OLIVIA BROWN

RIGHT-WING EXTREMISM ONLINE:
CAN WE USE DIGITAL DATA TO MEASURE RISK?

The internet plays an important role in the rising threat of right-wing terrorism. Olivia Brown and colleagues have combined psychology and computational science methods to identify whether online behaviour can be used to infer the risk of offline action.

BACKGROUND

The threat of right-wing extremism is growing globally, with statistics showing a 320% increase in right-wing terrorist offences in the past six years. Evidence suggests the internet is playing a key role in this growth, with online forums and social networking sites providing the opportunity for individuals to share ideas, recruit new members, as well as offer a medium through which to acquire ideology and plan attacks (Scrivens, Gill, and Conway, 2020). This can be illustrated in recent high-profile incidents such as the Christchurch terrorist attack and Pittsburgh Synagogue Shooting, in which the perpetrators posted about their intentions online in the weeks and moments preceding their violent attacks.

In an increasingly digital world, the role of the internet in the planning and execution of terrorism presents law enforcement with an opportunity to build technological tools to assess online communications and detect risk. Supported by research in social psychology, there is a growing consensus that digital data may indicate when and how interactions online might lead to right-wing extremist violence offline. However, the volume of extremist content makes it challenging to identify which individuals pose a risk to public safety. The challenge of identifying these 'needles in the haystack' has been exacerbated by the pandemic, in which we have witnessed an unprecedented rise in extreme-right wing content, with members of the far-right exploiting anti-vaccine and anti-authority sentiment.

RESEARCH

With right-wing extremist content on the rise across mainstream and dark-web platforms, questions remain as to whether there are specific markers of online behaviour that can be used to infer risk. Our research begins to address this question by modelling the online interactions of right-wing extremists across three far-right platforms. Existing methods have tended to focus on large scale quantitative analysis of entire platforms, identifying patterns of posting and indicators of extreme content. While this provides an overview of the far-right online context, it cannot offer any indication as to how to identify users who may be at risk of committing a violent offence. Unique to our approach is the inclusion of data from individuals who have been convicted of a terrorism-related offence and those who have not. By using conviction as an independent variable and tracing our digital data back to specific individuals, our data presents a unique opportunity to develop insight on risk.

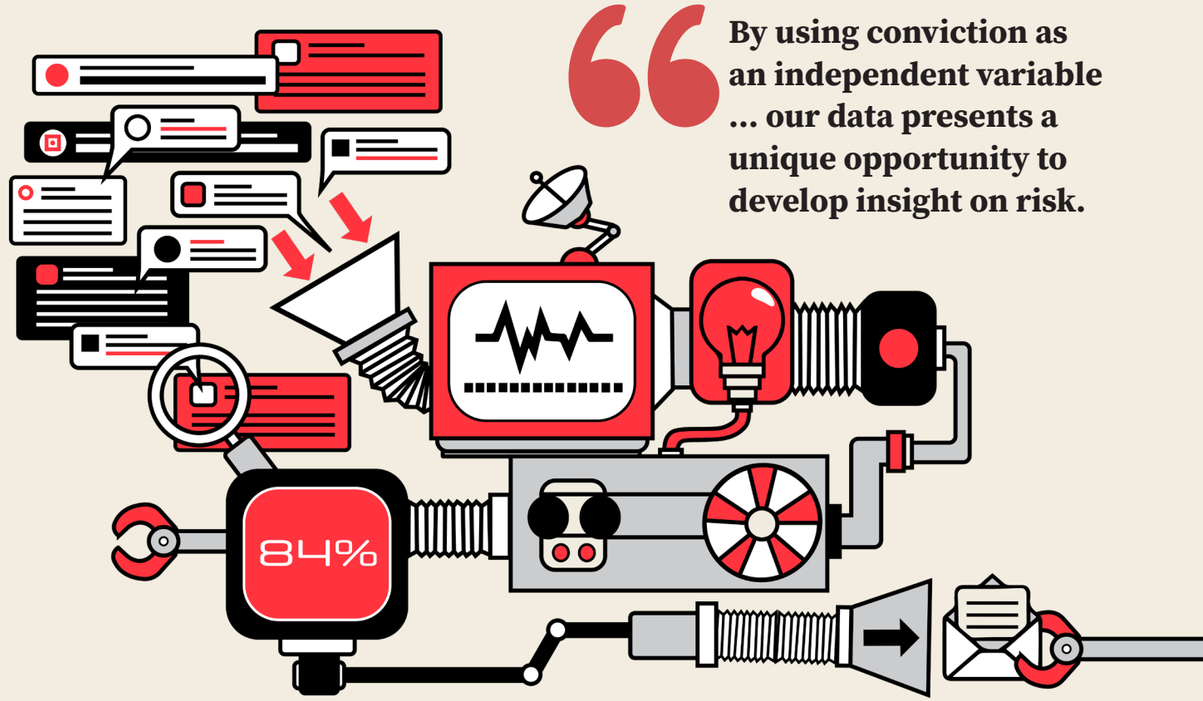
To compare convicted and non-convicted right-wing extremists, we obtained a sample of online postings and metadata that could be matched to individuals from their online aliases. All data obtained were publicly available and identified through open-source intelligence. We adopted strict inclusion and exclusion criteria to ensure that individuals (either convicted or not) were correctly matched to their online aliases. Our sample included 180,000 posts across three far-right forums (Gab, Discord, and Iron March) from 26 convicted and 54 non-convicted right-wing extremists.

