



Tracking the 2023 Wave of Anti-Trans and Anti-Drag Mobilisation in Australia

Matteo Vergani
Andrea Giovannetti
Dan Goodhardt



Vergani, M., Giovannetti, A., & Goodhardt, D. (2025). Tracking the 2023 wave of anti-trans and anti-drag mobilisation in Australia. Deakin University.

This report contains discussions of hateful and distressing language, including explicit examples, as well as themes of violence and discrimination. Some content may be confronting or upsetting for readers. Please engage with this material at your own pace and seek support if needed.

We would like to acknowledge funding from the Commonwealth Government, together with project support from the NSW Government.

Contents

3	Executive Summary
5	Introduction
11	Linking Online Activity To Offline Events
18	Insights From Australian Tweets
26	Conclusion

Executive Summary

This report analyses the relationship between online hate and offline events targeting transgender and gender-diverse communities in Australia. Drawing from a dataset of over 86,000 tweets, we examine patterns in digital discourse, the role of far-right networks, and potential early warning indicators of violent mobilisation. The findings offer insights for policymakers, law enforcement and grassroots organisations seeking to mitigate online hate and prevent violent mobilisation against gender diverse communities in Australia.

Key Findings

1. Simultaneous Surge in Online and Offline Anti-Trans Mobilisation

Our report shows that a significant increase in both online hate and offline mobilisation occurred around March 2023, coinciding with the Australian tour of British activist Posie Parker. We propose that the rise of this movement can be attributed to a convergence of factors, including transnational influence, post-pandemic opportunism, a coalition of diverse fringe groups, media amplification, and conspiratorial narratives.

2. Australian far-right networks coordinate information dissemination

Analysis of user networks reveals that Australian far-right groups played a key role in amplifying and coordinating anti-transgender and anti-drag narratives. These actors increased their online influence around major protest events, particularly the Let Women Speak rally in Melbourne on 18 March, which saw confrontations involving neo-Nazis, protesters, and police. Other far-right activity peaks were observed in Goulburn, Newcastle, Eltham, Oakleigh, and Perth, demonstrating a consistent pattern of online mobilisation around key events.

3. Anti-Politics Discourse as a Potential Early Indicator of Violence

Our analysis found a strong association between online hostility towards politicians (anti-politics discourse) and violent offline events. Prior to events characterised by violence, anti-politics content significantly increased. This suggests that heightened hostility towards politicians in the context of anti-trans and anti-drag mobilisation may serve as a potential early warning signal for offline violence in the context of a social movement mobilisation.

4. Persistence of Hate Beyond Events

While explicit discussions of events tend to fade quickly after they occur, broader hateful language persist and even grow over time. In some cases, hateful discussions about diverse genders first emerge within event-related discourse before diffusing into wider online discussions. This suggests that preventing the spread of hateful narratives requires longer-term monitoring and intervention, not just immediate responses to high-profile events.

5. Geographical Distribution of Online Hate Has Multiple Factors

Tweets expressing hate toward gender-diverse individuals were most concentrated in New South Wales and Victoria, aligning with population size and the number of events hosted. However, the proportion of hateful tweets did not always correlate with event frequency, suggesting that other factors—such as local social dynamics and the visibility of gender-diverse communities—also influence hate levels.

6. Global Anti-Trans Influencers and Information Sources Impacted Australian Mobilisation

The most frequently mentioned figures in hateful tweets were Posie Parker and American conservative commentator Matt Walsh, indicating the influence of international anti-trans movements in shaping Australian discourse. Additionally, international fringe media platforms and women's rights networks with anti-trans views were among the most frequently cited domains, reinforcing the role of targeted misinformation and ideological networks in driving hate.

7. Gender Dynamics in Online Hate

Female users posted a higher proportion of hateful tweets (35.2%) than male users (27.0%), possibly reflecting the engagement of exclusionary feminist groups in anti-trans discourse. Anonymous users also contributed significantly to hateful discussions, highlighting the role of anonymity in enabling online hate while also demonstrating that publicly identifying as female or male does not necessarily correlate with reduced hate.



Introduction

Aims

This report examines the wave of anti-transgender and anti-drag mobilisation that took place in Australia during the first half of 2023 by analysing online language in relation to 20 offline events, including protests and counter-protests, some of which escalated into violence. By studying social media language and its relationship with offline events linked to the anti-trans and anti-drag movement, we aim to identify the movement's characteristics and explore whether online signals can help predict future surges in its activity.

By analysing online language, this report explores why and how this anomalous wave of mobilisation may have emerged in Australia during this period. Hate movements are typically triggered by specific events, such as terrorist attacks or military conflicts. However, no clear catalyst was present in this case. In the conclusion section, we propose data-driven hypotheses to explain its formation.

More broadly, examining the intersection between online hate and offline mobilisation provides a deeper, more nuanced understanding of social and political tensions in Australia. As this movement demonstrates, such tensions can escalate into intimidation and violence, highlighting the need to monitor these dynamics to better anticipate and respond to emerging threats.

Context

During the first half of 2023, Australia experienced a pronounced wave of anti-trans and anti-drag activism, particularly targeting transgender individuals and drag queens. This period was characterised by a series of events that intensified the debate over queer rights. A notable example was the Let Women Speak tour, spearheaded by British activist Posie Parker. The tour began in Sydney on 11 March and concluded in Canberra on 23 March. Although these events were presented as platforms for free speech, they primarily served as catalysts for anti-transgender sentiment, highlighting the escalating tensions surrounding this issue.

In Melbourne on 18 March, the tour's impact was starkly visible. Outside Parliament House, neo-Nazis clad in black marched, displaying Nazi salutes and banners with slogans such as "Destroy Paedophilia", targeting LGBTQ+ communities under a guise of protecting children. Parker distanced herself from these extremists, yet her rhetoric appeared to provide a platform for their views.

During this period, Victorian MP Moira Deeming gained significant public attention due to her participation in the Melbourne event, which sparked significant controversy and led to calls for her expulsion from the Liberal Party. Her attendance highlighted the polarising nature of the discourse on gender and transgender issues within political circles and the broader community. This incident not only affected Deeming's political standing but also intensified the discussion on the influence of far-right elements in public demonstrations concerning gender rights in Australia.

Concurrently, community events were targeted with threats and protests. Drag storytime events, like the one at Manly Library in Sydney on 25 February, became targets of threats and protests. Despite a bomb threat and protesters accusing performers of grooming—a baseless claim mirroring American conspiracy theories—the event proceeded, supported by library staff, police, and counter-protesters. Similarly, in Wollongong on 24 June, an event drew hateful groups, including neo-Nazis, but was protected by a robust counter-demonstration of supporters of the queer community.

Significant events were also targeted by anti-transgender activists, including TasPRIDE in Launceston on 15 February. On 21 March, a pre-election community meeting tied to WorldPride in Sydney turned violent when a group, allegedly part of the Christian Lives Matter movement, attacked LGBTQ+ activists.

Significance

Despite being an unprecedented wave of mobilisation against gender-diverse communities in Australia, this movement remains largely understudied, with few exceptions (e.g., Squirrell & Davey, 2023; Thomas, 2024). While it did not escalate into widespread violence, it posed a serious threat to democratic institutions by pressuring them into restrictive actions. Local councils, for instance, cancelled events like Goulburn Library's Rainbow Storytime and Melbourne's Drag Queen Story Hour due to safety concerns. This highlights how fringe and extremist groups, even without large-scale violence, can undermine institutional resilience and social cohesion through intimidation and coercion.

Data and methods

In this report, we analyse tweets related to specific offline events that took place in Australia during the first half of 2023. The data was sourced from Brandwatch, comprising posts from multiple platforms: X (79,585), news sites (5,767), Reddit (487), Tumblr (130), blogs (38), forums (56), and other sources (29). We initially conducted broad searches for discussions of the events listed in Table 1 without restricting results by geolocation. We then restricted the searches to data identified by Brandwatch as originating from Australia using general anti-LGBTIQ+ hashtags (e.g., #lgbtmafia, #transwomenareconmen) as well as hashtags associated with LGBTIQ+ topics (e.g., #lgbtqpride). This means that all content in our dataset is either explicitly about Australian events, originates from Australia, or both.



The final dataset retrieved through these methods included 86,093 contents, of which 74,567 were identified as originating from Australia. Within this subset, 70,733 tweets were posted by Australian X users. As detailed in the following sections, some analyses examine the entire dataset, while others focus specifically on tweets from Australian X users, depending on the research question and analytical approach.

Table 1. Timeline of events included in this report, detailing dates and locations.

Date	Location	Event Description
15/02/2023	Launceston	TasPRIDE event proceeded with about 30 counter-protesters drowning out approximately 4 protesters.
25/02/2023	Sydney	Drag Storytime at Manly Library took place despite a bomb threat; the event featured substantial counter-protester support.
06/03/2023	Goulburn	Rainbow Storytime cancelled due to threats; relocated and hosted by a nearby council after significant support and security assurances.
11/03/2023	Sydney	First “Let Women Speak” event.
12/03/2023	Brisbane	“Let Women Speak” event.
14/03/2023	Perth	“Let Women Speak” event.
16/03/2023	Adelaide	“Let Women Speak” event.
18/03/2023	Melbourne	“Let Women Speak” event, notable for the presence of neo-Nazis, which drew significant media and public attention. Violence erupted between protesters, counter protesters, and police.
21/03/2023	Hobart	“Let Women Speak” event.
21/03/2023	Belfield, Sydney	Community meeting escalated into violence when a group attacked LGBTQ+ protesters; the altercation involved supporters of Christian Lives Matter.
23/03/2023	Canberra	Final “Let Women Speak” event.
25/03/2023	Mount Gambier	Drag Storytime Hour proceeded under heightened security measures after library staff received death threats.
25/03/2023	Perth	Despite opposition and anti-drag mobilisation, the Drag Storytime at Perth Library took place.

