1. Section 1: Fundamentals for Social Data Science

1.1 Polarizing Contributors on the "Politics" Stack Exchange

Although politicized online forums have the potential for fostering a productive discourse by engaging a diverse set of users, considerable research has found that they often devolve into echo chambers that amplify misinformation and polarization (Phillips et al. 2023). The following work will investigate the three most polarizing contributors to the 'Politics' stack exchange with reference to posts and comments made between 2013 and 2024. It will first assess the extent to which there are meaningful trends in the topics discussed by these users and then investigate similarities and differences in the comments they made. The analysis will find that all three users have posted on a broad range of topics and that although their comments vary in rhetoric from one another, they generally are not meaningfully distinguishable from the broader corpus.

Data on roughly 57,000 posts and 190,000 comments made by over 10,000 different users will be considered. The three most polarizing users were defined as those with the highest number of combined UpVotes and DownVotes under the condition that the ratio between the two values lies between 2:3 and 3:2. Also note that only answers (i.e., posts with a ParentPostId) were considered. Protected attributes, such as user location or names, were not revealed or factored into the analysis.

Table 1.1: User Activity Profiles

	User 101	**User 3135**	**User 18373**
Number of upvotes	2336	2937	4009
Number of downvotes	2122	3241	3983
Number of posts	235	444	1668
Number of comments	1182	2627	6414
Activity	2013–2022	2014–2024	2018–2024

Table 1.1 indicates that all three users have been very active on the forum for a prolonged period of time. The first part of this work now aims to identify common themes in the posts they create. Isabela Fairclough (2012) has grouped political discourse into three overarching narratives: policy, ideology, and current events. This distinction serves as a meaningful starting point since considerable literature has found that polarization in online forums is most prevalent in topics regarding ideology (Sinno et al. 2022). Testing this hypothesis for our users requires first grouping all posts into these categories. 20% of posts could be grouped with reference to their tags, after which a naïve Bayes classifier was trained on the actual content of the post and their newly assigned categories (labels) to classify the remaining 80%. Note that 86% accuracy was achieved when deploying the classifier on a test set.

Table 1.2: Topic Engagement by User

	Current Events	**Ideology**	**Policy**
User 101	50%	8%	42%
User 18373	48%	7%	45%
User 3135	36%	13%	51%

Generally, polarizers seem to engage less with topics that regard ideology than policy and current events. To quantify the significance with respect to the more general distribution of posts in each category, a Chi-squared test of independence was conducted. With a p-value of 0.07 the test shows that there is no statistically significant relationship between the type of topics that polarizing users engage with. It should be noted that users 101 and 18373 are remarkably similar in terms of their engagement across the three topics, while user 3135 created proportionally more ideological posts. Table 1.2 also reveals that polarizing users are certainly not confined to any individual topic, which Golf-Papez & Veer (2022) have shown is often indicative of 'troll-like behavior', especially when confined to a small subset of 'hot' topics, which would have been grouped into the current events category here.

Although analysing engagement across these three categories provides a meaningful starting point, assigning labels may also obscure meaningful trends in the similarity of their actual content. To this end, k-means clustering was applied to the TF-IDF representation of all posts. This required first pre-processing the text by removing links and HTML tags and tokenizing all remaining posts, after which stop words were removed and lemmatization was applied. The elbow method was then used to derive the number of relevant clusters.

