Rhetorical Strategies and Moral Signals in Partisan Youtube Content

Andrew Follmann
afollman@andrew.cmu.edu

December 13, 2019

Abstract

How do partisan Left and Right YouTubers differ in their rhetorical style and moral signaling? Comparing a broad spectrum of video content on Youtube, we use hierarchical aggolomerative clustering, multidimensional analysis, and Keyword Analysis to evaluate the transcripts from YouTube Vlogs, and more traditional News Channels and Talk Shows. We find substantial register variation in rhetorical content across media, especially tight clustering for Vlogs but not much within-medium partisan variation. Left/Liberal and Right/Conservative moral signals are not stable over register, but overal register variation suggests Vlogs reference Sanctity and Authority less than non-vlog media. We conclude with limitations and direction for future research.

1 Introduction

YouTube is a relatively young medium for media content, and is one which is significantly more democratic and diverse than radio, TV and Cable before it. One new video format that has proliferated on the platform is the vlog style. This is a very broad topic, but it generally follows the narrator/producer (YouTuber) musing on topics of their choice. Vlogs generally are directed in content domain, with sports vloggers, or make-up vloggers, or political vloggers generally staying in their chosen lane, similar to traditional media.

There is an increasing literature, primarily in journalism, documenting the radicalization of disaffected young, primarily white, men via youtube. [4] [11].

Much of this literature addresses Youtube's algorithm, and its recommendation algorithm promoting extreme channels and videos. In addition, there was recently a trend, especially among the self professed centrists in the lead up to the 2016 presidential primaries identifying a Horsehoe theory of politics, where the Extreme Left and Extreme Right are ideologically similar to eachother, especially in comparison with centrism. [8] This has fallen our of favor [3] and has been largely discredited in terms of ideology, but it is unaddressed whether arguments for two radical ideologies address the same disaffection.

How do YouTubers on the left and right wings use rhetoric in their content? Do these groups use the same rhetorical strategy to persuade or engage a similarly identifying audience? or do these ideological extremes fall more in line with their more mainstream ideological neighbors?

In addition, considering the Moral Foundations Theory, we further analyze the morally charged vocabularly in the transcripts to understand the moral framing of the content and issues addressed. Do the moral signals in the transcripts conform to what Moral Foundations Theory would suggest?

Moral Foundations Theory is a topic of social psychology and political science, holding Liberals and Conservatives to have dfferent moral predispositions. [6] There are criticisms theory, [12] especially any identification of innate predispositions, but the difference in which moral arguments are more effective in different groups of people is empirically well founded [5]. Considering the primary moral dimensions as: Sanctity/Purity, Loyalty, Authority, Fairness and Care, Moral Foundations Theory holds that Liberals are more swayed by moral arguments evoking Fairness and Care than Conservatives, while Conservatives are relatively evenly morally charged.

2 Methods

2.1 Data

Table 1: Counts by Political Partisanship and Media Type

	Media Type	Political Leaning	Channels	Transcripts	tokens
1	News Channel	Conservative	1	295	230443
2	News Channel	Left	1	260	521516
3	News Channel	Liberal	3	546	906853
4	News Channel	Moderate	5	701	400342
5	News Channel	Right	2	300	179322
6	Talk Show	Conservative	4	904	791096
7	Talk Show	Left	1	499	933134
8	Talk Show	Liberal	7	2262	2630981
9	Talk Show	Moderate	11	2145	4095251
10	Talk Show	Right	7	1651	5110820
11	Vlog	Apolitical	26	6339	10058221
12	Vlog	Conservative	13	2907	12349577
13	Vlog	Left	55	5735	13263062
14	Vlog	Liberal	25	4892	9748671
15	Vlog	Moderate	4	576	1237985
16	Vlog	Right	47	6010	18939641
17	Total		212	36022	81396915

Our data come from YouTube transcripts, scraped from the video page using Python

and Selenium. The videos selected for analysis were identifed via the YouTube Data API (https://developers.google.com/youtube/v3) as the 300 most recent uploads for a given channel, The text data were automatically constructed using YouTube's text transcription algorithm, which is which is has about a 60-70% accuracy rating. [9]. These error rates are likely more present in poor audio quailty recordings, and in those with accents, which may bias our data.

