Lauren Beard

Computational Content Analysis

Professor Evans

27 March 2020

Violence in Movies over Time

**Introduction**

A great deal of popular media attention espouses the idea that movies have become more violent over time. With public announcements warning parents against higher levels of weapon violence, explicit language, and more, greater attention has focused on what children and viewers generally should be consuming. Most of the studies that establish this phenomenon focus on select movies from various time periods, with content analyses being done on samples smaller than 50 movies. This paper aims to address this phenomenon utilizing novel computational methods, with a total sample of 150 movies, over three times the typical sample in this course of study. As such, themes of violence in 75 movies created during the time period of 1940-1960 are compared to 75 movies created in the time period of 2000-2020. This study provides further nuance to this phenomenon, while solidifying the findings and offering novel insight into the mechanics of violence in 1940-1960 in comparison to those of 2000-2020. Such insights provide further clarification and guidance for public audiences, parents who aim to develop psychologically sound and informed viewing habits, and researchers across diverse fields.

**Background**

Multiple studies highlight the fact that violence themes, words, and images have increased in movies over time. Such forms of increases include greater duration of violent scenes, higher levels of death and near-death encounters, and more instances of weapon use. Other researchers focus on expanded understandings of violence, including the amount of explicit language or instances of conflict between movie characters.

Organizations such as the American Academy of Family Physicians have made clear posting briefings with information such as, “studies have found that 91 percent of movies on television contained violence, even extreme violence.” Quite often these studies recognize that increased levels of violence in movies is related to increased acceptance and tolerance of violence and violent themes among public audiences, thus manifesting general public opinion. Still, facts such as “children and adolescents in the U.S. spend an average of about seven and a half hours a day using various forms of entertainment media” highlight that investigations into what children, adolescents, and even adults warrant attention (American Academy of Physicians 2019).

In their recent JAMA article, health researchers Fumie Yokota and Kimberly Thompson focused specifically on changes in duration of violent content, death, and weapons use in 74 G-rated animated feature films released in theaters between 1937 and 1999. These researchers defined violence as “intentional acts (e.g., to cause harm, to coerce, or for fun) where the aggressor makes some physical contact that has potential to inflict injury or harm” — thus indicating a focus on physical violence. Their analysis of time trends indicated a statistically significant increase in the duration of violence in the films over time, the association between violence and conflict between “good” and “bad” characters, and a wide range of weapons use (e.g. body, sword, poison, gun, explosive, etc.) (Yokota & Thompson 2000).

A similar study, published in the American Association of Pediatrics, identified the presence of violence in segments of the top 30 films since 1950 and the presence of guns in violent segments since 1985, the first full year the PG-13 rating was used. Here, Bushman et al. determined that violence in films has more than doubled since 1950 and that gun violence in PG-13-rated movies has tripled since 1985. As such, these researchers noted that PG-13-rated films have contained as much or more violence as R-rated movies in 2009 (Bushman, Jamieson, Weitz, & Romer 2013).

Increasing levels of violence are further significant given established studies on the relationship between viewing violence and increased violent or aggressive tendencies (Huesmann 2007). For example, psychologist Josephson randomly assigned 396 boys, ages 7–9 years old, to watch either a violent or a nonviolent movie before they played a game of floor hockey in school. For those boys who were rated as frequently aggressive by their teachers, the combination of watching a violent movie stimulated significantly more aggressive behavior. Randomized experiments have produced similar results (i.e. preschoolers who physically attack each other more often after watching violent videos (Bjorkqvist 1985) and for older adolescents with histories of delinquency who get into more fights on days that they see more violent films (Leyens, Parke, Camino, & Berkowitz 1975)).

In assessing the few studies above and associated content, researchers utilize multiple definitions of the term “violence” and vary what they consider to be of interest in association to violence. In general, researchers define media violence as visual portrayals of acts of physical aggression by one human or human-like character against another (Huesmann 2007). The World Health Organization, however, defines violence as “the intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, which either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment, or deprivation” (WHO 2011). Given that this definition offers a wide-ranging conceptual framework for understanding what may be deemed violent (as evidenced by the inclusion of the physical and psychological), this will serve as the guiding definition for this paper. Using this expanded definition and keywords from previous studies, three sub-domains of words that guide this analysis are 1) words related to state (i.e. ‘die’, ‘dead’, and ‘death’), 2) action (‘kill’, ‘shoot’), and 3) emotion (‘afraid’, ‘fuck’).