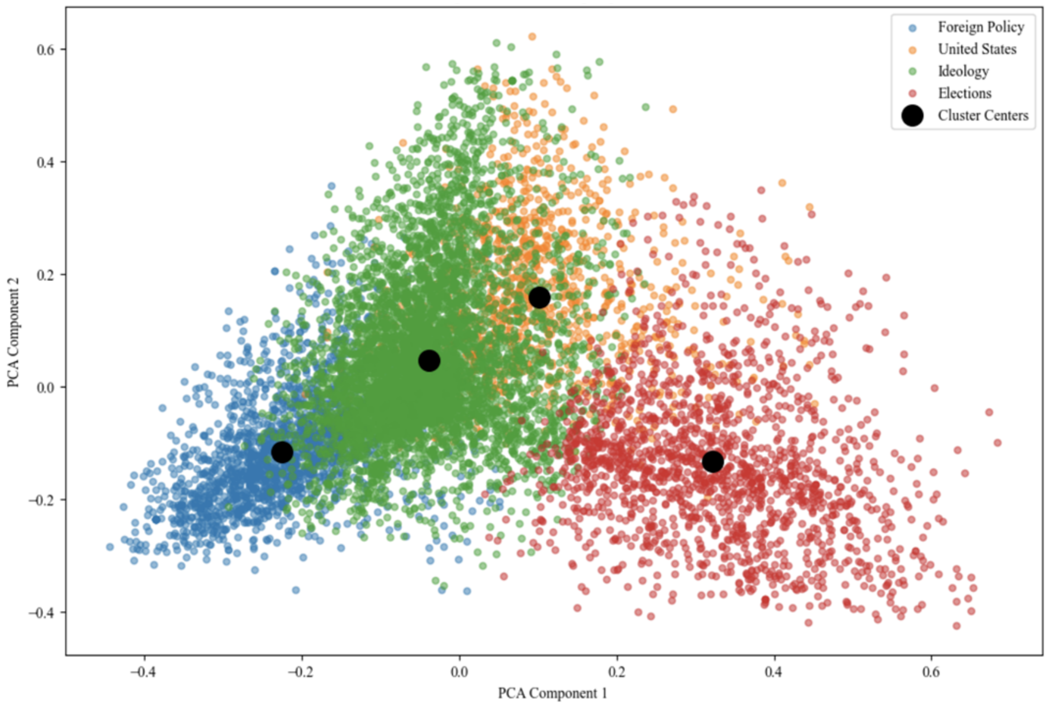

Figure 1.1: k-means Clustering for All Posts

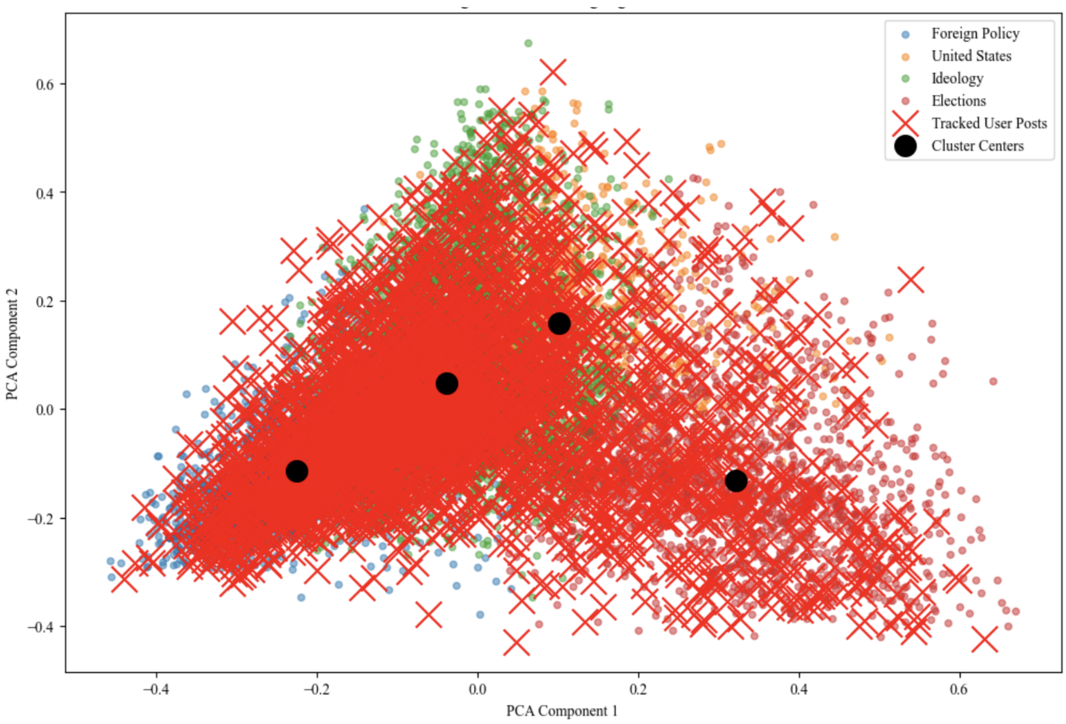

Figure 1.2: k-means Clustering with Target Users' Posts Marked

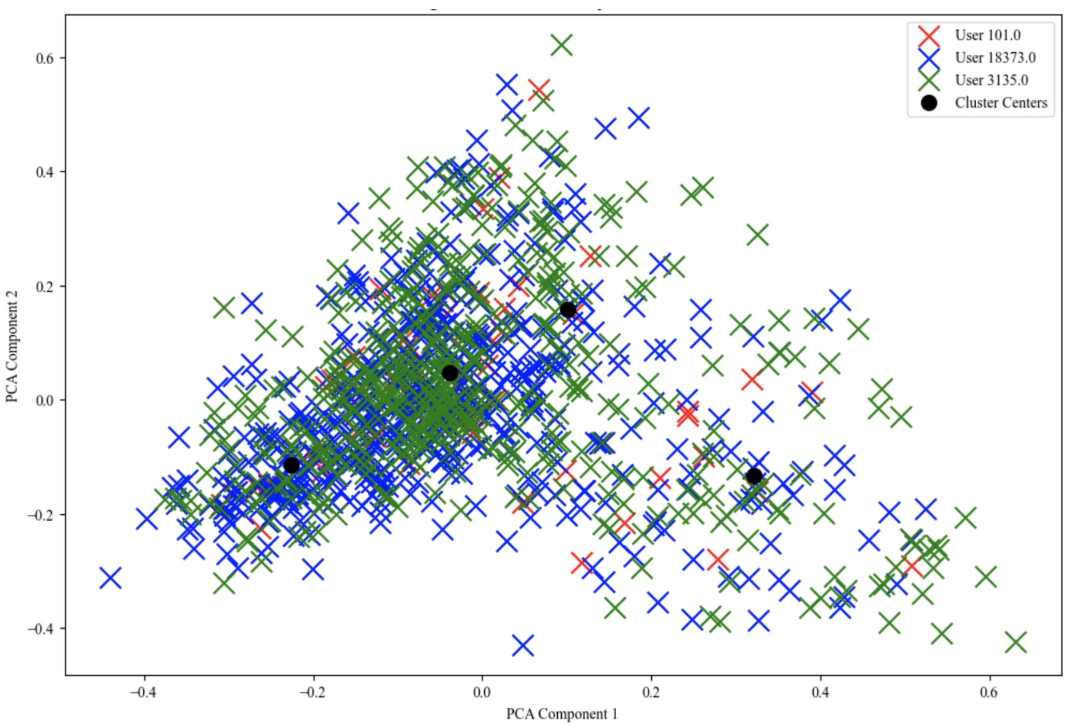

Figure 1.3: k-means Clustering With Sub-Sampled Target Users' Psts Marked

Figure 1.1 reveals clear subtopics discussed in the forum that roughly correspond to the three categories that posts were previously grouped into (note that cluster labels were assigned with Reference to TF-IDF Vectors). Figure 1.2 shows the same clustering, but with all posts created by the three users and Figure 1.3 marks 200 randomly sub-sampled posts made by the three users to provide insight into similarities between their posts without having user 18373's excessively large number of posts distort the analysis. The visualization further highlights that the three users were actively creating content on a broad range of topics, suggesting that polarizing users do not solely focus on a single thematic area, as has been found to be the case for extremists who often dominate discussions within online forums by fixating on specific topics (Reeve, 2019). These results may therefore be indicative of more productive engagement with the forum.

The second part of the investigation analysis comments made by the three users. Previous literature investigating why people receive upvotes and downvotes has shown that social media users that receive positive feedback are much more analytical and generally discuss topics of relevance with authority/confidence (Adaji et al. 2019), whilst dislikes have been linked to more emotionally charged but 'authentic' content.

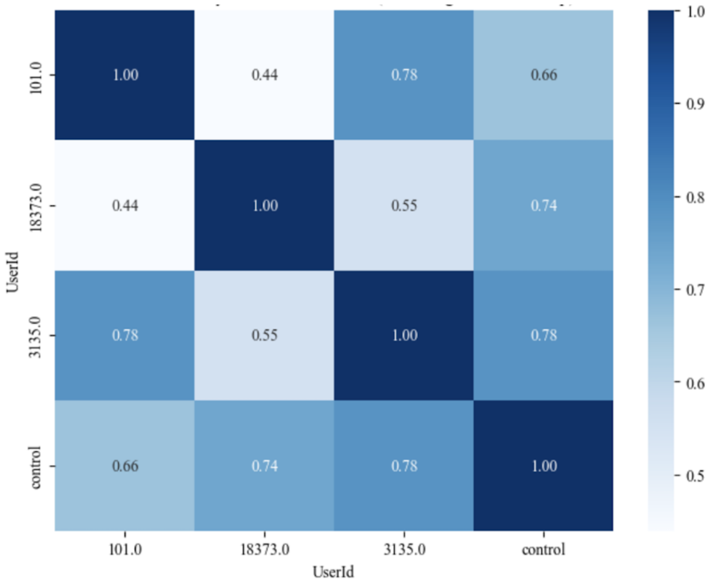

Figure 1.4: Cosine Similarity for Users' Comments

