced relationship between conspiracy engagement pathways and the radicalization process.

Additional Robustness Checks

1. User selection based on r/conspiracy: To characterize users' conspiracy radicalization process we select users that make at least 20 comments in r/conspiracy over at least 1 year of time. How is our user selection affected by this r/conspiracy constraint? To understand, we calculate the proportion of users' activity in other conspiracy related subreddits before their first comment in r/conspiracy. We find that 1,689 (4%) of the users have more than 10% of their total activity in other conspiracy subreddits. We remove the 1,689 users from entire analysis to ensure that r/conspiracy is a common starting point into Reddit's conspiracy world for the users in our study.

2. Coverage for conspiracy similarity scale: Our conspiracy similarity scale has conspiracy similarity value for only 4K subreddits. Does this mean we are missing out on modeling a large chunk of user activity while extracting conspiracy engagement pathways? We performed activity coverage analysis and found that just considering 4K subreddits in conspiracy similarity scale, 90% of authors have more than 80% coverage of their total post-conspiracy joining Reddit activity. We removed the rest of the 10% authors to ensure large coverage for all studied users.

3. Correlation between conspiracy similarity and generality scale: As we compare the generalist engagement for users on different conspiracy pathways, we wanted to ensure that two scales operating underneath are not correlated. Meaning, we wanted to check whether high conspiracy engagement inherently result in high generalist engagement due to our operationalizations. We performed Spearman rank correlation between subreddit rankings of the two scales and found only a weak correlation of 0.23.

4. Effect of user removals: To maintain the integrity of analysis across all users, we split the user activity in 10 deciles of equal contribution volumes. Meaning, all users have measurable activity in all of the deciles of their Reddit trajectory. This ensures that if a user gets banned, all activity before the ban will be studied across 10 deciles, indicating that user bans could not explain the observed disengagement in the user's trajectory prior to the ban.

5. Effect of subreddit bans: To understand how subred-dit bans could affect the disengagement, we compiled a list of over 3000 banned subreddits. Specifically, we sourced the list of banned subreddits from r/reclassified sub-reddit which maintains an up-to-date list of banned sub-reddits. Since, our dataset contains user activities prior to 2020, we first identified subreddits that were banned prior to 2020. Out of the 4,756 subreddits in the dataset, 21 were banned prior to 2020. Out of the 21 banned subred-dits, 5 subreddits lie in top 500 subreddits on the conspir-acy similarity scale, potentially affecting the engagement and disengagement trajectories of users. Before bans, the above mentioned five subreddits produced the following per-cent volume in the dataset—r/greatawakening (0.1%), r/altright (0.01%), r/uncensorednews (0.008%), r/911truth (0.003%), r/sjwhate (0.001%). Together, these subreddits make up for less than 0.2% data before getting banned.

Conclusions

In this paper we investigate the association between online CT discussion engagement and radicalization. Through an ensemble of computationally derived features backed by theoretical models, we observe three radicalization phases—Reflection, Exploration, Connection—across four conspiracy engagement pathways. We find that high or increasing engagement in CT discussions online is also associated with successive phases of online radicalization symbolizing psychological predisposition, adoption of alternate worldviews and social bonds with others in CT communities. Conversely, users with decreasing engagement show qualitatively different CT interest compared to other users, limiting their CT discussion tenure to specialized conspiracy topics. Our results have implications in understanding the conspiracy recovery process.

Ethics Statement

We refer to the AAAI code of conduct and ethics guidelines that mention stakeholders, harm, privacy and confidentiality dimensions of ethical research and conduct. First, we acknowledge that all people, especially social media users and social computing researchers are stakeholders in this research. With this paper, we intend to contribute insights

that can be considered while building safer online spaces for all. Furthermore, given that this study is retrospective and involves no interaction with the studied population, we do not anticipate any direct harm resulting from this research. We take proactive steps to preserve user privacy. Specifically, by presenting results aggregated over thousands of users and by intentionally not reporting any exact quotes made by Reddit users, we reduce the risk of re-identification. Finally, throughout this study, we analyze non-confidential Reddit data that is available in the public domain, collected through the publicly accessible Reddit Pushshift API (Baumgartner et al. 2020). Yet, given the potential stigma associated with participating in CT discussions, we do not release any raw user data from this study.