**Methods**

This study differs from previous studies that focus on assessing violence in movies in two major ways. The first is that it uses computational content analysis techniques, rather than more traditional methods of trained human coders. While trained coders can be taught to code nearly identically, these methods remain susceptible to human bias. Computational techniques offer an alternative way to code and analyze media like movie scripts to gain similar insights.

Further, the approaches used here also benefit from managing a larger sample size than most previous studies. The vast majority of studies that analyze violence in movie scripts maintain a sample size of less than 50, so as to be digestible to human coders. As this is not a constraint for computational techniques, this study assesses a total of 150 movie scripts, three times the amount of larger, human-coder-based studies.

The 150 movie scripts analyzed in this report are derived from ‘The Movie Corpus’, as created by Mark Davies, Professor of Linguistics at Brigham Young University. This corpus alone offers 200 million words across movie scripts created from the year 1930 to 2018. For this study, the total corpus was truncated to only include movies created from 1940 to 2020, with a sub-sample being created for each 20-year time period (i.e. 1940-1960, 1960-1980, 1980-2000, and 2000-2020).

The final sample of 150 movie scripts consists of 75 scripts ranging from 1940 to 1960 and an additional 75 scripts from the years 2000 to 2020. This subsampling allowed for an in-depth investigation into the differences surrounding themes of violence in an earlier time period in comparison to the most current 20-year time period. The results listed below shed further insight into the changes in (and increased use of) violence in movie scripts from the time period of 1940-1960 in comparison to 2000-2020. The associated code for this project can be found at the following GitHub repository: <https://github.com/lrnbeard/Content-Analysis-2020/tree/master/final/code>.

**Results and Analysis**

The following results assess uses and themes of violence in a number of ways, including word associations, semantic associations, network analyses, and more.

1. **Exploratory Analysis: Establishing the phenomenon**

This first set of analyses offered insight into whether or not these questions on violence proved relevant to this corpus, particularly in relation to a 1940-1960 to 2000-2020 comparison. In this initial phase, words and phrases were counted across all movie scripts. Then, using the encoded scripts, words were disambiguated using part-of-speech (POS) tagging and normalized through stemming and lemmatization. This allowed for the creation of word clouds of top words for the two subsamples, along with a count of top nouns, adjectives, and more.

In assessing the word clouds for 1940-1960 and 2000-2020, figure one makes clear that there are some differences in what words maintain the highest level of frequency in comparing these two subsets. Of note is the decreasing usage of the word ‘right’ from time point one to time point two, indicating a possible trend in considering changing themes of violence, particularly when considering the conceptual relationships between violence-nonviolence and right-wrong. Here, the shrinking of the term ‘right’ indicates a possible shift in violence between these two time points.

This indication that this corpus may provide insight into shifting themes of violence highlighted the need to assess changes in terms like ‘right’ over the time period of the entire corpus, as it ranged from 1940 to 2020. As further discussed below, the results from using Word2Vec techniques were used to assess linguistic change in the words ‘right’ and a keyword for this project, ‘dead.’ In this case, the dimensions of multiple embeddings for the words ‘dead’ and ‘right’ were assessed over time, allowing for the identification of semantic change as the word vectors change their loadings for focal words. Figures two and three make clear drastic semantic change for both words in the mid-1990s. Such an immense change solidified the significance of comparing the years 1940-1960 to 2000-2020.

Using the techniques of part-of-speech tagging, this analysis further revealed a notable increase in the use of explicit language in the later time point. For instance, the 1940-1960 subsample included zero explicit nouns in the top 100 nouns for the subsample. The 2000-2020 subsample, however, included the words ‘shit’, ‘hell’, ‘fuck’, and ‘ass’ all within the top 100 nouns. Direct references to words like ‘fuck’ tie into the keywords developed for this project.

1. **Topic Modeling**

Before delving into a more focused word-level analysis of these movie scripts, topic modeling (a method of two-dimensional content clustering) was used to gain insight into topical differences between the two time points. These computationally-induced topics are sparse distributions over nonexclusive clusters of words, which allows for the description of the scripts via sparse mixtures.

