9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Political discussions in online oppositional communities in the non-democratic context

Keywords: cross-cutting disagreement, affective polarization, autocracy, Russia, YouTube

Extended Abstract

According to the theory of *affective polarization*, emotions become the basis for discerning *"us vs. them"* and increasing intolerance toward the other side (Iyengar, Sood, & Lelkes, 2012). On the one hand, this phenomenon finds its manifestation in digital media when the level of incivility follows offline events of contentious politics (Theocharis, Barberá, Fazekas, & Popa, 2020). On the other hand, the political talk itself on social media can increase polarization among its participants (Yarchi, Baden, & Kligler-Vilenchik, 2020).

However, political tensions in oppositional communities in the non-democratic context, where the state controls traditional media as the main source of political information but still allows relative freedom on the Internet, attract less attention. To contribute to this nascent literature, I studied discussions in the community of the most vocal Russian opposition politician, Alexei Navalny, on YouTube. Although Navalny contributed to the escalation of existing tensions rather than creating them, his activity played a significant role in the launch of not only massive propaganda but also repressive campaigns against dissent by the ruling elite. As a result, both sides of the conflict see each other as an existential threat, resulting in affective polarization with a strong *"us vs. them"* division between the ruling elite and the opposition.

My focus on Navalny's YouTube channel is also because of the twofold role of this platform in the Russian media system. On the one hand, YouTube facilitates the promotion of the opposition's agenda and enlarges the political capital of independent activists. On the other hand, YouTube is still legal and available in Russia even after the full-scale Russian invasion of Ukraine and the subsequent shutdown of foreign social media platforms.

I address two research questions: (1) What happens when the pro-government narrative finds its manifestation in the community of the most vocal opposition politician? (2) What are the potential and limits of incivility, as a characteristic of affective polarization, to engage in political discussions?

The pooled dataset contains 8,980,313 comments from 407 videos. I use a variety of methods to answer these research questions. First, to identify pro-government and pro-opposition comments and detect cross-cutting disagreement, I trained a supervised machine learning model—the class affinity model (Perry & Benoit, 2017) —based on a dictionary with derogatory words applied to Navalny and his supporters, Putin and the government. I detected such words on the basis of an iterated computer-assisted keyword selection approach suggested by King, Lam & Roberts (2017). Second, logistic regression models and local polynomial fits are used to study the relationship between conversations and incivility. Finally, I show how the level of toxicity changes over time.

The main empirical findings are the following. First, pro-government comments (1) attract Navalny's supporters, who respond to the out-group criticism, and (2) contribute to the emergence of pockets of a pro-government narrative (Table 1). Second, top-level comments that open discussions tend to be more uncivil than those without discussion threads. But toxicity has its limits. Users are not willing to dispute with those who spread extreme forms of incivility with a null potential to deliberate. Third, the level of incivility of comments gradually goes up (Figure 1) with time passing after a video release during the first 14 hours and then stabilizes for top-level comments that have discussion threads and thread comments themselves.

Table 1: Results of multinomial logistic regression of discussion type

	Reference 'No discussion underneath of a top-level comment'			
	Attacks from both sides	Attacks on government	Attacks on opposition	Discussion
	(1)	(2)	(3)	(4)
Pro-Government	1.152 (0.056)	0.296 (0.056)	1.191 (0.039)	0.432 (0.038)
	p = 0.000	p = 0.00000	p = 0.000	p = 0.000
Pro-Opposition	−0.685 (0.053)	−0.056 (0.029)	−0.893 (0.042)	−0.352 (0.027)
	p = 0.000	p = 0.297	p = 0.000	p = 0.000
Toxicity (binary)	0.404 (0.048)	0.204 (0.028)	0.209 (0.030)	0.061 (0.018)
	p = 0.000	p = 0.00002	p = 0.00002	p = 0.202
Comment length (log)	1.273 (0.021)	0.788 (0.019)	0.939 (0.016)	0.603 (0.013)
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Count of Likes (log)	1.634 (0.022)	1.124 (0.017)	1.045 (0.019)	0.837 (0.013)
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Second 2 hours	−0.081 (0.054)	0.074 (0.036)	0.001 (0.042)	0.065 (0.027)
	p = 0.133	p = 0.037	p = 0.979	p = 0.017
Third 2 hours	−0.016 (0.060)	0.137 (0.042)	0.102 (0.042)	0.153 (0.028)
	p = 0.793	p = 0.001	p = 0.015	p = 0.000
Fourth 2 hours	0.085 (0.078)	0.136 (0.050)	0.236 (0.061)	0.233 (0.046)
	p = 0.278	p = 0.006	p = 0.000	p = 0.000
Fifth 2 hours	0.250 (0.077)	0.226 (0.045)	0.426 (0.069)	0.336 (0.042)
	p = 0.001	p = 0.004	p = 0.00000	p = 0.00002
Sixth 2 hours	0.282 (0.101)	0.411 (0.072)	0.516 (0.093)	0.453 (0.068)
	p = 0.005	p = 0.00005	p = 0.00000	p = 0.00001
Seventh 2 hours	0.490 (0.090)	0.418 (0.064)	0.563 (0.081)	0.536 (0.075)
	p = 0.00000	p = 0.00001	p = 0.000	p = 0.000
After more than 14 hours	0.569 (0.091)	0.458 (0.078)	0.699 (0.085)	0.599 (0.082)
	p = 0.000	p = 0.00000	p = 0.000	p = 0.000
Intercept	−12.636 (0.097)	−8.112 (0.070)	−8.589 (0.061)	−5.142 (0.043)
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
Akaike Inf. Crit.	3,527,597	3,527,597	3,527,597	3,527,597
Bayesian Inf. Crit.	3,528,304	3,528,304	3,528,304	3,528,304
Observations	5,965,458	5,965,458	5,965,458	5,965,458

