Ambivalence of LGBT Visibility in Croatian Video-Based Social Media Content

Zoran Fijavž

Faculty of Arts, University of Ljubljana
Aškerčeva 2, SI-1000 Ljubljana

Abstract

We examine the influence of video content on the presence of socially inappropriate discourse in a dataset of annotated Facebook comments, representative of publicly accessible LGBT-related posts from major news sources in Croatian.
Even when controlling for post source and topicality we find posts with video content received significantly more negative LGBT-oriented than negative other-oriented comments. We interpret the results through previous discussions of LGBT visibility as both a necessity for effective advocacy via norm contestation and contact theory pathways as well as an opening for potential backlash in the form of abusive language.

1. Introduction

Discussions of visibility have been central to theorizing in LGBT and queer studies and created accounts of visibility that overreach the (un)desirability of visibility and foreground its ambivalence (Edenborg, 2019).[1] On one hand, visibility creates effects consistent with the contact hypothesis: intergroup interaction may, under optimal contact conditions, reduce intergroup bias (Allport, 1954). While the original formulations of the contact hypothesis focused on racial and ethnic groups, research within the paradigm has expanded to various minority groups and generally finds a inverse relalationship between prejudice and intergroup contact, even in non-optimal contact conditions (Pettigrew and Tropp, 2000). Furthermore, similar changes in prejudice have been demonstrated in imagined interactions (Crisp and Turner; 2009), parasocial (mediatized) video contact with non-fictional (Cooley and Burkholder, 2011) as well as fictional LGBT people (Schiappa, Gregg, and Hewes, 2007). The extended contact hypothesis posits and finds evidence for improved attitudes from merely knowing that in-group members have cross-group friendships (Zhou et al., 2018).

However, while public visibility of LGBT people may facilitate changes in attitudes via contact mechanisms, it may also cause various forms of backlash. In an exhaustive examination of LGBT norm diffusion in the European context, visibility was found to be central to the ability of activists to place demands on states and change attitudes, while simultaneously provoking contestation from various actors (Ayoub, 2016). In theorizing about backlash occurences, it is furthermore important to define in their particular form, as Bishin, et al. (2015) find no shifts in public opinion following the US supeme court ruling on gay marriage in 2015 as well as in experimental conditions: a "negative, large, and enduring reaction against a policy or group", as the authors define backlash, was thus not observed following legislative change. In addition, backlash in the form of banning pride parades may, provided there is external leverage, further increase the salience of the LGBT rights norm and mobilize new allies (O'Dwyer, 2012).

While research on LGBT visibility often focuses on the relationship between nation states and a transnational LGBT norm (Bosia, 2013; Lind, 2014; Edenborg, 2019), this article will adopt the lens of visibility and potential backlash to analyze engagement with LGBT content in online interactions. Drawing on discoursive theories, the discourse produced by nation states may be conceived as a single instance of speech with productive capacities among others, such as in cases of hate speech by individuals (Butler, 1996), which suggests that theories of visibility and backlash may also be applied to online user-generated content. Indeed, literature on computational propaganda demonstrates that, given the proliferation of possibilites to author public texts, discourse hyperproduction is a practicable mode of indirect state-orchestrated censorship (Wu, 2018): individual users can produce discourse that is ultimately pooled into a particular mode of state speech, ranging from innocuous-yet-strategic distraction (King et al, 2017), to targeted online harassment (Kargar and Rauchfleisch, 2019) and technological amplification (Abokhodair, Yoo, and McDonald, 2015). Furthermore, social media platforms serve a similar function as traditional media, with 55% of respondents in Croatia listing social media as a news source and 89% reporting receiving news online, which includes both social media and the websites of online news outlets (Newman et al., 2019). The functional similarity also warrants the expansion of (critical) discourse analysis approaches typically applied to traditional media (e.g. Gabrielatos and Baker, 2008) to user-generated content.

