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

Isaac Smith | Department of Visual Art & Design | 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 (LIWC) software for analysis. 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. theSeraphim17’s comments scored highly for both of these positive categories, which was the opposite of what the hypothesis predicted. The RadBrad’s comments clearly had more errors in spelling, grammar, and punctuation compared to theSeraphim17’s, but that isn’t a determining factor of civility.

Tone: Radbrad mean score 62.952, Seraphim mean 44.18.
Authenticity: Radbrad mean score 54.399, Seraphim mean 31.81.

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.

Dhiraj, M.; Sanjay, S. Visualizing YouTube’s comment space: online hostility as a networked phenomena. New Media & Society. 2019.