The video channels selected for analysis were collected using a blend of heuristic methods. First, in the literature review step I found several high profile or paradigmatic youtube accounts noted, and included these [10]. The majority of the literature focused on Right wing channels, so I also found a survey of the contemporary popular leftist and liberal channels through a well sourced and documented list on the popular social media forum site Reddit [2]. Then, I used the channel neworks from these high profile accounts to select neighbor accounts, which I assumed had the same ideological dimension. These were conservative, right wing, liberal or left wing.

I also identified several traditional Talk Show and News Channel outlets, such as Fox News, ABC and CNN. The shows which air on these respective stations offer ideological comparison groups. I also found the channels of several Apolitical YouTube vlogs, ranging from sports to video-games to make-up to comedy.

Table 1 identifies the number of tokens and channels for each political orientation and content medium, and the appendix shows a more detailed breakdown by channel.

2.2 Rhetorical Content

To conduct the rhetorical analysis of the video transcripts, I used DocuScope tagged data, courtesy of Dr. David Brown, and the DocuScope developers of Dr. David Kaufer and Dr. Suguru Ishizaki. [7] The DocuScope dictionary is designed to reveal rhetorical strategy, offering a very high level of detail in analyzing the rhetorical moves occuring in the text.

I used hierarchical agglomerative clustering as well as K-means clustering to evaluate the similarity of the target evaluative groups - that is, of Political Orientation and video medium. In addition, I used Multidimensional Analysis, which further dimensionally reduces the docuscope categories into a few factors, composed of several highly correlated docuscope tags. Once factors are identified, we are able to score the salient ideology-medium groups and compare them along the scale of the factor loadings, as well as review the statistical differences of the groups according to an analysis of variance (ANOVA) test, as well as examine their R^2 component, highlighting the amount of variance each factor component explains.

2.3 Moral Signaling

In the Moral Foundations analysis we also used hierarchical agglomerative clustering to find morally near groups. In addition, we use keyness of the moral positioning to evaluate the primary differences between the different orientations, and the different mediums. This allows us to directly compare the few moral categories in view of the

Moral Foundations Hypothesis. Due to our relatively blunt instrument of dictionary lookups and our time constrained inability to collocate the moral signifier, we combined both positive and negative charges of the moral signal words to reflect any attention to a moral subject.

3 Results

3.1 Rhetorical Strategy

3.1.1 Hierarchical Agglomerative Clustering

First, using quanteda [1], we tokenize the docuscope tagged data and normalize each tag within document to reflect the percentage each tag composed of the transcript. We filtered the results to only include transcripts which were accurately tagged at a rate of 80%, and which had no single rhetorical tag as greater than 40% of the transcript, so to mitigate against outliers.

Grouping over all Channels to the Political Leaning and Type categories, we construct a dendrogram with seven clusters, as suggested by a silhouette plot (see appendix). This is a bit more than we would expect given the relatively few political ideological medium combinations.