While there is an extensive body of research on various forms of abusive language, it is often motivated with the goal of (automatic) detection and or online coversation moderation (Pavlopoulos et al., 2017; Waseem and Hovy, 2017). However, the aforementioned line of research often aims at optimizing classifiers according to a particular typology or harmful speech, without investiganting the patterns in the sources used to train classifier in the first place. In particular, the notion of user-generated content may obscure the fact that there may not be a unitary, homogeneous user: according to Eurostat 2019 data, 56% of respondents aged 16 to 74 and 88% of those aged under 25 used social media EU-wide. In spite of wide-spread social media use, user-generated content (or behavioral data) can not be considered equivalent to population-level survey data. In Croatia, for example, 40% survey respondents report sharing news

[1] While some variant of LGBTI, LGBTIQ+, or LGBTIAQ+ or queer as an umbrella term could be used, we will use the acronym LGBT as used in Ljubešić et al. (2019) for the purpose of terminological consistency.

online and 25% comment on social media or websites (Newman et al., 2019).

2. Goal of the paper

This paper seeks to further analyze the effect of using video content as opposed to text in Facebook posts about LGBT people. We analyzed the Croatian LGBT section of the FRENK dataset (Ljubešić et al., 2019), which contains manually annotated Facebook comments. This subsection was particularly interesting because it contained both posts with text and links to news articles as well as embedded videos from three major Croatian news sites (Jutranji list, Index and 24sata).[2] The hypothesis was that posts containing video had a different distribution of comment tags with more negative LGBT-oriented posts compared to non-video posts.

3. Dataset decription and methodology

We analyzed the manually annotated LGBT-related Croatian FRENK dataset of socially unacceptable discourse. The following subsections explain the adjustments made to the dataset for analysis as well as post metadata we included in the analysis.

3.1. FRENK dataset and comment selection procedure

It is important to note that the FRENK dataset gathered Facebook posts of news sites in a relatively exhaustive manner. The Facebook posts of the most visited news sites (defined by web traffic to their main webpages) were retrieved using the public Graph API, and samples of 100 LGBT as well as non-LGBT-themed posts were used to train a SVM classifier. All retrieved posts were then classified. A sample of the classification results was checked manually, added to the posts from the first iteration and used to train the second version of the SVM classifier, giving the final selection of LGBT-related posts, which were then manually annotated for speech type and targets. While Ljubešić et al. (2019) do not analyze the content of the posts classified as LGBT-themed, we find in section 3.2. that most of the posts are related to pride parades in Croatia, which is consistent with previous studies that find these events draw large amounts for public LGBT visibility, particularly when organized for the first time (Fejes and Balogh, 2013).

In total, the dataset contained 5,787 comments from 22 Facebook posts. Out of which we removed 316 comments with multiple type-target annotations. Out of all possible type-target combinations, we only retained comments with the six most common type-target tags, comprising 97% of all comments in the dataset, which were:

1. Acceptable speech, No target (Ac);
2. Background offensive, LGBT (BO);
3. Background violence, LGBT (BV);
4. Inappropriate, No target (In);
5. Other offensive, Commenter (OC);
6. Other offensive, Other (OO).

For describing our results, we group these tags further into: neutral comments (*Acceptable speech, No target*), negative LGBT-oriented comments (*Background offensive, LGBT* or *Background violence, LGBT*) and negative other-oriented comments (*Other offensive, Commenter* or *Other offensive, Other*).

3.2. Metadata: video presence, post source, topicality

In addition to the comment tags, we examined the presence of embedded videos in the posts, the post source and post topicality. We marked whether posts included directly **embedded videos,** playable with Facebook's domain, as a binary category. The embedded videos varied from lengthy live stream recordings and short video interviews to a 360-degree video montage with snapshot moments. 10 out of 22 posts contained embedded videos. The posts originated from the Facebook profiles of three different **sources**: Jutranji list, Index.hr, and 24sata, all of which are major Croatian news sites. Comparing the source and video presence data reveals that in our dataset, 24sata exclusively posted video posts, Index.hr predominantly posted text posts (which nonetheless may link to video content outside of Facebook), and Jutranji lists posted the same number of both video and non-video posts. Notably, all except one were related to various pride parades in Croatia, which is why we further examined the **topicality** of the posts.