Date	Location	Event Description
05/04/2023	Narre Warren	Children's drag workshop cancelled due to threats and disruptive behaviour during a council meeting on 26/04.
09/04/2023	Chelsea	Easter event Colour me 'Egg cited' occurred amid heightened anti-LGBTQ+ activism, proceeding without reported disruptions.
27/04/2023	Newcastle	Drag queen story hour faced threats but proceeded.
17/05/2023	Eltham	Pop-up Drag Queen Storytime attracted both protesters and a larger number of supporters. The event proceeded.
18/05/2023	Oakleigh	Drag Queen Story Hour planned for International Day Against Homophobia was cancelled due to threats and safety concerns.
17/06/2023	Perth	Drag Queen Storytime at Maylands Library saw protests and counter-protests; however, the event was successfully held.
24/06/2023	Wollongong	Drag queen story hour in Thirroul faced neo-Nazi protesters and substantial counter-protest, including police presence to ensure it proceeded without major disruptions.

The analyses in this report rely on two types of data. The first is metadata provided by Brandwatch, which includes the geographical origin of content, the platform where it was sourced, and the gender of the user posting the content where available. The second is the actual text of tweets, which we examined from multiple angles, including URLs, handles, hashtags, and the tweet's content. The text was then used to categorise tweets into discrete categories, as explained in next section.

Approaches to automated classification of hate targeting diverse genders

The most important classification that we used in this report distinguishes between content expressing hate toward diverse genders and content that does not, as perceived by the Australian transgender and nonbinary community. To create this variable, we worked with three community annotators—a transgender man, a transgender woman, and a nonbinary person—who were supported by two research assistants.

The annotation process followed a structured approach. It began with a training session that introduced key academic definitions of relevant concepts (e.g., hate speech, dehumanising language, and vilification). During the training session, the team reviewed examples of online language to establish a consensus on how to apply the definitions with consistent criteria.

The definition of hate speech used during the annotations training is a widely accepted one: *"Hate speech is language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics such as physical appearance, religion, descent, national or ethnic origin, sexual orientation, gender identity, or other, and it can occur with different linguistic styles, even in subtle forms or when humour is used"* (Fortuna & Nunes, 2018, p. X). This definition originates from one of the most cited papers in the field, with over 1,300 citations at the time of writing (Fortuna & Nunes, 2018).

In addition, we considered key definitions from a systematic review of hate speech definitions across ten countries (Vergani et al., 2024). This work enabled us to synthesise core characteristics of hate speech, which include incitement to violence, direct attacks based on group identity, and additional elements such as public defamatory statements, intolerance, subtle or humorous forms, and implicit or explicit manifestations.

Importantly, annotators were instructed to assess only what is explicitly manifested in the text, without inferring intent, motivation, or harm. All decisions were made strictly based on linguistic expression.

Next, we conducted inter-rater reliability training. Each week, annotators independently coded 100 pieces of content, followed by meetings to resolve disagreements. A shared document was created to document key decisions for consistency in future annotations. This process continued until the team achieved 85% agreement in the seventh round.

Following this, annotators coded 6,775 pieces of content, of which 4,581 were classified as not containing hate targeting diverse genders and 2,194 as containing hate. This dataset was deemed sufficient for developing a machine learning classifier.

The following table contains examples of what was coded as hate targeting diverse genders.

Table 2. Examples of language that meets the threshold for hate targeting diverse genders.

Fuck queers faggotstranys bicuriousand dick suckers Kill them all	Examples of blatant hateful language targeting queer people
let me guess a fat ugly fugly dyke and a less fucking ugly man in drag pervert	
Men in girls toilets are OK mutilating children is OK Grooming & pedophilia is OK 17-18yo selling their naked bodies online us OK GOD FORBID girls having some modesty! #LeaveTheKidsAlone	Examples of content expressing hate in more subtle ways, for example by attributing criminal behaviour (e.g. mutilating children, paedophilia) to transgender people, by suggesting all trans people are 'mentally ill', etc.
The contagious insanity is both infuriating & sickening. Matt Walsh's documentary is a must watch. #WhatsAWoman #WomensRightsAreHumanRights	

Our classifier used a transformer-based language representation model that processes text in a case-insensitive manner for classification. The model's performance was evaluated using the following metrics:

1. Accuracy: 0.899

The model correctly classified 89.9% of all examples in the test set, indicating strong overall performance.

2. Precision: 0.849

Of all instances the model predicted as positive (hateful), 84.9% were actually positive. This high precision indicates the model effectively avoids false positives.

3. Recall: 0.814

The model correctly identified 81.4% of actual positive cases in the dataset, demonstrating strong recall in capturing hateful content.

4. F1-Score: 0.831

The F1-score, a harmonic mean of precision and recall, balances the trade-off between the two metrics. At 83.1%, it indicates that the model effectively identifies hateful content while minimising both false positives and false negatives. This is particularly important for imbalanced datasets.

Finally, we tested the model on an additional 400 unseen data points annotated by our community annotators. The model correctly identified hate toward diverse genders 93% of the time. These results provide confidence in using this classifier to detect content expressing hate toward diverse genders within the dataset.

In addition to categorising tweets as hateful or non-hateful toward diverse genders, we classified all tweets into the categories described in the following table. One content can be categorised with multiple labels because it can contain many messages.

Table 3. Categories used for tweet classification with examples of tweets falling into each category.

Category Name	Label	Description	Example (de-identified)
Anti-Politics	ANTIPOLITICS	Attacks on Australian (e.g. Dan Andrews, Albanese) or international (e.g., Biden, Starmer) political figures, including councillors, premiers, and MPs.	Never forget what #DictatorDan and the rest of the gov did. All of that for a mRNA therapy that never stopped transmission #MedicalTyranny #healthgestapo (Labels: ANTIPOLITICS, NOVAX)
Anti-Vax	NOVAX	Opposition to vaccines or COVID policies, such as lockdowns.	Less than 7% of under 5 years old have been injected with the MRNA technology. It is imperative that we say NO to any further experiments on our children! #Genocide #crimesagainsthumanity #bioweapon #protectthekids (Labels: NOVAX, CONSPIRACY)
Conspiracy Theories	CONSPIRACY	Promotion of conspiracy theories, including those about 5G, the New World Order (NWO), 15-minute cities, QAnon, or COVID.	#TransTerrorism is happening, but it's a psyop creating collective trauma like all the rest, exploding in all our western countries in exactly the same week. Divide and conquer working brilliantly. (Labels: CONSPIRACY, ANTITRANS)
Anti-Drag	ANTIDRAG	Attacks on drag queens, often framing them as groomers, paedophiles, or sexual abusers.	Watch the huge drag queen punch an elderly man. Watch the men try to get at KJK. It chills me to think of what would have happened if they succeeded. #LetWomenSpeak #WarOnWomen #IStandWithPosieParker (labels: ANTIDRAG)
Anti-Trans	ANTITRANS	Criticism of transgender individuals, often portraying them as groomers, paedophiles, or societal threats.	The tech sector wants to take over human reproduction. It makes sense now why young girls are the worst affected and targeted by transideology. They want to use our wombs in male pregnancy experiments. #TransTerrorism #LetWomenSpeak #StandingForXX (Labels: ANTITRANS, CONSPIRACY)
Parents' Rights	PARENTSRIGHTS	Advocacy for parental control over children's exposure to specific materials, lifestyles, ideologies, or beliefs.	Can we go back to the times when grooming children and paedophilia was a crime? This is what our world has come to, clear proof that some people do not deserve to be parents. How tf do you expose your child to this traumatic scene?? #TransWomenAreConMen #podcastandchillwithmacg (Labels: PARENTSRIGHTS, ANTITRANS)
Pro-Christian	PROCHRISTIAN	Promotion of Christian religious values and support for integrating Christianity with civic life in Australia, the US, or the UK.	Transgender has reached the bottom of human depravity. The next step downward is eternal judgment & the Lake of Fire. When a society of people promote & support this dark path, that society has unleashed darkness upon itself. #transgenderSicko #Devil #pedophilia #childabuse (Labels: PROCHRISTIAN, ANTITRANS)

Category Name	Label	Description	Example (de-identified)
Mobilisation	MOBILISATION	Calls to action, such as protests or confrontations with ideological opponents.	It's time for a massive rally. If all the women in Australia predominantly in service industries like nurses, age care, hospitality, teachers, this country would ground to a halt. Just one day, and then we might finally be seen and heard. #equity #WomensRightsAreHumanRights (Labels: MOBILISATION, EXCLFEMINISM)
Exclusionary Feminism	EXCLFEMINISM	Advocacy for women's rights in opposition to transgender rights.	No decent adult male would violate the single-sex rights of females. Those that do are misogynists who (like most TRAs and LGBT+ groups) want male rapists who self-ID as women placed in women's prisons. #KeepPrisonsSingleSex #SexNotGender #SaveFemaleSports (Labels: EXCLFEMINISM, ANTITRANS)
Anti-Trans LGB	ANTITRANSLGB	Criticism of transgender identities while supporting gay, lesbian, and bisexual identities (e.g., using hashtags like #lgbwithoutthet).	A huge number of gay people are starting to speak out about the ridiculousness of the world we find ourselves in these days. Read up on some of the threads under this hashtag: #LGBWithoutTheT or #LGBwithoutT (Labels: ANTITRANSLGB)

To perform this task, we automated the classification of tweets using OpenAI's GPT-4o model. Specifically, we developed a classification function that sent each tweet to the AI model, instructing it to categorise the content based on these themes. The AI processed each tweet, identifying relevant themes based on the text and associated hashtags, and assigned one or more category labels accordingly. This automated approach enabled the systematic and scalable identification of key narratives within the dataset, allowing for a detailed examination of the relationship between online discourse and broader societal debates. A small random subsample of tweets was cross-checked by the team to ensure correct classification, and confirmed that the model classified tweets correctly 96% of the times.



Linking Online Activity To Offline Events

In this section, we analyse the full dataset (n = 86,093) to explore the relationships between online language and offline events.

Descriptive Analysis

The next figure presents the total volume of content across the dataset, with 20 offline events marked as vertical lines, coloured according to event type. Blue lines indicate events where a counter-protest occurred, orange lines represent threats of violence against drag storytime events, red lines mark protests that resulted in violent clashes with counter-protesters and police, and white lines denote peaceful demonstrations. The accompanying bar chart, which displays daily tweet volumes, is coloured according to the categories assigned during classification.