In assessing topic modeling, ten topics proved to provide interesting insights into each sample. Figures four and five provide visualizations via stacked bar charts of the ten topics produced for ten movies in each sample, and figures six and seven do the same via heatmaps. The primary result of interest in performing in these two subsamples was the notable rise in violent words in topics for 2000-2020 in comparison to 1940-1960.

Figures eight and nine, however, provide direct insight into the word-based differences between these two time periods. In assessing violent words across the 1940-1960 topics, violent words include ‘shoot’ (2), ‘police’ (3), ‘gun’ (1), and ‘hit’ (1). The topic that matches to most of the movies in figure eight is topic #7, which only includes the keyword ‘police’. This exists in stark contrast to the topic modeling results for time point two. Keywords resulting for this topic include ‘fuck’ (10), ‘death’ (1), ‘shoot’ (5), ‘die’ (2), and ‘fight’ (1). Note that this amounts to not only more keywords, but also the presence of keywords at a much higher frequency.

1. **Network Analysis**

This notable difference and increasing presence of violence over time indicates that it may be linked to more concepts over time. Network analysis (with the aid of *networkx,* a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks) was utilized to address this question.

The first network technique employed in this project was a simple visualizing of the relationships between the movie scripts for each of the two time periods. A starkly different network in one time period compared to the other could indicate a sub-par comparison, as the two microcosms of movie scripts could be too structurally different to offer valuable insight into micro-relationships. However, figure ten demonstrates that these two timepoints offer quite similar network structures between scripts, with each time period maintaining one central cluster of most movies and three to four outlier movies.

This verification thus supported the investigation of micro-level word trends, as made possible via co-occurrence graphing. Using co-occurrence graphs, it is possible to visualize the above-chance that two terms from a text corpus will appear alongside each other in a certain order. Given that this corpus is so large, all edges with weight below 25 were dropped from the co-occurrence graph, along with all isolates. The 1940-1960 co-occurrence graph (figure 11) offers little that is unexpected: there is no significant cluster indicating violence, and primary words associated with the time period appear (i.e. ‘sir’, ‘gentleman’, ‘lady’). The only keyword associated with this project that appears in this graph is the term ‘afraid.’ What is interesting, however, is that a notable ‘violent’ cluster appears in the 2000-2020 co-occurrence graph (figure 12), including words like ‘tick’, ‘tock’, ‘frantic’, and ‘deathstyle.’ The emergence of this cluster marks an explicit instance in greater levels of violence in the later time point than in the earlier time point. The word ‘afraid’ re-appears as well.

With this document and cluster-level analysis, word-level relationships offer an additional layer of insight. One way of doing this is through measures of centrality, where the concept of centrality is that some nodes (being words or documents) are more central to the network than others. This means that nodes with the highest number of connections are the most central. Running centrality measures between these two time points further solidified increasing levels of violence, as evidenced by an increased measure for the term ‘afraid’ (time point one: 0.013, time point two: 0.014). Ego networks for keywords also highlighted increasing levels of violence, as is evident in the two-fold increase from three connections to six connections for the term ‘die’ from 1940-1960 to 2000-2020 (figure 13), the increase from eight connections to eighteen connections for the term ‘kill’ (figure 14), and the increase from six to seven connections for the term ‘shoot’ (figure 15).

1. **Word Embeddings**

In this next section, the movie scripts are represented as densely indexed locations in dimensions instead of the sparse mixtures of topics explored in section II. Here, the process included 1) specifying an underlying number of dimensions for the samples, 2) training a model with a neural network auto-encoder that best described script words in their local linguistic contexts, and 3) assessing the locations of those the words in the space in order to make sense of how they were produced. These spaces were assessed to produce three major findings: script words that are similar to the predefined keywords, T-distributed Stochastic Neighbor Embedding plots, and semantic categories relating to themes of violence. Such analyses provide insight into the cultural background that informed the construction of these movie scripts.