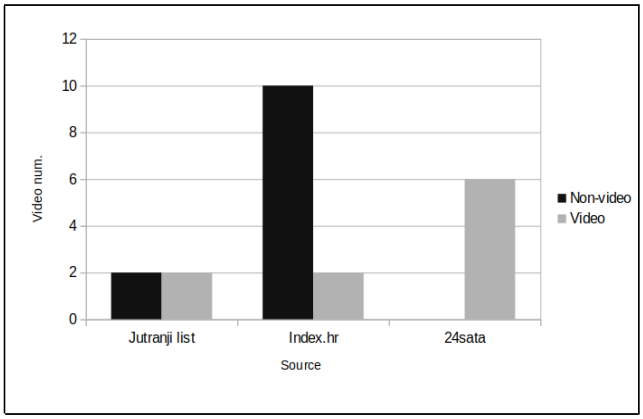

Figure 1: Video presence in posts by source

The topicality of posts was introduced to distinguish between posts focusing directly on pride parades and those not. As the video posts were nearly universally related to pride parades (with one exception), any differences identified between video and non-video posts may simply be due to the events they report and not video presence. Indeed, as seen in Figure 2, 18 of the 22 posts were published in June, when pride parades typically take

[2] Jutranji list: https://www.jutarnji.hr/
Index: https://www.index.hr/
24sata: https://www.24sata.hr/

place[3]. Manually examining the posts revealed that all posts in June were related to pride parades, while none in July or February were. While the number of video and non-video posts in June were the same, video posts in June received far more comments (3,261) compared to non-video comments (1,441), together comprising 81,3% of all comments in the dataset.

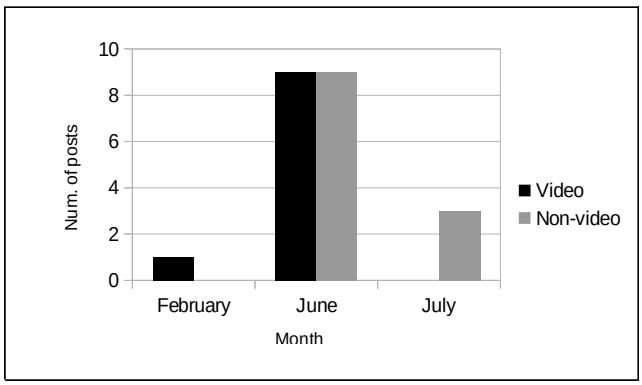

Figure 2: Video presence by month

4. Analysis

We tested whether the distribution of comment tags was statistically significantly different between posts with or without embedded videos, using a chi-square test of independence and calculated the standard residuals to determine which categories in particular differ between the two subsets and are over- or underrepresented. We used Cramér's V to measure the effect size of the results.[4]

The chi-square test of independence, Cramér's V and standardized residuals for each comment tag were also calculated on comment subsets to account for the effect of the difference between the sources and topicality of posts.[5]

4.1. General results

The results show a weak significant relationship between the video presence and comment tag distribution, $\chi^2(5, n = 5{,}357) = 175.133$, $p < 0.05$, $V = 0.180$. The distribution of all comment categories except *Inappropriate, no target* differed significantly. The differences were directional, with LGBT-oriented negative comment categories overrepresented in video posts, while neutral and other-oriented negative comments overrepresented in the non-video posts, as visualized in Figure 3 (i.e. their standardized residuals were positive or negative).