Acknowledgments

We thank the members of Social Computing and Algorithmic Experiences (SCALE) Lab at University of Washington for their valuable feedback on this work. This research was supported by Office of Naval Research (ONR-YIP #N00014-21-1-2748), a US Navy/DOD Minerva (#N00014-21-1-4001), and an NSF grant IIS (#2041068).

References

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In In Proc. of ICWSM, volume 14, 830–839.

Bessi, A.; Coletto, M.; Davidescu, G. A.; Scala, A.; Caldarelli, G.; and Quattrociocchi, W. 2015. Science vs Conspiracy: Collective Narratives in the Age of Misinformation. PLOS ONE 10(2): e0118093. ISSN 1932-6203. doi: 10.1371/journal.pone.0118093. URL http://dx.plos.org/10.1371/journal.pone.0118093.

Blake, A. 2021. anti-Vax theorists become desperate after full FDA authorization. https://tinyurl.com/3292ab2s. Accessed: 2022-01-01.

Bouma, G. 2009. Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL 30: 31–40.

Butter, M.; and Knight, P. 2020. Routledge handbook of conspiracy theories. Routledge.

Chandrasekharan, E.; Samory, M.; Srinivasan, A.; and Gilbert, E. 2017. The Bag of Communities Approach: Identifying Abusive Behavior Online with Preexisting Internet Data. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '17. ISBN 9781450346559. doi:10.1145/3025453.3026018.

Conway, M. 2012. From al-Zarqawi to al-Awlaki: The emergence of the internet as a new form of violent radical milieu. CTX: Combatting Terrorism Exchange 2(4): 12–22.

Crossett, C.; and Spitaletta, J. 2010. Radicalization: Relevant psychological and sociological concepts. JHU .

Dalgaard-Nielsen, A. 2010. Violent radicalization in Europe: What we know and what we do not know. Studies in conflict & terrorism 33(9): 797–814.

Douglas, K. M.; Sutton, R. M.; and Cichocka, A. 2017. The psychology of conspiracy theories. Current directions in psychological science 26(6): 538–542.

Eisenstein, J.; Ahmed, A.; and Xing, E. P. 2011. Sparse additive generative models of text. In Proc. ICML-11.

Franks, B.; Bangerter, A.; Bauer, M. W.; Hall, M.; and Noort, M. C. 2017a. Beyond "monologicality"? Exploring conspiracist worldviews. Frontiers in Psychology 8(JUN). ISSN 16641078. doi:10.3389/fpsyg.2017.00861.

Franks, B.; Bangerter, A.; Bauer, M. W.; Hall, M.; and Noort, M. C. 2017b. Beyond "monologicality"? Exploring conspiracist worldviews. Frontiers in psychology 8: 861.

Garrett, M. 2021. Capitol riot exposes reach of QAnon disinformation: "It was a drug" - CBS News. https://www.cbsnews.com/news/qanon-capitol-riot-reach/. Accessed: 2022-01-01.

Genolini, C.; Ecochard, R.; Benghezal, M.; Driss, T.; Andrieu, S.; and Subtil, F. 2016. kmlShape: an efficient method to cluster longitudinal data (time-series) according to their shapes. Plos one 11(6): e0150738.

Goertzel, T. 1994. Belief in conspiracy theories. Political psychology 731–742.

Habib, H.; Musa, M. B.; Zaffar, F.; and Nithyanand, R. 2019. To act or react: Investigating proactive strategies for online community moderation. arXiv preprint arXiv:1906.11932 .

Hamilton, W. L.; Zhang, J.; Danescu-Niculescu-Mizil, C.; Jurafsky, D.; and Leskovec, J. 2017. Loyalty in Online Communities URL http://arxiv.org/abs/1703.03386.