The figure illustrates the complexity of the dataset, given the number of events and categories analysed. However, at first glance, the most prominent coloured segments highlight the dominant conversations within the dataset. The bottom blue segment represents hate toward

diverse genders, while the orange segment above it captures tweets attacking Australian politicians. The top yellow segment identifies discussions promoting exclusionary feminist views. The data clearly illustrate a surge in both online and offline mobilisation against transgender and gender-diverse identities around March 2023. This period saw the highest volume of online activity and the most intense offline events, suggesting a reciprocal relationship between digital discourse and real-world mobilisation. Notably, this peak coincided with Posie Parker's tour.

Importantly, the two events marked by violent clashes occurred at the peak of both online and offline hate, indicating a significant escalation in mobilisation.

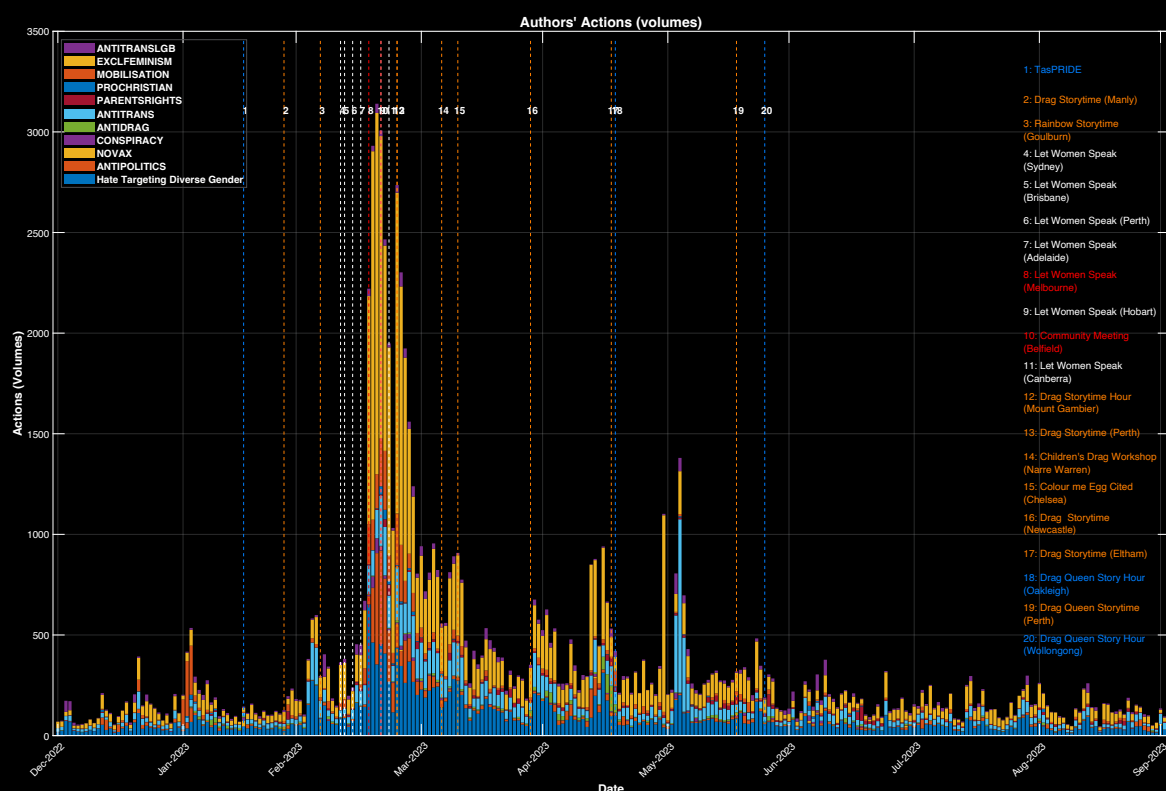


Figure 1. Daily volume of tweets over time, with different categories of language represented by different colours. Vertical dashed lines indicate events, colour-coded by event type.

User Networks and Far-Right Influence

We conducted an analysis of user and network characteristics. We began by categorising users based on their association with far-right fringe groups (e.g., Antipodean Empire, Australian Nationalist Socialists, PEGIDA) and more mainstream far-right organisations (e.g., Reclaim Australia, Firearms Lovers Australia, Australian MAGA group). This classification included both individuals known to be affiliated with these groups and official pages associated with them.

In total, we identified 37 accounts in the dataset, which collectively generated 512 pieces of content. We then constructed a network of all users ($n=14,684$), defining connections based on interactions such as tagging, retweeting, or responding to tweets ($n=25,361$). To understand the influence of far-right actors within this network, we calculated measures of centrality, particularly outdegree centrality—a measure of how actively a node (user) disseminates information or connects with others.

Some far-right users hold disproportionately high centrality in discussions. The most central ranks 12th overall by betweenness centrality, with others following at 25th, 55th, 65th, and 74th positions.

Our findings indicate a peak in the outdegree centrality of far-right users coinciding with the surge in online activity around the Let Women Speak event in Melbourne on 18 March. This event, which involved the presence of neo-Nazis and confrontations between protesters, counterprotesters, and police, marked the height of anti-trans and anti-drag mobilisation. The elevated outdegree centrality of far-right actors at this time suggests that they played a key role in spreading information about the event online, both before and after the event.

Additionally, we observed smaller peaks in far-right outdegree centrality corresponding to events in Goulburn, Newcastle, Eltham, Oakleigh, and Perth. These findings underscore the Australian far-right's role in amplifying and coordinating online discourse around anti-transgender and anti-drag mobilisation, particularly in relation to certain key events.

Betweenness Centrality: Measures the extent to which a node lies on the shortest paths between other nodes in a network, indicating its potential role in controlling or facilitating information flow.

Outdegree Centrality: Refers to the number of outgoing edges from a node, representing its activity or potential influence in disseminating information within a directed network.

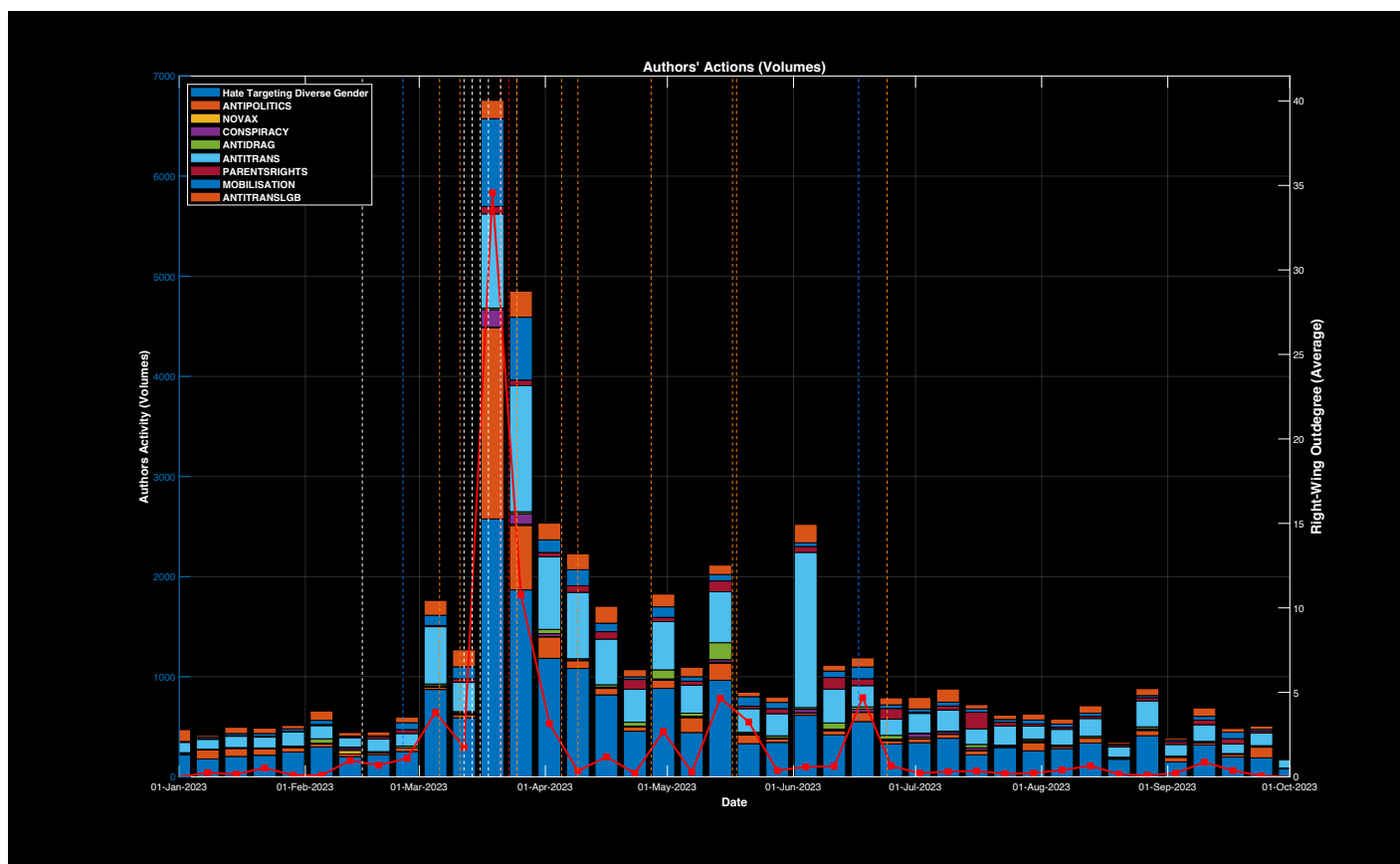


Figure 2. Weekly volume of tweets over time. The red line represents the outdegree of far-right extremists in the dataset.