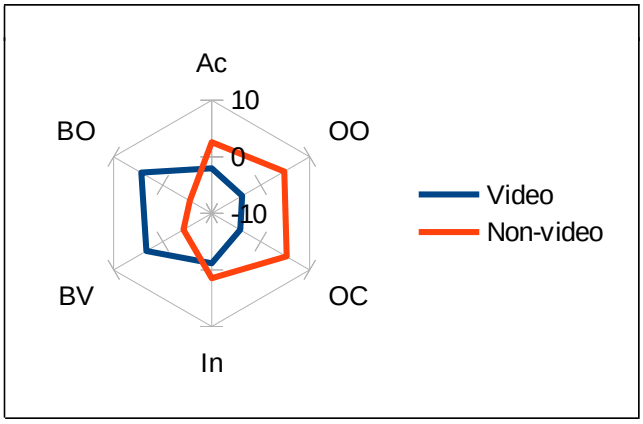

Figure 3: Standardized residuals per comment category

4.2. Controlling for source: Analysis of Jutranji list

As noted, not all sources used video posts to the same extent, so the results of 4.1. could influenced by the post source (e.g. more or less favorable reporting or social media audience). That is why we performed a separate analysis for Jutranji list, which had the same number of video and non-video posts. The difference in comment tag distributions were significant, $\chi^2(5, n = 1{,}705) = 58.304$, $p < 0.05$, $V = 0.184$. The directionality of the residuals was comparable to the results in 4.1. as seen in Figure 4: the video posts contained more negative LGBT-oriented comments and less other-oriented negative comments compared to non-video posts. In spite of the similar contours, the standardized residuals for neutral comments as well as comments expressing LGBT-related violence were close to zero and the differences overall were less pronounced (e.g. the overall residuals were smaller), with only 4 out of 12 residuals being over 1.645 (compared to 10/12 for the whole dataset).

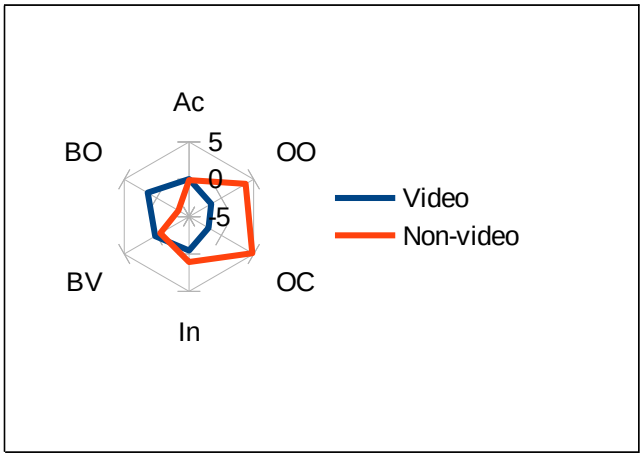

Figure 4: Standardized residuals per comment category (Jutranji list)

[3] For the purpose of our analysis, the particular year the article was published is not relevant, as the month of publishing corresponds to content related to pride parades. All posts were published in either 2011, 2016 or 2017.

[4] For df = 1 for all comparisons in this paper, the threshold value of V for small effects is 0.10, 0.3 for medium effects and 0.5 for large effects (Gravetter and Wallnau, 2016; 586).

[5] A standardized residual is a ratio: the difference between the observed count and the expected count divided by the standard deviation of the expected count in chi-square testing. Standardized residuals can be interpreted in the same manner as z-scores and were considered significant if greater than 1.645.

4.3. Controlling for topic orientation and moth of publishing

We also examined the influence of topicality on the distribution of comment tags. As discussed in section 3.2. nearly all video posts were related to pride parades and posted in June and none of the posts in July contained video. To disentangle the effect of video presence from topical orientation, two comparisons were made: firstly between the video and non-video posts published in June exclusively and secondly between non-video posts in July and June.

The results of separately analyzing posts in June mirror the main results. The difference between video and non-video posts was significant, $\chi^2(5, n = 4{,}379) = 66.198$, $p < 0.05$, $V = 0.122$. As seen in Figure 5, negative LGBT-oriented comments were above the expected frequencies in video posts and negative other-oriented comments were below it, while the inverse was true for non-video posts. One notable difference to the main results is that the absolute value of standardized residuals for neutral comments (*Acceptable speech, No target*), was less than 1, and the standardized residuals were smaller overall with 7/12 reaching a critical value over 1.645.