The cosine similarity matrix displayed in figure 1.4 measures the textual similarity between comments made by the three users, as well as a control group, which consists of 10,000 randomly sub-sampled comments in the forum. Since cosine similarity is derived with reference to TF-IDF it can capture overlap in word choice and topic similarity, but it neglects synonyms and word embeddings. The matrix reveals that user 3135 and user 101 are most similar, but values for the control also demonstrate that a similarity of 0.78 shouldn't be seen as meaningful. Instead, more attention should be given to the amount by which user 18373 differs from the other two.

The first part of this work found that there is no statistically significant relationship in the topics posted on by the three most polarizing users. The second part then assessed their comments, which revealed a considerable difference in semantics between the three users. These findings indicate that the polarizing nature of their content should not be attributed to the topic discussed, but rather to the opinions expressed or the sentiment with which they are conveyed online. Further exploring why their content generates mixed feedback may provide insights into drivers of ideological division and the spread of contentious narratives.

1.2 Citations

Phillips, S. C., Uyheng, J., Carley, K. M. (2023). A high-dimensional approach to measuring online polarization. *Journal of Computational Social Science, 6*(1), 1147–1178.

Golf-Papez, M., Veer, E. (2022). Feeding the trolling: Understanding and mitigating online trolling behavior as an unintended consequence. *Journal of Interactive Marketing, 57*(1), 90–114.

Reeve, Z. (2019). Engaging with online extremist material: Experimental evidence. *Terrorism and Political Violence, 33*(8), 1595–1620.

Fairclough, I., Fairclough, N. (2012). *Political discourse analysis: A method for advanced students* (1st ed.). Routledge.