Online Activity Patterns Before Violent and Non-Violent Events

We analysed the presence of discussions classified into specific categories in the lead-up to offline events to determine whether certain types of language or narratives could indicate the risk of such events occurring. The only category that stood out in our analysis was content labelled as ‘Anti-politics’—that is, posts containing attacks against Australian or international politicians.

In the next figure, we present a boxplot comparing the volume of anti-politics content in the days leading up to the 20 offline events. Each boxplot represents the distribution of activity levels, with the interquartile range (the middle 50% of data) shown as the main box, the horizontal line indicating the median, and the dot representing the mean.

To distinguish between events with and without violence or threats of violence, we use two colours:

- **Red boxplots** represent the average level of anti-politics content before events involving violence or threats of violence.
- **Green boxplots** represent the average level of anti-politics content before events that remained peaceful.

For example, the red boxplot in figure 3 labelled “3” shows the distribution of anti-politics content in the three days preceding events characterised by violence. Comparing the green and red distributions, we observe a notable shift: the entire interquartile range of the red boxplots is elevated, not just the mean.

This pattern suggests a structural difference in online discourse before violent versus non-violent events. Specifically, events involving violence tend to be preceded by a broader and more sustained wave of content attacking politicians, with consistently higher activity levels across multiple users. In contrast, peaceful events may experience occasional spikes (tail events), but their overall distribution remains flatter, indicating less sustained engagement with this narrative.

These findings highlight a strong association between heightened online anti-politics discourse and the likelihood of violence or threats of violence at offline events. This pattern may indicate a potential early warning signal for increased violent mobilisation, warranting further investigation in future research.

Interestingly, as the following figure shows, the average levels of hate targeting diverse genders (as captured by our community-led classifier) are very similar across both violent and non-violent events.

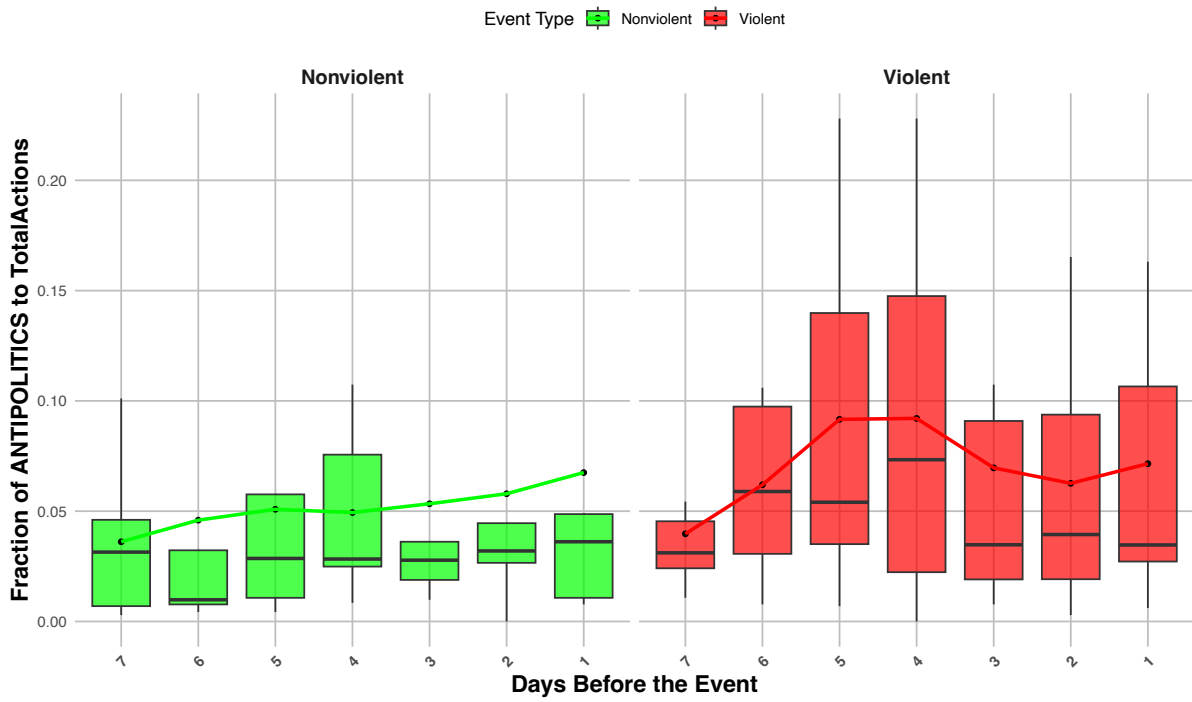


Figure 3. Boxplot comparing the average volume of anti-politics content in the days leading up to the 20 offline events.

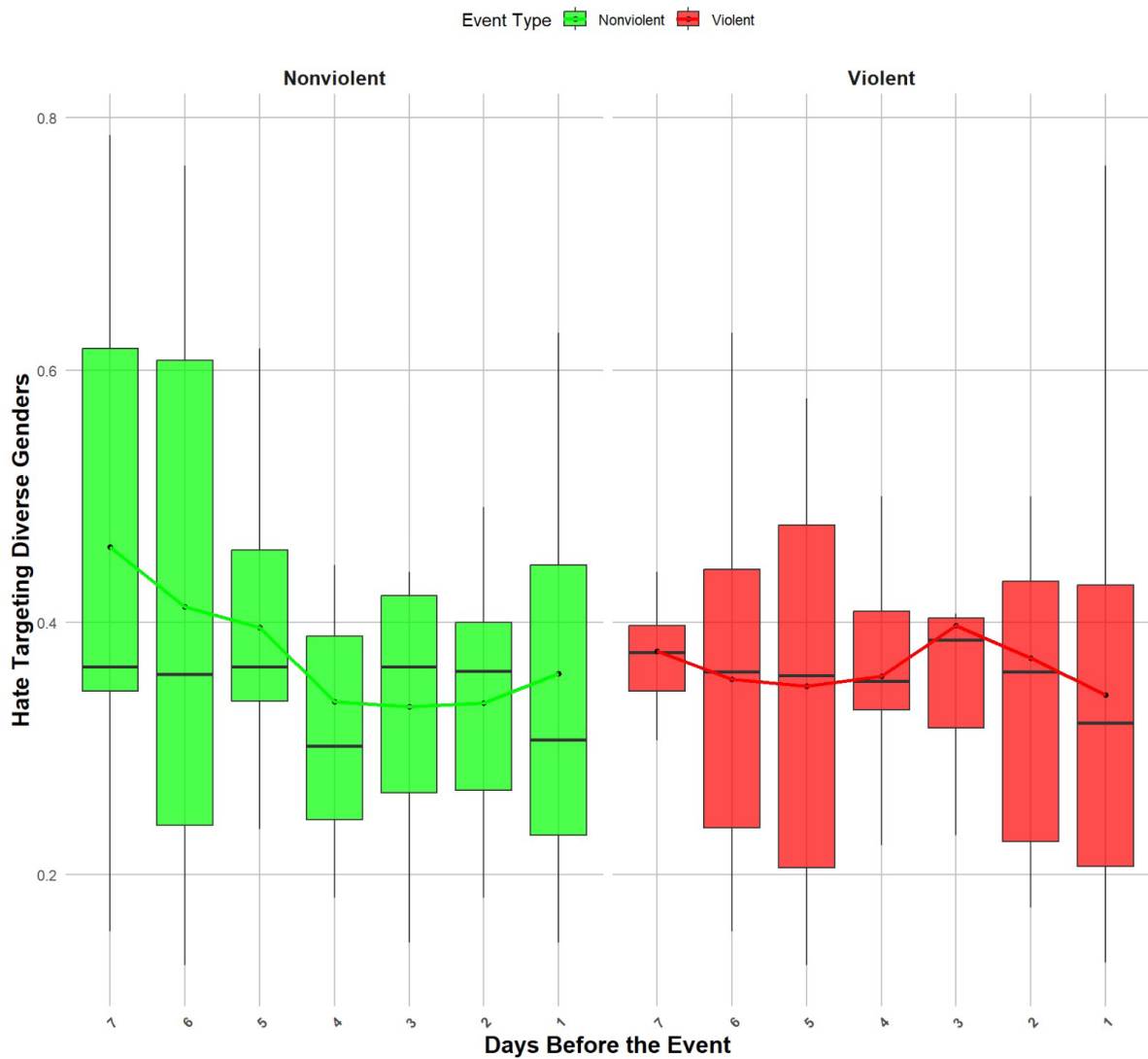


Figure 4. Boxplot comparing the average volume of hate speech targeting diverse genders in the days leading up to the 20 offline events.

Event Dynamics and the Persistence of Online Narratives After the Events

We conducted an event study analysis using a two-week window—one week before and one week after each event. For each event, we distinguished between two types of online activity: (1) actions explicitly discussing the event and (2) actions where the event was not directly mentioned. We then calculated the volume of activity for each of these categories across all events and datasets, averaging the results and computing standard deviations.

This analysis reveals several key patterns. Notably, while discussions explicitly referencing an event (red series) tend to decline rapidly after the event, related topics can remain popular in broader

discourse. Additionally, in some cases, a topic first emerges within event-related discussions before appearing in conversations that do not directly mention the event, suggesting a diffusion of narratives beyond the immediate context of the event itself.

For example, figure 5 shows the average volume of content expressing hate toward diverse genders before and after the events. While there is a small spike in hateful discussions that explicitly mention the event—which quickly declines to zero—the overall level of hate remains high and continues to increase over time, even after the events.