Cosine similarity was used to determine which words were most similar to the keywords for each subsample:

1. In 1940-1960, the word ‘kill’ was most similar to words like ‘fight’ (0.999), tough (0.999), ‘fella’ (0.999), and ‘horse’ (0.999). In 2000-2020, the word ‘kill’ was most similar to words like ‘die’ (0.999), ‘today’ (0.999), and ‘run’ (0.999).
2. In 1940-1960, the word ‘shoot’ was most similar to words like ‘fast’ (0.999) and ‘risk’ (0.999). In 2000-2020, the word ‘shoot’ was most similar to words like ‘money’ (0.999) and ‘motherfucker’ (0.999).
3. In 1940-1960, the word ‘die’ was most similar to words like ‘hurt’ (0.999). In 2000-2020, the word ‘kill’ was most similar to words like ‘today’ (0.999), ‘tomorrow’ (0.999), and ‘kill’ (0.999).
4. Lastly, the word ‘fuck’ was not present in the 1940-1960 subsample. However, in 2000-2020, the word ‘fuck’ was most similar to words like ‘shit’ (0.998), ‘kill’ (0.996), and ‘afraid’ (0.996), all of which are major keywords for this analysis.

In looking at the word similarities, the violent keywords are associated with more immediate and similarly violent words (i.e. ‘today’, ‘die’, and ‘shit’) in 2000-2020 compared to 1940-1960.

In further looking at how words are embedded across the movie scripts, dimension reduction was used to visualize word vectors (with PCA to reduce the dimensions to 50, and T-SNE to project those to the two visualized). The word vectors for the top 50 to 150-word vectors are visualized in figures 16 and 17, given that the first 50-word vectors in both subsets represented mostly commonplace, less meaningful words. Here again we see that the visualization for 1940-1960 includes little on themes of violence, essentially only represented as the word kill in the space (circled in blue). The visualization for the 2000-2020 subset, however, includes two separate violence-related clusters, one consisting of ‘die’-‘stay’-‘help’ and another consisting of ‘shoot’-‘fuck’.

Word embeddings can also be projected to semantic dimensions, which allows for insight into cultural associations of words. In this instance, two primary dimensions were constructed to assess themes of violence: nonviolence and violence, one the basis of words like ‘right’, ‘wrong’, ‘shoot’, ‘kill’, ‘stop’, ‘save’, etc. Words associated with gender and incentive were then mapped onto these two dimensions, in order to assess changes between (non-)violence, gender, and incentive over time (figures 18 and 19). A greater elaboration on these results is assessed in the discussion section.

1. **Deep Neural Networks: Language Prediction**

Lastly, deep neural network techniques allow for the possibility of text generation. That is to say, it is possible to train a language generation model using both the 1940-1960 and 2000-2020 subsets. The models were trained using code derived from the Hugging Face Team, which allows writing with a transformer. Example predictions associated with themes of violence were made possible via the text generation feature. These examples, as followed below, demarcate increased usage and ease with violence concepts among the 2000-2020 movie scripts:

"The definition of violence is"

1940-1960: The definition of violence is to kill. - What's the use of killing? - I don't know. I don't know what it means to kill. (-0.999)

2000-2020: The definition of violence is that it's a way of life. (+0.709)

"Guns are"

1940-1960: Guns are not allowed in the house. - No, no, no. (-0.999)

2000-2020: Guns are not the only thing that makes you feel good. (+0.992)

"Guns are"

1940-1960: Fuck who? - I'm sorry. (-0.999)

2000-2020: Fuck who? - I'm not gonna be able to find you. (-0.999)

Note that these examples also include sentiment scores, with sentiment analysis being the process of examining a piece of text for opinions and feelings. The result is that scores for the different predictions can be compared over time to assess change in opinion about violence (which indicates a stark shift to positive scores from 1940-1960 to 2000-2020 in the first two examples). These results are further analyzed in the discussion section.

**Discussion**

These various sub-analyses of themes of violence in 1940-1960 compared to themes of violence in 2000-2020 produce several interesting points of discussion. The first is that this corpus replicates the findings of previous studies: movies have in fact become more violent (linguistically and conceptually) over time. This project demonstrates that this increase in violence occurs across all three sub-domains of investigation, as violence relates to state (‘die’, ‘dead’, ‘death’), action (‘kill’, ‘shoot’), and emotion (‘afraid’, ‘fuck’).