We adopted a novel methodological approach to our analysis by combining qualitative and quantitative tools. First, a qualitative content analysis was conducted on 28,000 posts from eight convicted and eight non-convicted users. The qualitative analysis helped establish an in-depth understanding of the context of the research and began comparing users according to their conviction status. Notably, the results from the qualitative analysis were then used to inform our quantitative analysis. We adopted a computational approach to the quantitative analysis, in which we ran topic models to acquire features that could be used in a machine learning algorithm to predict conviction status based on post content.

WHAT WE FOUND AND WHAT IT MEANS

In the qualitative content analysis we identified 9 higher-order categories representative of the data:

1. Hateful content
2. Group Formation
3. Information Sharing
4. Intra-Group Debate
5. Operational and Security focused content
6. Threats of Violence
7. Offline Action
8. Weapons
9. Inciting Violence



> **By using conviction as an independent variable … our data presents a unique opportunity to develop insight on risk.**

For example, Group Formation represented content focused on forming groups, recruitment and connecting different online users together:

> *"Hi mate, I'm one of the main organisers with [REDACTED] in the UK. Who's currently in charge of [REDACTED]? I want to establish contact. Thanks"*

When comparing the two groups (convicted and non-convicted right-wing extremists), we found significant differences in each category of posts apart from Inciting Violence. Convicted users posted more of each category of content, apart from Intra-group Debate. Interestingly, this finding demonstrates that the convicted users were much more likely to be participating in direct action – whether that be through sharing files and websites (Sharing Information), discussing ways to avoid detection by the authorities (Operational and Security), working to recruit and connect others (Group Formation) or engaging in offline activities (Offline Action). Non-convicted users, on the other hand, were much more likely to discuss ideological positions, strategise about the movement going forward and debate historical events (Intra-Group Debate).

Next, we conducted topic modelling on the full dataset of 180,000 posts. We iterated through the topic models (ranging from 3-10 topics) to identify topics that were coherent with the qualitative analysis. Four topics were retained for further analysis - Hateful Content, Ideological Debate, Violence, and Intra-Group Connections. In the next step, we ran machine learning models to predict the conviction status of users (convicted versus non-convicted) using the degree to which their posts fit the topics (i.e., the probability distribution of each topic for each post) as features in the model. Our model currently demonstrates 84% accuracy when detecting whether a post belongs to a convicted or non-convicted user. When accounting for our unbalanced dataset (i.e., we have more posts from non-convicted than convicted users), the accuracy reduces to 65%. Notably, the findings demonstrate very high accuracy when detecting whether someone is non-convicted – 92%. This would suggest that our findings could be used to reduce the volume of digital data that requires close monitoring by the authorities by highlighting posts indicative of reduced risk (i.e., posts by individuals who, while expressing extremist views, have not engaged in activities that meet the threshold of illegal action).

CONCLUSION

With the volume and accessibility of extremist content online increasing, government agencies must continue to grapple with the challenge of identifying individuals who might be a risk to public safety. The methodology and findings presented here demonstrate promise and suggest cautious optimism in developing technological tools to narrow down the number of individuals in need of close monitoring online. We intend to continue building on this research and work towards developing algorithms that can identify 'bigger needles' and create 'smaller haystacks'.

Dr Olivia Brown is a lecturer at the University of Bath. She is interested in how intra- and inter-group processes influence individual and group behaviour.