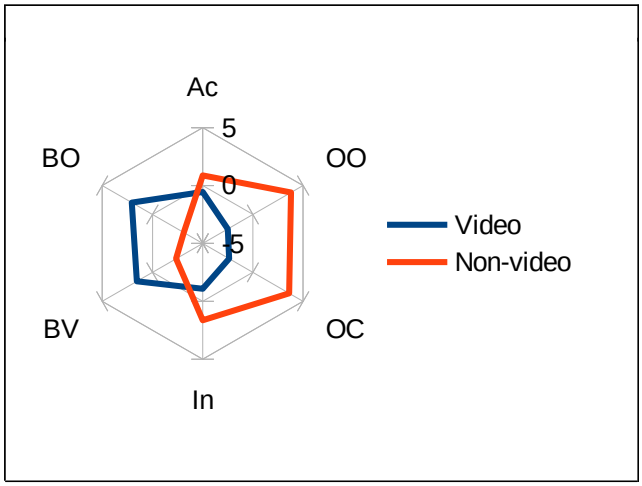

Figure 5: Standardized residuals (June only)

The comparison between non-video posts in June and July also had significant results, $\chi^2(5, n = 2{,}058) = 33.847$, $p < 0.05$, $V = 0.128$. As shown in Figure 6, there was a significant difference in offensive negative LGBT-oriented comments and the standardized residuals for LGBT-oriented comments including violence in June were slightly under the significance threshold at 1.545. Other-oriented negative comments, however, were more consistent across the months. In comparison the results for June alone, there was also a larger difference in the presence of neutral comments.

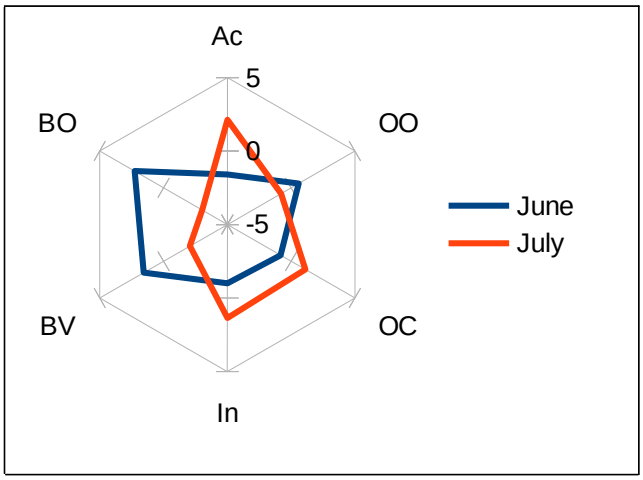

Figure 6: Standardized residuals (June and July, text)

The results from both comparisons in 4.3. provide further analysis of the main results. The results from June reflect a form of trade-off between negative other- and LGBT-related comments depending on the presence of video content, as observed in the main results. The comparison of text posts between June and July suggests that the main results can not be explained by topicality alone, as a different topicality (pride-related content and other content) lacks the difference in negative other-oriented comments seen in the main results. However, there is a significant overrepresentation of offensive negative LGBT-related comments and nearly significant overrepresentation of negative LGBT-related comments tagged as violent, which may be due to the overall increased visibility during a pride month.

5. Conclusion

This article hypothesized that LGBT-related posts in with video content received a different distribution of comment types compared to non-video ones in the FRENK dataset of socially unacceptable speech for Croatian. When comparing video and text posts, we found a weak positive and significant association between videos posts and negative LGBT-oriented comments and a weak and significant negative association between video posts and negative other-oriented comments. The directionality of these effects persisted when analyzing the source with the fewest posts and same number of video and non-video posts (Jutranji list), with the caveat that differences in standardized residuals did not reach critical values for all type-target tags. The main finding persisted when we only analyzed posts related to pride parades (all published in June). We also compared text posts published in June and in July (that is pride-related and non-pride-related text posts) and found increased presence of negative LGBT-related comments as well as a lower difference in other-oriented negative comments in comparison with the main results. Overall, the results show that what we considered a form of backlash in online LGBT visibility is more frequent in content related to pride parades and in particular to video content showing pride parades.