The findings bolstered these previous studies in a number of ways. It demonstrated that there are 1) conceptual changes in words like ‘right’ and ‘dead’ over the time from 1940 to 2020, 2) greater numbers of violent-based topics in the 2000-2020 subsample than in the 1940-1960 subsample, 3) higher frequencies of violent words from time point one to time point two, and 4) greater relationships between violent words and other words and concepts from time point one to time point two. Each of these findings point to the fact that not only is there increased violence in the 2000-2020 subsample, they also highlight the fact that violence is intertwined with a greater number of concepts -- and thus more embedded in the cultural reality of these more recent movies.

This analysis further probed into the cultural changes surrounding the relationship between violence and gender. For instance, in looking at the relationships between violence, nonviolence, and gender, the results indicate that nonviolence is consistently more associated with words gendered female (i.e. ‘lady’, ‘woman’), and violence is consistently more associated with words gendered male (i.e. ‘man’). This is made clear through the fact that the top words for nonviolence are ‘lady’ in 1940-1960 and ‘woman’ in 2000-2020 (with ‘gentleman’ and ‘man’ being least associated); the top words for violence are ‘man’ in both 1940-1960 and 2000-2020 (with ‘lady’ and ‘woman’ being least associated).This indicates that despite some fluctuations in the relationship between gender and violence, violence remains a largely masculine concept in movie scripts.

This project also offered potential insight into the cultural changes surrounding the relationship between violence and incentives for violence over time. Here, the results demonstrate that both nonviolence and violence are consistently associated with the term ‘good’, which points to interesting understandings of what violence and nonviolence mean to the larger social audience of these movie scripts. Of additional interest is that nonviolence becomes much more associated with the word ‘lose’ from time point one to time point two, whereas violence becomes much more associated with the word ‘win’ from time point one to time point two. As such, one may infer that these more recent movie scripts indicate a greater emphasis on violence as a possible means of winning in this subsample’s contemporary social landscape, whereas selecting nonviolence more often lends to a loss. Further, these movie scripts are primarily written by English-speakers, and, given the immense value placed on winning in many English-speaking countries (i.e. the U.S.), one could infer that the heightened relationship between violence and the word ‘win’ indicate a cultural valuing of violence as a means over nonviolence among these speakers. Such questions require greater investigation.

Of final cultural relevance are the language predictions noted in section five of the results. Here, it becomes clear that cultural understandings of violence are much more flexible and accepted in 2000-2020, as demonstrated in the shift from “The definition of violence is to kill. - What's the use of killing? - I don't know. I don't know what it means to kill.” to “ The definition of violence is that it's a way of life.” This is reflected in the large shift in the sentiment score, which goes from -0.999 in 1940-1960 to +0.709 in 2000-2020, buttressing this increased acceptance. This is also reflected in the predictions on guns, which go from “Guns are not allowed in the house. - No, no, no.” (-0.999) to “Guns are not the only thing that makes you feel good.” (+0.992); this indicates that whereas guns are taboo in the home in 1940-1960, they offer a source of pleasure in 2000-2020 (again represented in the stark positive increase in sentiment scores over time). Lastly, when asking the model “Fuck who?” it apologizes “Fuck who? - I'm sorry” in 1940-1960 versus responding “Fuck who? - I'm not gonna be able to find you” in 2000-2020. This again highlights increased cultural acceptance of violence words and concepts over time, in relation to violence in daily life, feelings associated with weapons, and violence against other people.

In considering broader associations between the findings of this project and grounded social life, several intriguing questions come to the fore. Of the most pressing: if violence is supposedly lessening in prevalence in the social world, then why is violence increasing in movie spaces? Some postulate that this presence and increase in violence merely represents the manifestation of traditional U.S.-centric values in movies, those which are firmly aligned to violence and thus offer palpable and marketable tropes to U.S. viewers (Theodore-Vachon 2016). One could similarly postulate that these themes are reflective of a sort of collective memory playing out in these films that captivates intergenerational audiences.

Of additional note is increasing tolerant attitudes towards the representation of violence and explicit language in particular as driving greater violence in movies. This may be related to the general relaxing of public opinion around various previously taboo topics, such as homophobia or interracial marriage. In this vein, some point to related attitudes around taboo topics like swearing as reflective of increasing individualism and emphasis on freedom of expression in social spaces (Twenge, VanLandingham & Campbell 2017). More work on these larger questions deserves attention in order to glean a fuller understanding as to why these contradictory and increasing media-based states of violence are emerging.