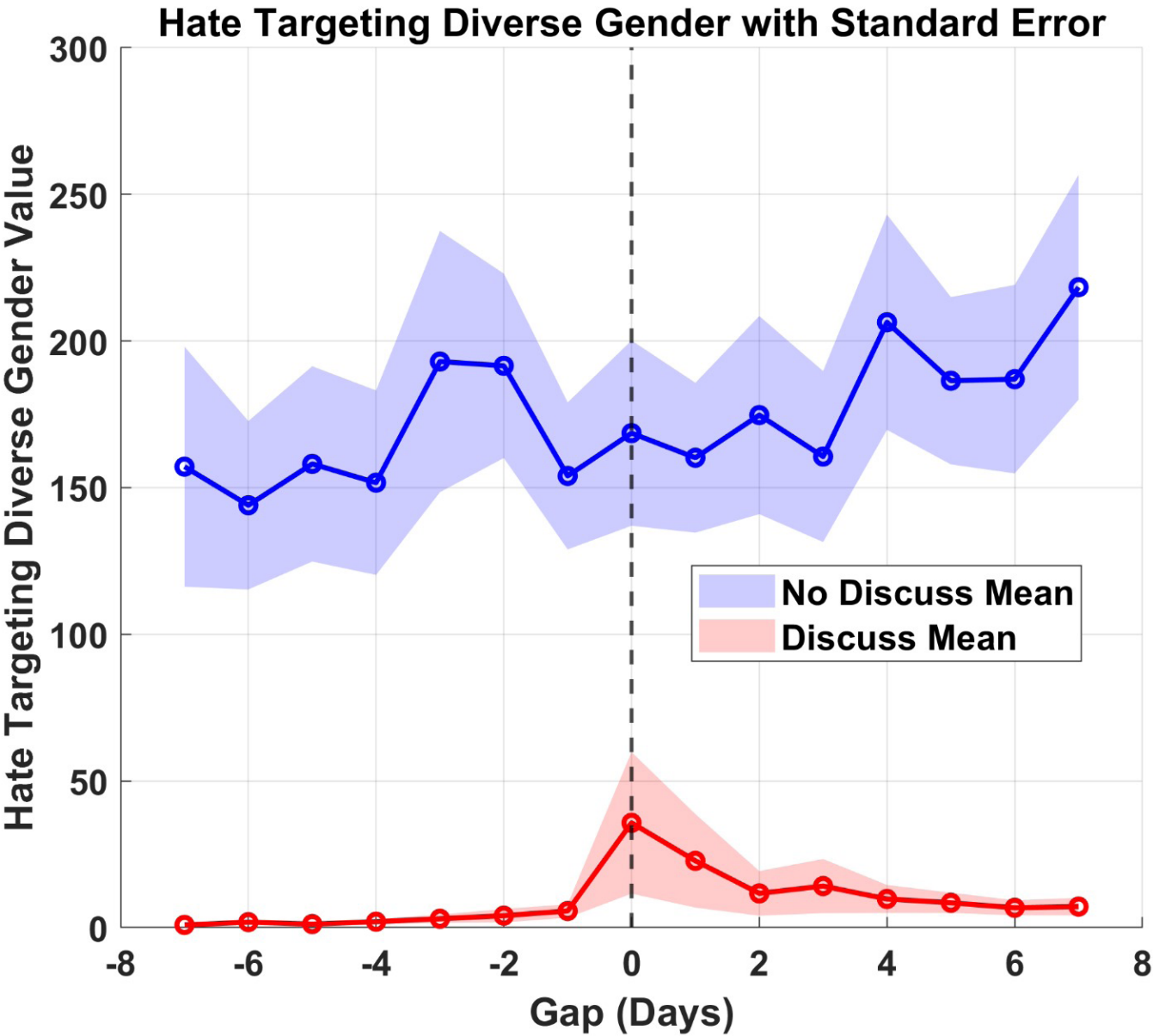


Figure 5. Average volume of hate speech targeting diverse genders before and after the events.

Figure 5 stands out from the others we analysed, as it shows that anti-drag hate tends to spike, on average, before both violent and non-violent events. This pattern may indicate a potential early warning signal for increased anti-trans mobilisation, warranting further investigation in future research.

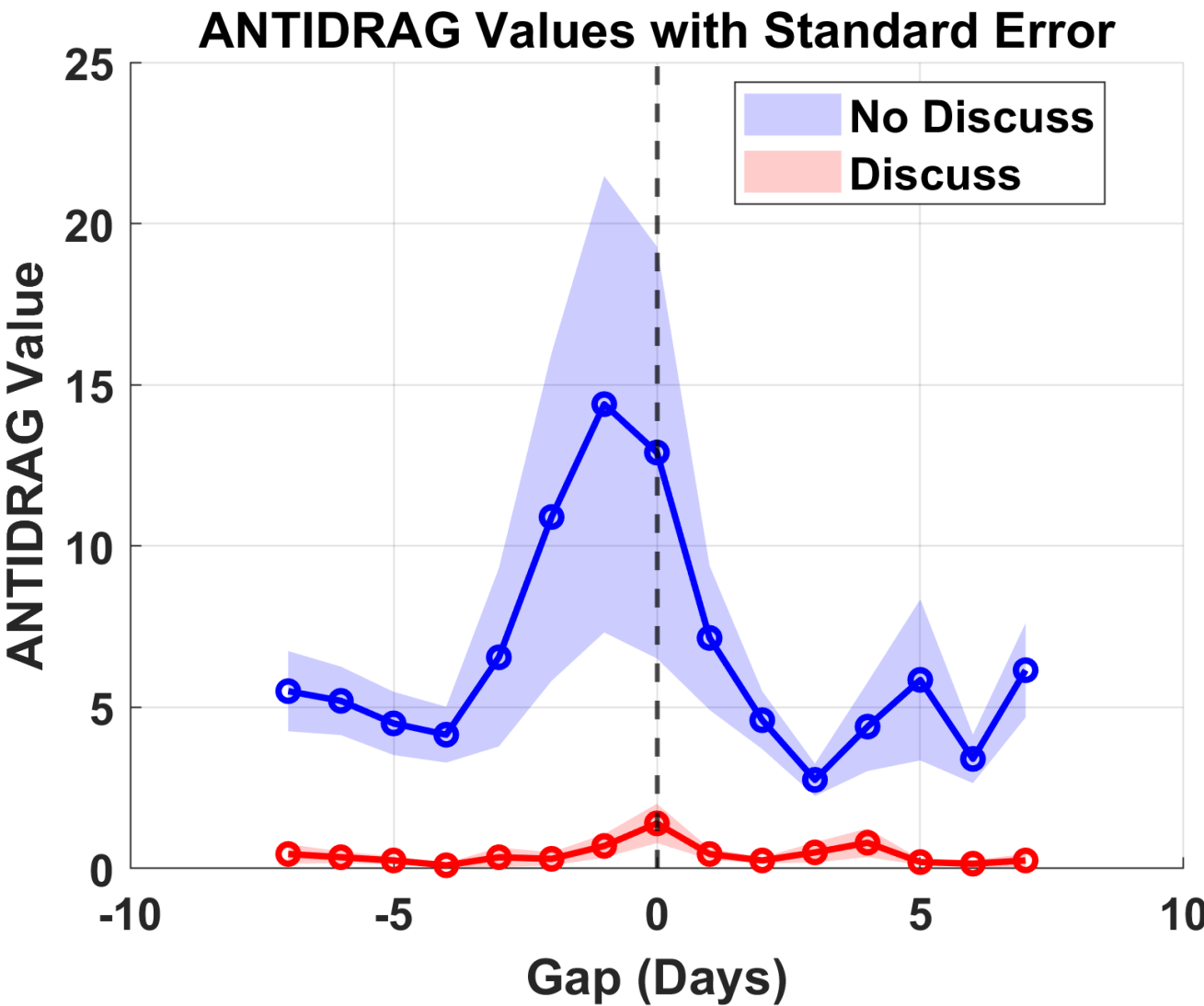


Figure 6. Average volume of anti-drag content before and after the events.

Figure 6 illustrates how different event characteristics influence the volume of anti-politics content over time. Specifically, it examines whether online hostility toward politicians changes after an event, whether this change differs for events involving violence, and whether the discussion explicitly mentions the event.

The analysis reveals that violent events tend to be preceded and followed by a significant increase in anti-politics content that does not directly mention the event. In contrast, for discussions that explicitly reference violent events, the surge in anti-politics discourse occurs only after the event has taken place.

These findings reinforce previous results suggesting that anti-politics discourse may serve as a potential indicator of the risk of violence, highlighting a possible link between offline violence and increased hostility toward politicians in the lead-up to an event. Importantly, they also underscore the need to examine not only content that directly discusses an event but also broader discussions originating from Australia or addressing Australia-relevant issues and topics.

ANTIPOLITICS Values with Standard Error and Violence

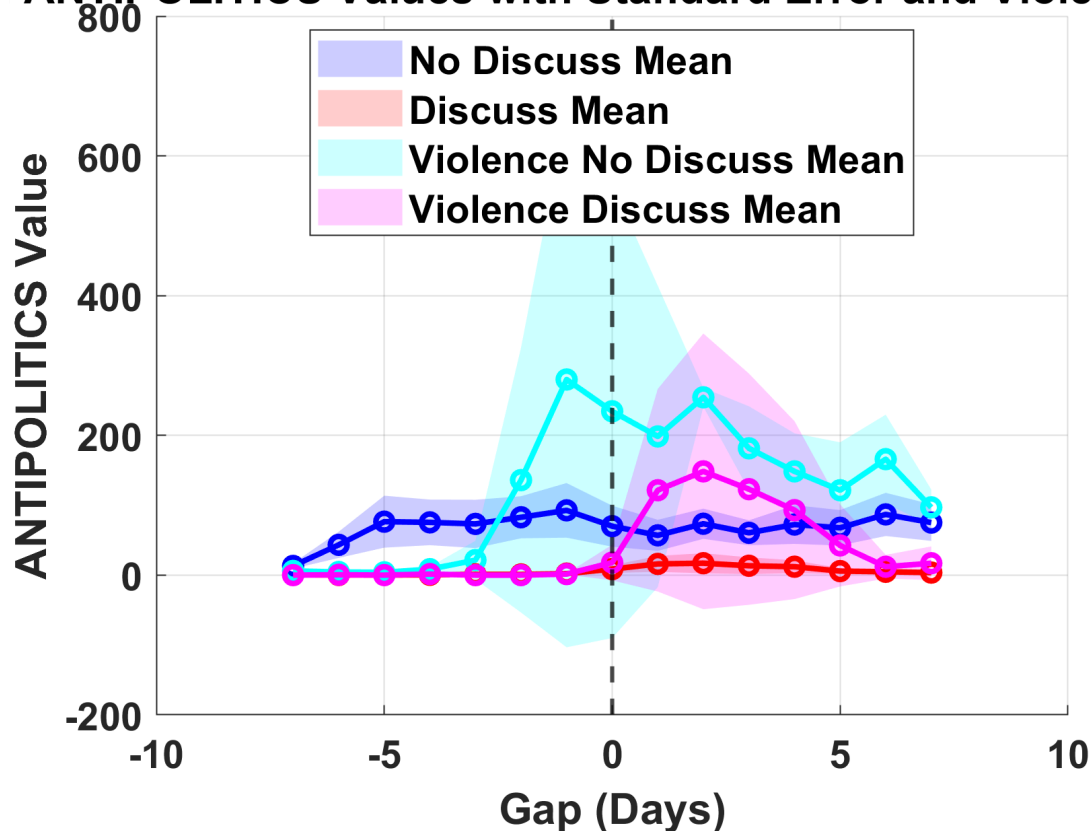


Figure 7. Average volume of anti-politics content before and after the events, separated by violent and non-violent events.