Such an interpretation woul be consistent with previous accounts of visibility (Ayoub, 2016) although on a micro-level in user-generated content: Visibility both allows for norm diffusion through multimedia content on social networks as well as potentially producing backlash in the form of socially unacceptable discourse including background-related threats. It is imporant to stress that these findings can not be read normatively as a reason to dissuade attemps at online LGBT visibility, as we merely analyzed the presence of negative LGBT-oriented comments and did not have external measures for their mid- or long-term effects: the results might well point towards a possibility to anticipate such comments and use them for countermobilization by LGBT and allied groups along the lines of O'Dwyer (2012).

Furthermore, the paper elaborates on the content of the posts gathered in the Croatian FRENK dataset: a large proportion of the posts was about pride parades in Croatia and thus published in June, consistent with qualitative studies on the role of pride parades in making LGBT communities visible in terms of media presence (Fejes and Balogh, 2013). This is an important fact for further use of the Croatian section of the FRENK dataset (and possibly others), as even the relatively exhaustive approach used for the FRENK dataset to gather LGBT-related posts mostly captured season-bound occurences (or alternatively, media articles on the topic are published nearly exclusively in relation to pride parades in June). 81,3% of all comments in the Croatian FRENK dataset originate from posts related to pride parades and the comments on pride-related videos alone comprise 56,4% of all comments in the dataset. This could be exploited in further research to compare other linguistic contexts, where the public online communication is less season-bound and more diffused and integrated into the reporting of news sites. This finding is also important for efforts regarding comment classification through machine learning approaches. Firstly, it is a cautionary tale against interpreting data gathered through automated means as a form of public opinion, as in our case the dataset clearly captures discourse of a non-random population around pride parades as a yearly peak of LGBT visibility in the context of Croatia. Secondly, the fact that the dataset is bound up with pride parades raises the question of generalizability of classifiers trained in this manner to other contexts: e.g. it is conceivable that headlines on a highly discussed legal case would produce comments that are not comparable to the ones in the FRENK dataset.

This paper may also provide some methodological insights. While the multi-variable analysis of categorical data is hardly novel, the development of the hypothesis on the effect of video content was formulated through the authors involvement in the comment annotation of the Croatian LGBT as well as other sections of the FRENK dataset, which required close reading of the comments and consideration of their pragmatic context, while theoretical frameworks of LGBT visibility, contact theory and backlash were incorporated after the statical analysis was performed. The author also had extensive background knowledge on LGBT-related topics, which eased the selection and interpretation of post metadata (i.e. introducing post topicality to further examine the main hypothesis). In this sense, the research process resembled grounded theory approaches, which favor hypothesis construction based on observations rather than fixed prior theoretical frameworks (Strauss and Corbin, 1994).

However, there are some limitations to this study. Firstly, there is a number of questions on the generalizability of the results. The results cannot be generalized outside of social media posts, as social media users differ significantly from the general population (Mellon and Prosser, 2017; Eurostat, 2019), a relatively low number of users actually write comments and there are demographic-related preferences for consuming content in video versus text form (Newman et al., 2019). Secondly, chi-square tests were used on data subsets and there was no assessment of the family-wise error rate (false discoveries while performing multiple hypothesis tests), which may be beyond the type-I error rate of each individual test at $p = 0.05$. That could be avoided by applying logit models (e.g. logistic regression), although the topicality and video presence would need to be added manually, as they are not part of the original FRENK dataset.

6. References

Amy Lind. 2014. 'Out' in International Relations: Why Queer Visibility Matters. *International Studies Review*, 16(4):601–604.

Anselm Strauss and Juliet Corbin. 1994. Grounded Theory Methodology: An Overview. In: Norman K. Denzin and Yvonna S. Lincoln, eds., *Handbook of Qualitative Research*, 1st edition, pages 273–285. Thousand Oaks, California, Sage Publications.

Benjamin Bishin, Thomas Hayes, Matthew Incantalupo, and Charles Anthony Smith. 2015. Opinion Backlash and Public Attitudes: Are Political Advances in Gay Rights Counterproductive? *American Journal Of Political Science*, 60(3):625–648.

Conor O'Dwyer. 2018. The Benefits of Backlash: EU Accession and the Organization of LGBT Activism in Postcommunist Poland and the Czech Republic. *East European Politics And Societies: And Cultures*, *32*(4):892–923.

