A Glance into the ‘Manosphere’: An Analysis of User Language, Radicalization & Engagement in r/Incels

Lydia Blum, Alexander von Janowski & Alexander Sobieska

Problem & Research question

- New platforms like **Reddit** enables users to radicalize, spread radical ideologies, incitement for terror, violence & harm
- Only limited research on **individual users and their development** over time in radicalized online spaces

Focus **Incels** – Involuntary Celibates – a misogynistic, violent, & fatalistic online movement.

Research question:

In what ways do users radicalize over time by participating in the subreddit r/incels?

Theoretical Framework

Grover and Mark (2019)

- **Fixation**: increasingly pathological preoccupation with a person or cause;
- **Identification**: identification with radical action and/or role model, and identification with their ideological in-group;
- **Leakage**: declaration of intent to do harm to a desired target.

Bilewicz & Soral (2019)

- Process of **desensitization**, hate speech reduces people's ability to recognize the offensive character of such language
- cycle of intergroup contempt and outgroup degradation

References

Baumgartner Jason, Zannettou Savvas, Keegan Brian, Squire Megan, Blackburn Jeremy (2020): The Pushshift, in: *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*, Association for the Advancement of Artificial Intelligence

Bilewicz Michał, Soral Wiktor (2020): Hate Speech Epidemic. The Dynamic Effects of Derogatory Language on Intergroup Relations and Political Radicalization, in: *Advances in Political Psychology*, 41(1)

Grover Ted, Mark Gloria (2019): Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit, in: *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media (ICWSM 2019)*, Association for the Advancement of Artificial Intelligence

Hypothesis

- Usage of radical, group specific vocabulary per user will increase over time (**H1**)
- Increased interaction leads to higher levels of group specific radical vocabulary (**H2**)

Data

- Collected from march 2016 to November 2017
- *n* = 878,195 posts by 9,140 authors
- Mean number of posts (PN) per author was 17.98 (*SD* = 79.15)
- Time difference between authors’ first and last comment: *M* = 44.21, *SD* = 89.02
- Average comments length 158.5 characters long (*SD* = 259.32, *Range* = 1-10,000)

Methodology & Analysis

- NLP with corpora for word frequencies:
  - list with the 4,000 most commonly used terms with manual sorting for *fixation, identification and leakage*
- Result: **dictionary** with **85 Incel-specific terms**, e.g. *‘hypergamy’, ‘chad’, & ‘suifuel’*
- Calculating **mean keyness score** (MKS) per author (*M* = 0.6, *SD* = 1.55) & **time difference** (TD) between individual users’ first and last post
- **Multiple regression analysis** with TD and PN as predictor variables and MKS as response variable
- **Cluster analysis** with four clusters (elbow method) was performed

Discussion

- **Reject H1, keep H2**.
- **Active interaction (PN)** is the best predictor for radicalization
- **Cluster analysis:** notable variances show that users are not very homogenous
- **Outliers:** Do some of these people radicalize outside of subreddits?
- **Limitations:** Only scraped reddit (incels.co); language as the best predictor?
- **Follow-up research:** Track individual user behavior across different subreddits
- Incels not originally violent radicalization -> **digital self-harming**, novel form of hate movement; more research is needed

Results

Table 1.

Regression analysis for number of posts and time difference predicting mean keyness score per author

	Estimate	Standard Error	*t*-Value	*p*-Value
Intercept	0.416	0.015	27.71	<2e-16*
Number of Posts	-0.0001	0.0002	-0.89	0.373
Time difference between Posts	0.011	0.0002	62.96	<2e-16*

Note. R^2 adjusted = 0.33. * $p < 0.05$

Table 2.

Cluster Summary for number of posts, time difference between posts and mean keyness score with included means and standard deviation (sd) for all clustered variables

	Number of Posts (sd)	Time Difference between Posts (sd)	Mean Keyness Score (sd)
Cluster 1 *(n = 7485)*	5.46 (11.68)	8.03 (16.27)	0.44 (1.25)
Cluster 2 *(n = 65)*	774.97 (347.01)	119.48 (106.13)	7.55 (5.49)
Cluster 3 *(n = 1204)*	52.64 (83.18)	113.75 (47.94)	1.25 (1.96)
Cluster 4 *(n = 665)*	22.12 (55.88)	318.28 (69.63)	0.58

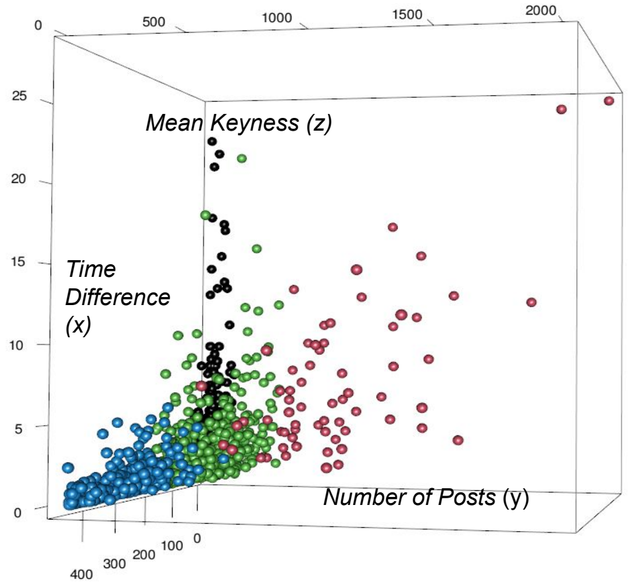

Figure 1. *Plotted values for mean keyness, time difference and number of posts per User. Clusters are displayed in different colours.*