Understanding the Role of Anti-Politics Discourse in Violent Mobilisation

Our analysis found a strong association between online hostility towards politicians (anti-politics discourse) and violent offline events within the specific context of social movement mobilisation that targets diverse genders. This pattern suggests that anti-politics discourse may serve as an early warning indicator of escalating mobilisation in this context. Several hypotheses could explain this relationship:

1. Focusing Mobilisation and Justifying Action

Anti-politics discourse channels frustration by framing politicians as corrupt, oppressive, or existential threats, providing a clear adversary for mobilisation. In the case study of the anti-trans and anti-drag social movement, politicians are blamed for enacting or supporting policies perceived as harmful, making them focal points for protests and direct action. This framing can legitimise radical measures when conventional political processes are seen as ineffective, potentially encouraging violent tactics.

2. Reinforcing Grievances

Anti-politics discourse often intersects with broader ideological grievances. In anti-trans and anti-drag mobilisation, online hostility towards politicians is frequently linked to misinformation, far-right narratives, and cultural backlash against social change. As grievances compound, individuals may escalate their participation from protests to direct confrontations, perceiving institutional avenues as ineffective.

3. Eroding Institutional Trust and Encouraging Alternative Action

Widespread hostility towards politicians contributes to declining faith in democratic institutions. In movements mobilising against diverse genders, this erosion of trust can lead individuals to reject legal and democratic mechanisms. In extreme cases, it fosters support for vigilante justice or direct attacks on political and civic institutions.

It is important to emphasize that this relationship between anti-politics discourse and violence is observed in the specific context of a social movement mobilisation targeting diverse genders. This does not imply that all instances of anti-politics discourse are predictive of violence, nor should this finding be generalised across different social movement contexts. The potential for violence in response to political hostility depends on a range of factors, including the nature of the movement, its ideological drivers, and broader societal conditions. Future research should explore whether similar dynamics are present in other mobilisation contexts and under what conditions anti-politics discourse may contribute to escalation.

Insights From Australian Tweets

In this section, we provide a detailed analysis of tweets expressing hate toward diverse genders, focusing on their geographical origins, the top trending users mentioned, the web domains shared, and the gender of users posting hateful content.

Content published on X by Australian users represents the largest subset of our dataset, accounting for approximately 82% of the total. This section offers insights into the characteristics of this movement as it manifests online.

There are several reasons why we present this focus on X. Firstly, X is an interesting platform because, following Elon Musk's acquisition, many fringe users previously expelled for violating terms of service—particularly regarding misinformation and hate speech—were reinstated. This has created a unique environment where mainstream and fringe groups coexist, although no clear data is available on the impact of Musk's acquisition. However, the platform's pro-free speech policies appear to have amplified the presence of fringe political discourse.

Additionally, X is one of the few platforms where, for some users, geographical data can be identified down to the state, city, or even postcode level. Analysing Australian X users, therefore, allows for a more detailed geographical understanding of where these users—and their discourse—originate.

The following figure presents a network representation of all users included in our dataset. Connections between users are defined by interactions such as commenting on posts, tagging other users, or retweeting content. Users who have published at least one hateful tweet are highlighted in red. The visualisation reveals that the network of hate-spreading users is highly cohesive, with frequent interactions both within the group and with unrelated non-hate networks. Notably, this hate network occupies a central position in broader discussions, influencing diverse user groups and attracting engagement from non-hate communities.

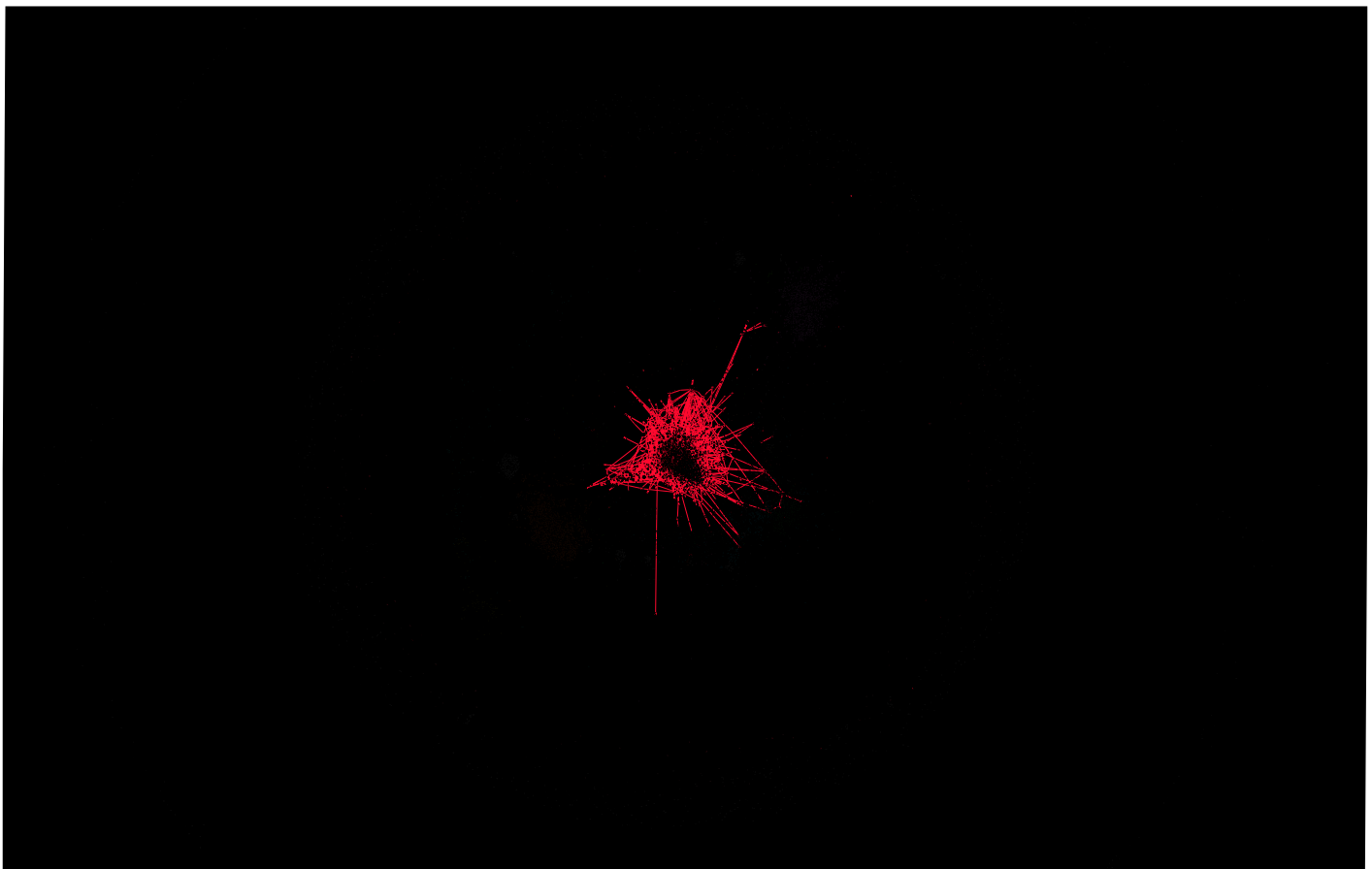


Figure 8. Network visualisation of all X users in the dataset. Connections represent all interactions, including tweets, retweets, quotes, and responses. Actions identified as hate targeting diverse genders are highlighted in red.

In the following pages, we examine the characteristics of users who have shared hateful content, focusing on their geographic origins, gender distribution, and their engagement patterns. Additionally, we analyse the most frequently mentioned users within hate-related discussions and the web domains most commonly shared in hateful content.

Pinpointing hate against diverse genders: a geographic breakdown

Our dataset included city-level location metadata for 30,701 of the total 70,733 tweets posted by Australian X users.

Considering that NSW is the most populous state in Australia, followed by Victoria, Queensland, and WA, our analysis reveals a broad association between events location, population size and the volume of hate-related tweets. The largest numbers of such tweets were recorded in NSW (11,998) and Victoria (11,309), which align with their larger populations. Queensland, WA, SA, and TAS followed, with totals of 6,834, 5,229, 1,773, and 1,076 tweets, respectively. The Northern Territory (NT), which did not host any of the events under consideration, yielded the fewest tweets, totalling 69, of which 21 were identified as hate.

Table 4 outlines the number and proportion of hateful tweets originating from each state within our dataset. Notably, despite Victoria hosting 5 of the 20 events analysed in this report (see Table 1), it exhibits a comparatively lower proportion of online hate relative to New South Wales (NSW), which hosted 6 events. Moreover, Queensland (QLD) and South Australia (SA), hosting only 1 and 2 events respectively, display significantly higher proportions of online hate compared to Victoria.

This data suggests that the occurrence of online hate is not solely dependent on the number of events hosted in a state, indicating other influencing factors at play.