Costas Gabrielatos and Paul Baker. 2008. Fleeing, Sneaking, Flooding. *Journal Of English Linguistics*, 36(1):5–38.

Edward Schiappa, Peter B. Gregg, and Dean E. Hewes. 2005. The Parasocial Contact Hypothesis. *Communication Monographs*, 72(1):92–115.

Emil Edenborg. 2019. Russia's spectacle of "traditional values": rethinking the politics of visibility. *International Feminist Journal Of Politics*, 22(1):106–126.

Eurostat, the statistical office of the European Union. 2020. ICT usage in households and by individuals [Data file]. Retrieved from: http://tiny.cc/c5z4oz (Accessed 15. 5. 2020)

Frederick Gravetter and Larry Wallnau. 2016. *Statistics for the Behavioral Sciences, Tenth Edition*. Boston: Cengage Learning.

Gary King, Jennifer Pan and Margaret Roberts. 2017. How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument. *American Political Science Review*, 111(3):484–501.

Gordon Willard Allport. 1954. *The nature of prejudice.* Boston: Addison-Wesley.

John Pavlopoulos, Prodromos Malakasiotis, Ion Androutsopoulos. 2017. Attention to Abusive User Content Moderation. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1125–1135, Copenhagen, Denmark, Association for Computational Linguistics.

Jonathan Mellon and Christopher Prosser. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*, 4(3).

Jonny Cooley and Gary Burkholder. 2011. Using Video and Contact to Change Attitudes Toward Gay Men and Lesbians. *Journal of Social, Behavioral, and Health Sciences*, 5(1):83–90.

Judith Butler. 1996. *Excitable speech*. New York: Routledge.

Nárcisz Fejes and Andrea Balogh. 2013. *Queer visibility in post-socialist cultures*. Bristol: Intellect.

Nic Newman, Richard Fletcher, Antonis Kalogeropoulos, and Rasmus Kleis Nielsen. 2019. *Reuters Institute Digital News Report 2019*. Oxford: Reuters Institute for the Study of Journalism.

Nikola Ljubešić, Darja Fišer, Tomaž Erjavec. 2019. The FRENK Datasets of Socially Unacceptable Discourse in Slovene and English. In: *Proceedings of International Conference on Text, Speech, and Dialogue, TSD 2019*, pages 103–114, Ljubljana, Slovenia.

Norah Abokhodair, Daisy Yoo, David Wright McDonald. 2015. Dissecting a Social Botnet. *Proceedings Of The 18th ACM Conference On Computer Supported Cooperative Work and Social Computing - CSCW '15*, pages 839–851, New York, United States, Association for Computing Machinery.

Phillip Ayoub. 2016. *When States Come Out: Europe's Sexual Minorities and the Politics of Visibility*. New York: Cambridge University Press.

Richard Crisp and Rhiannon Turner. 2009. Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychologist*, 64(4), 231–240.

Shelly Zhou, Elizabeth Page-Gould, Arthur Aron, Anne Moyer, Miles Hewstone. 2018. The Extended Contact Hypothesis: A Meta-Analysis on 20 Years of Research. *Personality And Social Psychology Review*, 23(2):132–160.

Simin Kargar and Adrian Rauchfleisch. 2019. State-aligned trolling in Iran and the double-edged affordances of Instagram. *New Media and Society*, 21(7):1506–1527.

Thomas F. Pettigrew and Linda R. Tropp. 2000. Does intergroup contact reduce prejudice? Recent meta-analytic findings. In: Stuart Oskamp, ed., *Reducing prejudice and discrimination: Social psychological perspectives*, pages 93–114. Mahwah, New Yersey, Lawrence Erlbaum Associates.

Tim Wu. 2018. Is the First Amendment Obsolete? *Michigan Law Review*. 117(3).

Zeerak Waseem and Dirk Hovy. 2016. Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In: *Proceedings of the Student Research Workshop,* pages 88–93, San Diego, California, Association for Computational Linguistics.