**Conclusion**

While these findings mark an initial exploration into cultural changes around violence over time, this project makes clear that studying violence in movies offers numerous interesting insights. While it remains contested how violence in media affects viewers, it appears clear that violence has become more acceptable among movie writers, producers, and audiences over time (Huesmann 2007). Furthermore, the relationship between violence and gender and violence and incentives like winning and losing offer several future routes of study. For instance, one may be interested in major social or political events that have emboldened violence as a means of winning over time, as possibly related to war or economic trade.

It is of further note to mention that several limitations do exist in this project. First off, these analyses depend on the subsample of 150 movie scripts pulled from the Davies Corpora, which may not be the most relevant to those engaged with movies. Further, it only begins a very initial exploration into violence as it relates to the initial keywords, being ‘die’, ‘dead’, ‘death’, ‘kill’, ‘shoot’, ‘afraid’, and ‘fuck’. However, there are several other ways one could go about such analyses, such as focusing exclusively on weapons-based or explicit language. Lastly, this project does not include any analyses on audio, video, or movie images, which could enrich such a study as this a great deal more.

Beyond that, the project did succeed in highlighting the increased presence over violence in more recent movie scripts than older movie scripts. It further brought several interesting routes of future study to the fore, which could be additionally unpacked using the methods offered in this paper. Such analyses offer great potential in making sense of how cultural understandings of violence are infused into and produced in movies, lending to work done across sociology, psychology, media studies, and more.

**References**

Björkqvist Kaj. (1985). *Violent films, anxiety and aggression: experimental studies of the effect*

*of violent films on the level of anxiety and aggressiveness in children*. Helsinki, Finland:

Soc. Scientiarum Fennica.

Bushman, B. J., Jamieson, P. E., Weitz, I., & Romer, D. (2013). Gun Violence Trends in Movies.

*Pediatrics*, *132*(6), 1014–1018. doi: 10.1542/peds.2013-1600d

Definition and typology of violence. (2011). Retrieved from

<https://www.who.int/violenceprevention/approach/definition/en/>.

English-Corpora.org. Retrieved from <https://www.english-corpora.org/>.

Huesmann, L. R. (2007). The Impact of Electronic Media Violence: Scientific Theory and

Research. *Journal of Adolescent Health*, *41*(6), S6–13.   
doi: 10.1016/j.jadohealth.2007.09.005

Josephson, W. L. (1987). Television violence and children's aggression: Testing the priming,

social script, and disinhibition predictions. *Journal of Personality and Social Psychology,*

*53*(5), 882–890. https://doi.org/10.1037/0022-3514.53.5.882

Leyens, J.-P., Camino, L., Parke, R. D., & Berkowitz, L. (1975). Effects of movie violence on

aggression in a field setting as a function of group dominance and cohesion. *Journal of*

*Personality and Social Psychology*, *32*(2), 346–360. doi: 10.1037/0022-3514.32.2.346

Theodore-Vachon, R. B. (2016). PG-13 Blockbusters and the Sugarcoating of Violence.

Retrieved from <https://www.nytimes.com/roomfordebate/2016/06/30/pg-13-blockbusters-and-the-sugarcoating-of-violence/violence-in-movies-reflects-a-cultural-reality>.

Twenge, J. M., Vanlandingham, H., & Campbell, W. K. (2017). The Seven Words You Can

Never Say on Television: Increases in the Use of Swear Words in American Books,

1950-2008. *SAGE Open*, *7*(3). doi: 10.1177/2158244017723689.

Yokota, F., & Thompson, K. M. (2000). Violence in G-Rated Animated Films. *Jama*, *283*(20),

2716–2720. doi: 10.1001/jama.283.20.2716

Violence in the Media and Entertainment (Position Paper). (2019). Retrieved from

<https://www.aafp.org/about/policies/all/violence-media.html>.