Table 4. Distribution of Online hate by State

State	No Online hate	Online hate	Total	% Hate
Australian Capital Territory	278	116	394	29.44
New South Wales	7,513	4,485	11,998	37.38
Queensland	4,479	2,355	6,834	34.46
South Australia	1,088	685	1,773	38.64
Tasmania	750	326	1,076	30.30
Victoria	8,019	3,290	11,309	29.09
Western Australia	3,960	1,269	5,229	24.27

Our analysis reveals no significant differences in the proportion of hateful tweets between capital cities and regional cities, with the sole exception of Victoria. In Victoria, Melbourne exhibits a online hate proportion of 29.3%, compared to 23.4% in regional cities. In New South Wales (NSW) and Queensland (QLD), where sufficient data from regional cities exists to enable a meaningful comparison, the differences in online hate proportions between capital and regional areas are minimal (less than 1%).

Tasmania (TAS), and Western Australia (WA) due to insufficient data. Specifically, these regions either lacked sufficient tweet volumes from regional cities or did not have adequate representation of both capital and regional cities in the dataset to enable a meaningful comparison.

The tables below provide a detailed breakdown of the counts of Hateful and non-hateful tweets, geolocated at the city level. This data further supports the findings of minimal regional disparities in online hate distribution, with the notable exception of Victoria. The analysis comparing the proportion of hateful tweets between capital and regional cities was not conducted for the Australian Capital Territory (ACT), Northern Territory (NT), South Australia (SA),

Table 5. Hateful and non-hateful tweets from Australian Capital Territory (ACT)

City	Non hate	Hate
Canberra	278	116

Table 6. Hateful and non-hateful tweets from New South Wales (NSW)

City	Non hate	Hate
Albury	4	0
Armidale	3	2
Ballina	1	0
Batemans Bay	2	0
Bathurst	6	1
Central Coast	11	3
Cooma	1	0
Dubbo	2	0
Goulburn	0	1
Kiama	5	0
Maitland	2	0
Newcastle	119	117
Orange	0	1
Port Macquarie	10	0
Richmond	6	2
Singleton	4	0
Sydney	6326	3865
Wagga Wagga	2	0
Wollongong	119	50

Table 7. Hateful and non-hateful tweets from Northern Territory (NT)

City	Non hate	Hate
Alice Springs	5	1
Darwin	38	16
Katherine	5	4

Table 8. Hateful and non-hateful tweets from Queensland (QLD)

City	Non hate	Hate
Brisbane	1653	825
Bundaberg	10	2
Caboolture	1	0
Cairns	37	27
Gold Coast	441	194
Hervey Bay	11	3
Mackay	2	0
Rockhampton	30	7
Sunshine Coast	3	0
Toowoomba	15	3
Townsville	155	99

Table 9. Hateful and non-hateful tweets from South Australia (SA)

City	Non hate	Hate
Adelaide	715	225
Gawler	1	0
Mount Gambier	5	0
Murray Bridge	3	0

Table 10. Hateful and non-hateful tweets from Tasmania (TAS)

City	Non hate	Hate
Burnie	4	0
Hobart	566	122
Launceston	57	96

Table 11. Hateful and non-hateful tweets from Victoria (VIC)

City	Non hate	Hate
Ballarat	84	31
Bendigo	22	0
Cranbourne	4	0
Geelong	195	55
Melbourne	7614	3163
Melton	46	21
Mildura	8	3
Shepparton	3	0
Wangaratta	2	1

Table 12. Hateful and non-hateful tweets from Western Australia (WA)

City	Non hate	Hate
Albany	1	1
Busselton	5	0
Esperance	1	0
Geraldton	3	0
Kununurra	2	1
Mandurah	24	20
Perth	2147	804

From this data, we can infer a general observation that there tends to be a higher concentration of online hate in cities where events occurred, particularly in larger urban centres. However, the proportion of online hate does not strictly increase with the number of events hosted in a city, as seen in Sydney, Perth, and Melbourne, which have different percentages of online hate with Melbourne being significantly lower than Sydney and Perth, suggesting that other factors such as city population, social dynamics, presence and visibility of target communities (e.g. gender diverse visibility) and possibly the nature of the events themselves also influence the volume of online hate.

To further explore the relationship between the proximity of online hate to event locations and the presence of specific communities,

such as gender diverse individuals, additional research could be conducted. A common assumption might be that regions with fewer visibility of transgender individuals would exhibit higher levels of hate due to lack of familiarity and prevailing ignorance. However, the preliminary findings from our dataset suggest a contrary pattern—areas with higher proximity to events, which are often more urban and presumably more diverse, tend to show higher levels of hate. This observation leads to a hypothesis that physical closeness and perhaps frequent encounters or heightened tensions in more diverse or densely populated areas might exacerbate online expressions of hate. This pattern warrants further investigation to assess whether similar dynamics occur with other forms or targets of hate, to identify the underlying drivers of the effect, and to test the hypothesis across a broader temporal scale.

Top trending users mentioned in the tweets

This analysis is not about the authors of the tweets, but rather the handles being mentioned by the tweets contained in our dataset. They are mostly public figures, with few exceptions. The bar chart below provides a detailed breakdown of the top 40 most cited authors. Each bar represents one author, colour-coded to show the nature of the mentions: blue for non-hate and red for hate. This chart shows also the context of these mentions, giving insight into the tone of the discourse surrounding each author.

The two accounts that appear significantly more than any other account are linked with two anti-trans influencers based overseas: Posie Parker and Matt Walsh.

Kellie-Jay Nyishie Keen-Minshull, known as Posie Parker, is a British anti-trans activist who appears in 1876 contents, of which 235 (12.5%) are hate. She conducted seven anti-transgender rallies across Australian capital cities in March 2023, which significantly contributed to her being the most cited author in a dataset capturing hate targeting diverse genders in Australia during the first six months of 2023.

Matt Walsh is an American conservative commentator who appears in 1469 contents, of which 77 (5.2%) hate. He is known for his strong opposition to transgender rights. Host of The Matt Walsh Show and a prominent figure at The Daily Wire, Walsh has produced content critical of transgender healthcare, particularly for minors. His works include the documentary “What Is a Woman?” and the book “Johnny the Walrus,” which controversially compares being transgender to pretending to be a walrus.

The third most cited account is Moira Deeming, who appears 532 times, of which 72 (13.5%) hate. Moira is a Victorian politician and member of the Victorian Legislative Council for Western Metropolitan Region. Known for her anti-transgender views, Deeming actively participated in and helped organise an anti-trans rally at the Victorian Parliament in March 2023. Her involvement in this event, which saw neo-Nazi groups in attendance, led to her expulsion from the parliamentary Liberal Party. Deeming was readmitted to the parliamentary Liberal Party in December 2024.

The fourth account in order of appearance is @speakupdotell, which was suspended by X. Interestingly, all 466 tweets where the account is named are hate. We could not find any trace of this account on the internet.

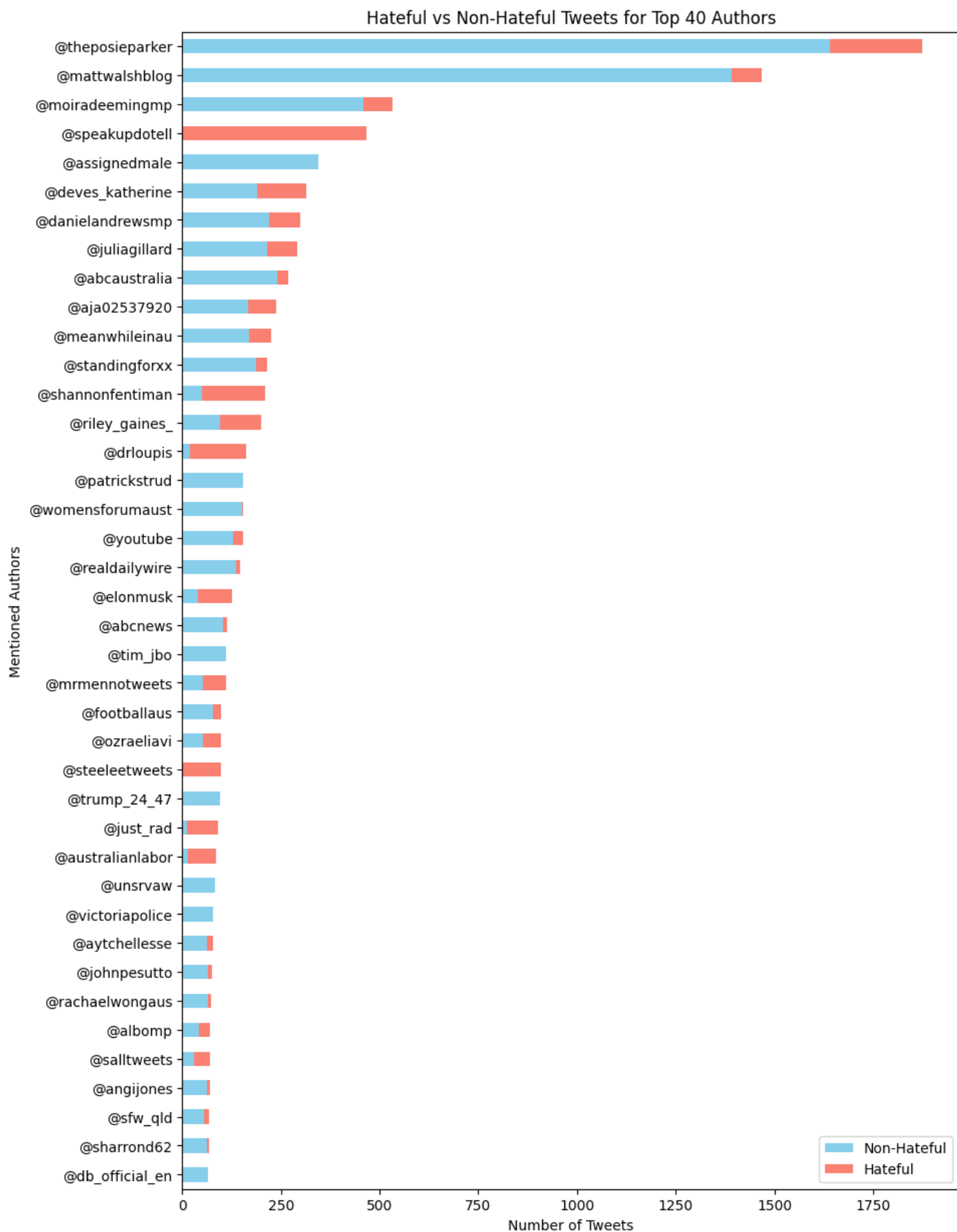


Figure 9. Top trending users mentioned in the dataset.