Horne, B. D.; Adali, S.; and Sikdar, S. 2017. Identifying the social signals that drive online discussions: A case study of reddit communities. In 2017 26th ICCCN, 1–9. IEEE.

Hutto, C.; and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the AAAI ICWSM, volume 8.

Javed, A.; Lee, B. S.; and Rizzo, D. M. 2020. A benchmark study on time series clustering. Machine Learning with Applications 1: 100001.

Klein, C.; Clutton, P.; and Dunn, A. G. 2019. Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit's conspiracy theory forum. PloS one 14(11).

Kuhn, S. A. K.; Lieb, R.; Freeman, D.; Andreou, C.; and Zander-Schellenberg, T. 2021. Coronavirus conspiracy beliefs in the German-speaking general population: endorsement rates and links to reasoning biases and paranoia. Psychological medicine 1–15.

Kumar, S.; Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2018. Community interaction and conflict on the web. In Proceedings of the 2018 WWW conference, 933–943.

Leone, M. 2017. Fundamentalism, Anomie, Conspiracy: Umberto Eco's Semiotics against Interpretive Irrationality. In Umberto Eco in his Own Words, 221–229. De Gruyter Mouton.

Marwick, A.; and Lewis, R. 2017. Media Manipulation and Disinformation Online. Data & Society Research Institute 1–104.

Neo, L. S. 2016. An Internet-Mediated Pathway for Online Radicalisation (January 2016): 197–224. doi:10.4018/978-1-5225-0156-5.ch011.

Neo, L. S. 2020. Detecting markers of radicalisation in social media posts: the role of person-centric and psychosocial risk factors, and protective factors. Ph.D. thesis, Nanyang Technological University, Singapore.

Phadke, S.; Samory, M.; and Mitra, T. 2021a. Characterizing Social Imaginaries and Self-Disclosures of Dissonance in Online Conspiracy Discussion Communities. In CSCW .

Phadke, S.; Samory, M.; and Mitra, T. 2021b. What Makes People Join Conspiracy Communities? Role of Social Factors in Conspiracy Engagement. In Proc. CSCW 4.

Reedy, J.; Gastil, J.; and Gabbay, M. 2013. Terrorism and small groups: An analytical framework for group disruption. Small group research 44(6): 599–626.

Samory, M.; and Mitra, T. 2018. Conspiracies online: User discussions in a conspiracy community following dramatic events. In Proceedings of the 12th AAAI ICWSM, volume 12.

Swami, V.; Coles, R.; Stieger, S.; Pietschnig, J.; Furnham, A.; Rehim, S.; and Voracek, M. 2011. Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories. British Journal of Psychology 102(3): 443–463. ISSN 20448295. doi:10.1111/j.2044-8295.2010.02004.x.

Tausczik, Y.; and Pennebaker, J. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. Journal of language and social psychology 29(1).

Thomas, P. B.; Riehm, D.; Glenski, M.; and Weninger, T. 2021. Behavior Change in Response to Subreddit Bans and External Events. IEEE Transactions on Computational Social Systems .

Van Raemdonck, N. 2019. The echo chamber of anti-vaccination conspiracies: mechanisms of radicalization on Facebook and Reddit. IPAG Knowledge Series .

Vermeule, C. A.; and Sunstein, C. R. 2009. Conspiracy theories: causes and cures. Journal of Political Philosophy .

Vidgen, B.; Yasseri, T.; and Margetts, H. 2021. Islamophobes are not all the same! A study of far right actors on Twitter. Journal of Policing, Intelligence and Counter Terrorism 1–23.

Waller, I.; and Anderson, A. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In The WWW Conference.

Wood, M. J.; and Douglas, K. M. 2015. Online communication as a window to conspiracist worldviews. Frontiers in psychology 6: 836.

Xiao, S.; Cheshire, C.; and Bruckman, A. 2021. Sensemaking and the Chemtrail Conspiracy on the Internet: Insights from Believers and Ex-believers. Proceedings of the ACM on Human-Computer Interaction 5(CSCW2): 1–28.