**Appendix**

|  |  |
| --- | --- |
| **Figure 1. Word Cloud Comparison** | |
| 1940-1960 | 2000-2020 |
|  |  |

|  |
| --- |
| **Figure 2. Conceptual Change over time of the Word ‘Right’** |
|  |

|  |
| --- |
| **Figure 3. Conceptual Change over time of the Word ‘Dead’** |
|  |

|  |
| --- |
| **Figure 4. Topics Across Ten Movies (1940-1960)** |
|  |

|  |
| --- |
| **Figure 5. Topics Across Ten Movies (2000-2020)** |
|  |

|  |
| --- |
| **Figure 6. Topic Heatmaps for 10 Movies (1940-1960)** |
|  |

|  |
| --- |
| **Figure 7. Topic Heatmaps for 10 Movies (2000-2020)** |
| A screenshot of a cell phone  Description automatically generated |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Figure 8. Words within Topics (1940-1960)** | | | | | | | | | | |
|  | Topic 0 | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 |
| 0 | majesty | uh | johnny | uh | doc | majesty | uh | uh | johnny | majesty |
| 1 | uh | captain | picture | mary | shoot | highness | picture | gentle-man | ma | uh |
| 2 | shoot | police | lie | john | wan | child | majesty | john | gilbert | john |
| 3 | dr | school | uh | child | gun | johnny | child | captain | pa | picture |
| 4 | king | paint | madam | police | sister | goodbye | dr | paint | circus | child |
| 5 | pete | alright | eye | lie | alright | uh | wan | lie | gentle-man | gentle-man |
| 6 | lie | gentle-man | tommy | captain | uh | steve | rock | police | child | circus |
| 7 | guard | horse | captain | gentle-man | mary | papa | ship | child | bye | papa |
| 8 | order | avalan- che | uncle | sister | ta | bye | gentle-man | mary | eye | tommy |
| 9 | hit | doc | mary | george | car | daddy | doc | bye | car | horse |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Figure 9. Words within Topics (2000-2020)** | | | | | | | | | | |
|  | Topic 0 | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 |
| 0 | buddha | fuck | jerry | song | fuck | fuck | fuck | fuck | johnny | fuck |
| 1 | fuck | ron | fuck | fuck | ron | dude | jerry | tick | billy | betty |
| 2 | george | ralph | lewis | jersus | christmas | shoot | dad | song | darren | shoot |
| 3 | violin | game | dad | vote | dad | dad | princess | dude | kristen | vampire |
| 4 | genre | shoot | sam | die | connor | send | love | ron | turkey | amy |
| 5 | teach | dad | eddie | child | shoot | party | center | shoot | edward | lose |
| 6 | mickey | win | laugh | thomas | truth | lady | lewis | lose | fuck | george |
| 7 | death | medal | honey | lie | dude | fight | charlie | life | die | ted |
| 8 | frank | drive | ride | sin | lose | lose | drive | nell | ali | rock |
| 9 | peter | alright | lose | lose | song | dance | whoa | death | dad | music |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | **Figure 10. Document Network Structure** | | | 1940-1960 | 2000-2020 | |  |  | |

|  |
| --- |
| **Figure 11. Co-occurrence Graph (1940-1960)** |
|  |

|  |
| --- |
| **Figure 12. Co-occurrence Graph (2000-2020)** |
|  |

|  |  |
| --- | --- |
| **Figure 13. Ego-network for ‘Die’** | |
| 1940-1960 | 2000-2020 |
|  |  |

|  |  |
| --- | --- |
| **Figure 15. Ego-network for ‘Kill’** | |
| 1940-1960 | 2000-2020 |
|  |  |

|  |  |
| --- | --- |
| **Figure 14. Ego-network for ‘Shoot’** | |
| 1940-1960 | 2000-2020 |
|  |  |

|  |
| --- |
| **Figure 15. T-distributed Stochastic Neighbor Embedding (1940-1960)** |
|  |

|  |
| --- |
| **Figure 16. T-distributed Stochastic Neighbor Embedding (2000-2020)** |
|  |

|  |  |
| --- | --- |
| **Figure 17. Gender-Violence Associations** | |
| 1940-1960 | 2000-2020 |
|  |  |

|  |  |
| --- | --- |
| **Figure 18. Incentive-Violence Associations** | |
| 1940-1960 | 2000-2020 |
|  |  |