Note: Video clustered standard errors are in brackets

This research contributes to the extant literature on affective polarization on social media, shedding light on patterns and peculiarities of online political discussions within an oppositional community in a non-democracy. The findings advance our understanding of the behavior of out-group commenters who attack the domain of their political opponents, eventually forming pro-government hotspots. Furthermore, I was able to identify a limited potential for incivility in order to initiate discussions and report on the dynamics of those discussions' incivility.

References

- Iyengar, S., Sood, G., & Lelkes, Y. (2012, sep). Affect, not ideology. Public Opinion Quarterly, 76(3), 405–431. doi: 10.1093/poq/nfs038
- King, G., Lam, P., & Roberts, M. E. (2017, apr). Computer-assisted keyword and document set discovery from unstructured text. American Journal of Political Science, 61(4), 971–988. doi: 10.1111/ajps.12291
- Perry, P. O., & Benoit, K. (2017). Scaling text with the class affinity model. arXiv. doi:10.48550/ARXIV.1710.08963
- Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020, apr). The dynamics of political incivility on twitter. SAGE Open, 10(2). doi:10.1177/2158244020919447
- Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020, jul). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. Political Communication, 38(1-2), 98–139. doi: 10.1080/10584609.2020.1785067

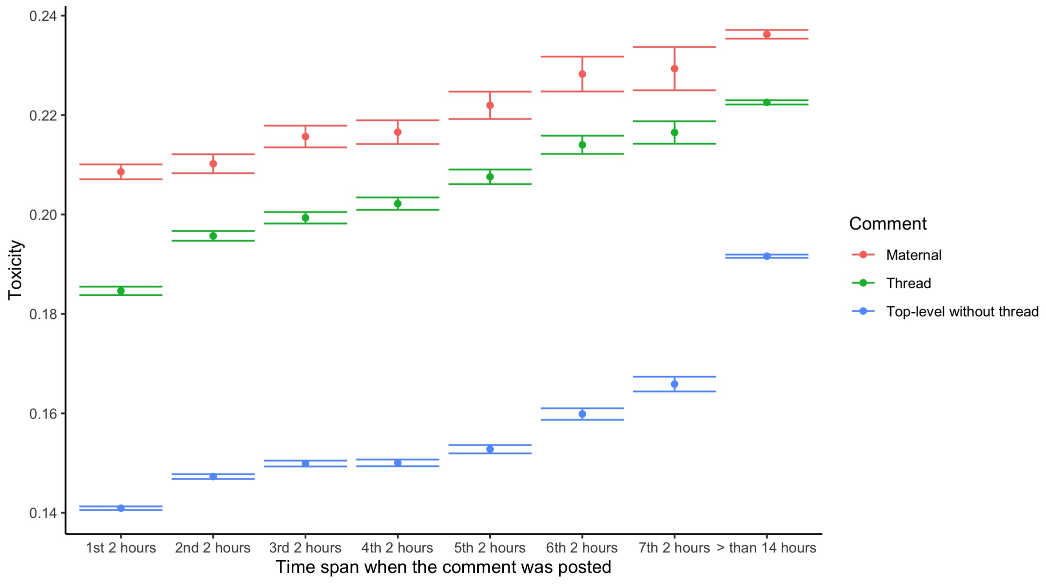

Figure 1: Toxicity score of comments by time of comment posting (2-hour bins)