Investigating the Effects of Popularity on Comment Civility:
A YouTube Case Study

Department of Visual Art and Design
J.N. Andrews Honors Program
Isaac Smith & Dr. Heather Day

Abstract

This study will explore the connection between the civility of internet comments and their popularity-through the example of two YouTube videos. By using a combination of manual review and computer analysis, readers will be able to discover if video popularity has a negative effect on comment civility. My findings will also examine correlations between the factors that indicate civility, and provide examples of notable and intriguing comments I found during my study.

Introduction

I am interested in pursuing research similar to Han et al. and Dhiraj and Sanjay, utilizing textual analysis to explore comment civility. The Law of Big Numbers is defined by Dhiraj and Sanjay as a tendency towards discourse becoming less civil on popular forums. The amount of comments on popular videos makes it difficult to regulate discussion. The focus of the study is a comparative analysis between the comment sections of two videos- theSeraphim17 with 31 comments, and theRadbrad with 23,758 comments. (which were reduced with representative random sampling to 419 comments.)

Methodology

The comments have been evaluated using textual analysis based on methods used by Walsh, and Dhiraj and Sanjay. I obtained the comment text and information with Youtube Comment Scraper. I also utilized the Linguistic Inquiry and Word Count (LWIC) software for analysis. The statistical data it provides (the presence of words indicating action and emotion) shows the presence of factors indicating civility. My primary hypothesis is that the smaller video (theSeraphim17's) will have more civil comments compared to theRadBrad's. The LIWC scores have been evaluated in addition to the raw, original comment text.

Results

The LIWC results are mostly inconclusive. Emotional Tone and Authenticity were the only two categories that showed statistically significant differences between the samples. theRadBrad's comments scored highly for both of these positive categories, which was the opposite of what the hypothesis predicted. theRadBrad's comments clearly had more errors in spelling, grammar, and punctuation compared to theSeraphim17's, but that isn't a determining factor of civility.

Conclusion

The sample size of two videos and 450 comments is small. LIWC is versatile software, but it does have its limitations- it works better with longer segments of data which contain more context. Examining comment chains instead of single comments is a possible solution. Additionally, LIWC supports the creation of custom libraries to more accurately reflect the subject of analysis. There are many more possibilities for research and investigation into this topic.

Selected Annotated Bibliography

Dhiraj, M.; Sanjay, S. Visualizing YouTube's comment space: online hostility as a networked phenomena.
New Media & Society. **2018.**
https://doi.org/10.1177/1461444818792393

Dhiraj and Sanjay used a similar methodology of numerical identification to distinguish and evaluate several factors found to be present in several related YouTube comment sections that were known to have large amounts of inflammatory discussion. Their focus was on tracking how different users spread hateful language and reacted to each other's actions, using network analysis (NodeXL) to create a large scale graph that represented users as nodes and their interactions as lines connecting nodes. They focused on providing a more balanced perspective on the issue instead of simply criticizing YouTube and it's users, and their detailed coding schema to determine comment factors was a great help in my decision-making. In addition to providing an example of procedure and precedent for my own comment focused research, I'm also using analysis software (LWIC) to aid in my information gathering and provide statistical support for my conclusions.

Soo-Hye Han, LeAnn M. Brazeal, and Natalie Pennington. Is Civility Contagious? Examining the Impact of Modeling in Online Political Discussions. *Social Media + Society.* **2018.**
https://doi.org/10.1177/2056305118793404

Han et al. experimented with the effect of discursive cues on online political discussions. In the course of introducing their experiment, they provided an interdisciplinary definition of civility and an explanation of the nature and consequences of online incivility. Civility includes politeness and respect towards the topics, opinions, and participants along with the discussion forum, while incivility is the opposite behavior. They predicted and found that participants who were exposed to civil comments became more likely to behave in a civil manner and that metacommunication is likely to have a positive effect on the overall quality of the situation. In addition to providing an example of statistically measuring and evaluating civility, it's results about the usefulness of moderation and metacommunication are extremely useful to support my own hypothesis. I believe that a smaller channel is more likely to experience metacommunication among its members, and that modeling will be more effective in those communities.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin. DOI: 10.15781/T29G6Z
http://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_LanguageManual.pdf
http://liwc.wpengine.com/wp-content/uploads/2015/11/LIWC2015_OperatorManual.pdf

Pennebaker, et al. developed the LIWC software to efficiently study and measure the emotional and structural components in samples of writing. LIWC is primarily a tool for psychometric analysis, examining large quantities of text for the presence It is an incredibly versatile and customizable software, using a customizable 'dictionary' (developed and evaluated by a committee of qualitative judges) that searches for words and word stems that fit into a wide range of significant categories- like 'power words', 'informal language', and 'quantifiers.' The dictionary can also be customized, to add specific words or terms unique to a certain field or discipline. I'm incredibly excited about the possibilities that LIWC provides, as it's specific statistical categories will make it far easier to evaluate the large quantities of information I'm working with.