Domains contained in the tweets

We analysed the origins of the information shared in tweets by extracting and opening all URLs included within the dataset. After excluding X, which is the most common source of URLs, we identified that the most frequently cited domain is womensforaustralia.org, appearing in 105 tweets. This was followed by YouTube with 32 mentions, theaustralian.com.au with 22, abc.net.au with 19, telegraph.co.uk with 17, womensrights.network with 13, and familyfirstparty.org.au with 10.

This analysis indicates that apart from referencing mainstream media outlets, some of the content originates from websites associated with women’s rights networks, notably those with anti-transgender stances such as womensforaustralia.org in Australia and womensrights.network in the UK. However, it is important to note that the vast majority of tweets containing URLs link back to other tweets, suggesting that most discussions are contained within X itself, based on content shared on the platform.

Gender and online hate

The dataset includes gender metadata for the authors of 33,395 tweets, revealing patterns in tweet behaviours associated with hateful language. Female users who disclosed their gender, representing 25.41% of total tweets, post or repost a higher proportion of online hate, with 35.2% of their tweets classified as hate. In comparison, male users who disclosed their gender, who account for 21.82% of the total tweets, display a lower incidence of hateful tweets at 27.0%. Importantly, the ‘Unknown’ gender category, comprising users who opt not to disclose their gender, constituting the majority with 51.73% of all tweets, also shows a significant presence of hate speech, with 36.0% of their tweets being hate. Conversely, the ‘Not Available’ category, making up just 1.06% of the total, has the lowest rate of hate speech at 19.5%.

Table 13. Gender breakdown by percentage of hateful tweets.

Total Tweets	Total Tweets	% of Total Tweets	Hateful tweets	% Hateful tweets
Female	17,967	25.41%	6,316	35.2%
Male	15,428	21.82%	4,166	27.0%
Unknown	36,590	51.73%	13,162	36.0%
Not Available	748	1.06%	146	19.5%

Several factors might contribute to the observed phenomenon where females exhibit a higher rate of hate speech compared to males in the dataset. This trend could be influenced by networks of feminist women groups that are anti-transgender, such as womensforaustralia.org or womensrights.network, epitomised by movements like that led by Posie Parker, which command significant online followings and influence. Contentious issues such as the perceived threats posed by transgender women and girls to women in sports, toilets, or prisons could further drive women’s engagement in hate discourse. These discussions often serve as flashpoints that escalate into expressions of hate speech.

This data enables a discussion on the relationship between anonymity and the propensity to engage in online hate. Notably, the ‘Unknown’ gender category accounts for over half of the total tweets and exhibits a hate speech rate of 36.0%. Users identified as female or male, whose gender is known, suggesting a lesser concern for online anonymity, also show significant levels of online hate, with females displaying a hate speech rate of 35.2%, very close to that of the ‘Unknown’ category. This suggests that disclosing key personal characteristics, such as gender, in social media profiles does not necessarily correlate with lower levels of online hate.

Conclusion

This report illustrates the value of computational research in examining the link between online hate and offline mobilisation. Our findings demonstrate how this approach can inform the development of risk indicators—such as anti-politics discourse—for early warning systems and support long-term monitoring strategies to track the persistence of hate beyond specific events.

This research shows how the anti-trans and anti-drag mobilisation in the first half of 2023 in Australia was highly diverse, reflecting characteristics of contemporary extremist and fringe movements globally (Meleagrou-Hitchens & Ayad, 2023). The anti-trans and anti-drag movement drew participants from a broad spectrum of ideological backgrounds, including exclusionary feminists, conspiracy theorists, and far-right actors. Many of these groups had prior roots in the anti-lockdown protests during the COVID-19 pandemic, demonstrating how movement structures, grievances from past movements can be repurposed and adapted to new issues.

Despite their differing motivations, these disparate actors coalesced under a broader umbrella movement focused on opposition to trans and drag representation. However, the movement was not driven by this single issue alone; rather, it served as a conduit for a variety of grievances, ranging from distrust in government institutions to broader conspiracy-laden narratives. The most potent and mobilising force within this ecosystem was anti-politics discourse, associated with distrust in democracy and mainstream institutions—an element commonly exploited by known extremist groups to fuel mobilisation and radicalisation.

Understanding this convergence of ideological currents is crucial for policymakers, as it highlights how seemingly distinct movements can coalesce around shared themes of grievance and distrust. By recognising these patterns, strategies can be developed to counter the grievances that drive mobilisation and prevent extremist groups from leveraging these movements for broader destabilising goals.

Based on the data collected in this research, several hypotheses can be proposed to explain the rise of anti-drag and anti-LGBTQ+ mobilisation in Australia at this particular time. One key factor is the influence of international narratives, including online US-based narratives and the presence of Posie Parker in Australia, with Australian activists adopting talking points and strategies from international activists. Importantly, similar movements emerged in US, UK and France in the same period (Squirrell & Davey, 2023). The Australian movement has also been shaped by the post-pandemic landscape, where former anti-lockdown and anti-vaccine activists have pivoted to anti-LGBTQ+ activism to retain their audiences and mobilisation potential. A coalition of fringe groups—including

Christian fundamentalists, conspiracy theorists, and far-right extremists—has coalesced around this issue, exploiting social divisions. Neo-Nazi groups, though numerically small, have used anti-drag protests to generate media attention. Our data shows that they played a key role in disseminating information in the network, especially around specific events.

Further research is needed to refine our understanding of the relationship between online hate and offline harm. While we have identified clear correlations, additional work is required to establish consistent risk indicators across different forms of hate and case studies. For instance, our research suggests that anti-drag hate may be a precursor to offline violence, but more in-depth analysis is necessary to confirm this pattern. Similarly, while online hate persists beyond major events, future research could explore the long-term social and psychological impacts of this continued hostility on gender-diverse communities.

Geographical analysis indicates that online hate does not simply correlate with the number of offline events but is shaped by complex social and demographic factors. Understanding these dynamics further could help in designing more targeted interventions at the local level. Our study raises important questions about the role of anonymity in online hate speech, particularly given that disclosed female users posted a higher proportion of hateful tweets than their male counterparts, potentially reflecting the influence of exclusionary feminist networks in driving hate targeting diverse genders.

By building on these findings, future research can help refine predictive models and develop more effective strategies for mitigating online hate before it escalates into offline violence. In the meantime, the insights from this report provide a crucial foundation for policymakers, law enforcement, and civil society organisations working to counter online and offline threats against gender-diverse individuals in Australia.

References

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, 51(4), 1-30.

Meleagrou-Hitchens, A., & Ayad, M. (2023). The age of incoherence? Understanding mixed and unclear ideology extremism. Program on Extremism, George Washington University, & National Counterterrorism Innovation, Technology, and Education Center. <https://digitalcommons.unomaha.edu/ncitereportsresearch/41>

Squirrell, T., & Davey, J. (2023). A year of hate: Understanding threats and harassment targeting drag shows and the LGBTQ+ community. Institute for Strategic Dialogue (ISD). <https://www.isdglobal.org/isd-publications/a-year-of-hate-understanding-threats-and-harassment-targeting-drag-shows-and-the-lgbtq-community/>

Thomas, E. (2024). A year of hate: Anti-drag mobilisation efforts targeting LGBTQ+ people in Australia. Institute for Strategic Dialogue (ISD). <https://www.isdglobal.org/isd-publications/a-year-of-hate-anti-drag-mobilisation-efforts-targeting-lgbtq-people-in-australia/>

Vergani, M., Perry, B., Freilich, J., Chermak, S., Scrivens, R., Link, R., Kleinsman, D., Betts, J., & Iqbal, M. (2024). Mapping the scientific knowledge and approaches to defining and measuring hate crime, hate speech, and hate incidents: A systematic review. *Campbell Systematic Reviews*, 20(2), e1397. <https://doi.org/10.1002/cl2.1397>

Authors Bios

Matteo Vergani

Matteo is Associate Professor in Sociology at Deakin University and specialises in radicalisation and hate crime, publishing in leading international journals and securing large research grants. He collaborates with numerous institutions and government agencies in Australia and Canada. His research advances the systematic consolidation of knowledge in hate and extremism studies through large-scale systematic reviews and the development of rigorous measurement tools of online and online hate and radicalisation. His research programme fosters multidisciplinary collaboration across social sciences, data science, econometrics and engineering, leveraging advanced technologies for analysing digital archives and social media big data.

Andrea Giovannetti

Andrea is a Lecturer in Economics and Quantitative Methods at the Australian Catholic University, with over a decade of experience in computational modelling of network dynamics. He has published in high-impact journals and secured major grants, including a Marie-Curie Postdoctoral Fellowship and funding from Leverhulme Trust and ESRC, totalling over AUD 1 million. He has held research positions at the University of Cambridge and collaborates with public stakeholders, including London Metropolitan Police and Merseyside Police, focusing on statistical and network-analysis techniques to analyse and predict deviant behaviour in networked offenders.

Dan Goodhardt

Dan is a researcher at Deakin University, where he has collaborated for over a decade on studies of hate crime and violent extremism in Australia and Southeast Asia. He has extensive experience in collecting antisemitism data through the Jewish Community's Community Security Group and currently works as a senior manager with Victoria Police. As co-convenor of the Practitioners Working Group on Tackling Hate in Victoria at the Centre for Resilience and Inclusive Societies, he connects security practitioners with researchers to advance efforts in countering hate and extremism.



Publication design by TAG.
www.wearetag.co