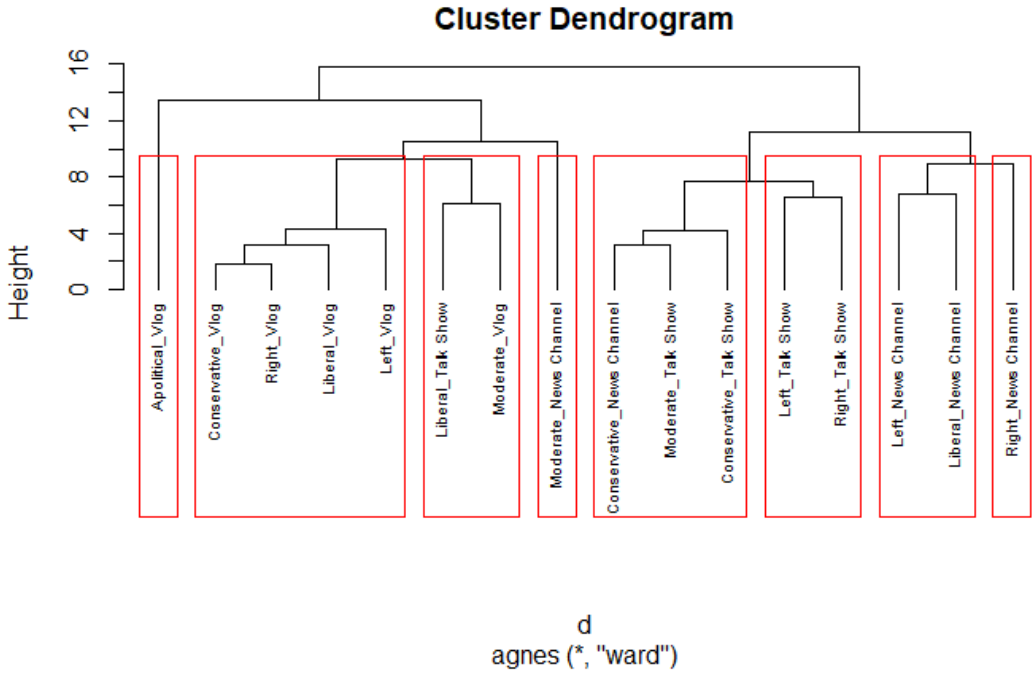

Figure 1: Docuscope Dendrogram

Our dendrogram appears to cluster in a coherent manner - Partisan Vlogs, News Channels, and Talk are all fairly near to one another. Moderate News Channel and

Apolitical Vlog however both appear to remain far from any other grouping, and suggests partisanship rhetoric is more alike between party than outside of the partisan political dimension altogether. In addition, Left/liberal and Right/Conservative vlogs are all closer than their partisan opposites. This is generally affirming of our data selection and labeling, and suggests that medium is more important than partisanship in rhetorical composition.

3.1.2 MultiDimensional Analysis

Finding an optimal number of factors at 3 (See Docuscope Scree Plot in appendix), we reduce rhetorical composition of the transcripts into collections of these variables to maximize explantory variation for channels.

	Factor1	Factor2	Factor3
AcademicTerms	0.335	0.409	0.095
Character	-0.23	-0.671	0.15
Citation	0.276	-0.069	0.158
Description	-0.074	0.061	-0.971
ForceStressed	-0.539	0.111	0.103
FirstPerson	-0.679	-0.07	-0.035
InformationExposition	0.002	0.382	0.014
InformationChange	0.193	0.235	-0.038
InformationPlace	0.403	-0.043	0.001
InformationStates	0.233	0.069	0.119
InformationTopics	0.317	0.209	0.022
Interactive	-0.577	-0.265	0.198
Metadiscourse	-0.174	0.319	0.201
Negative	0.026	-0.187	0.24
PublicTerms	0.549	-0.099	0.173
Reasoning	0.026	0.229	0.257
Strategic	0.236	-0.208	0.13
SyntacticComplexity	0.777	0.316	0.055
Uncertainty	-0.242	0.166	0.126
p.value	0	0	0
R^2	0.406	0.116	0.300

First we note the remarkably high R^2, which demonstrates that we can effectively reduce the rhetorical moves to only a few composite factors.

The first factor appears to be a polarity between detailed complex evaluation and engaged interpersonal discorse. It can account for 40.6% of the variation in the channel level data. That Apolitical Vlog is such a high outlier may explain some of the high variability explaned in this factor. Additionally, the clustering of the Vlogs and the News Channels reflects coherent register variation.

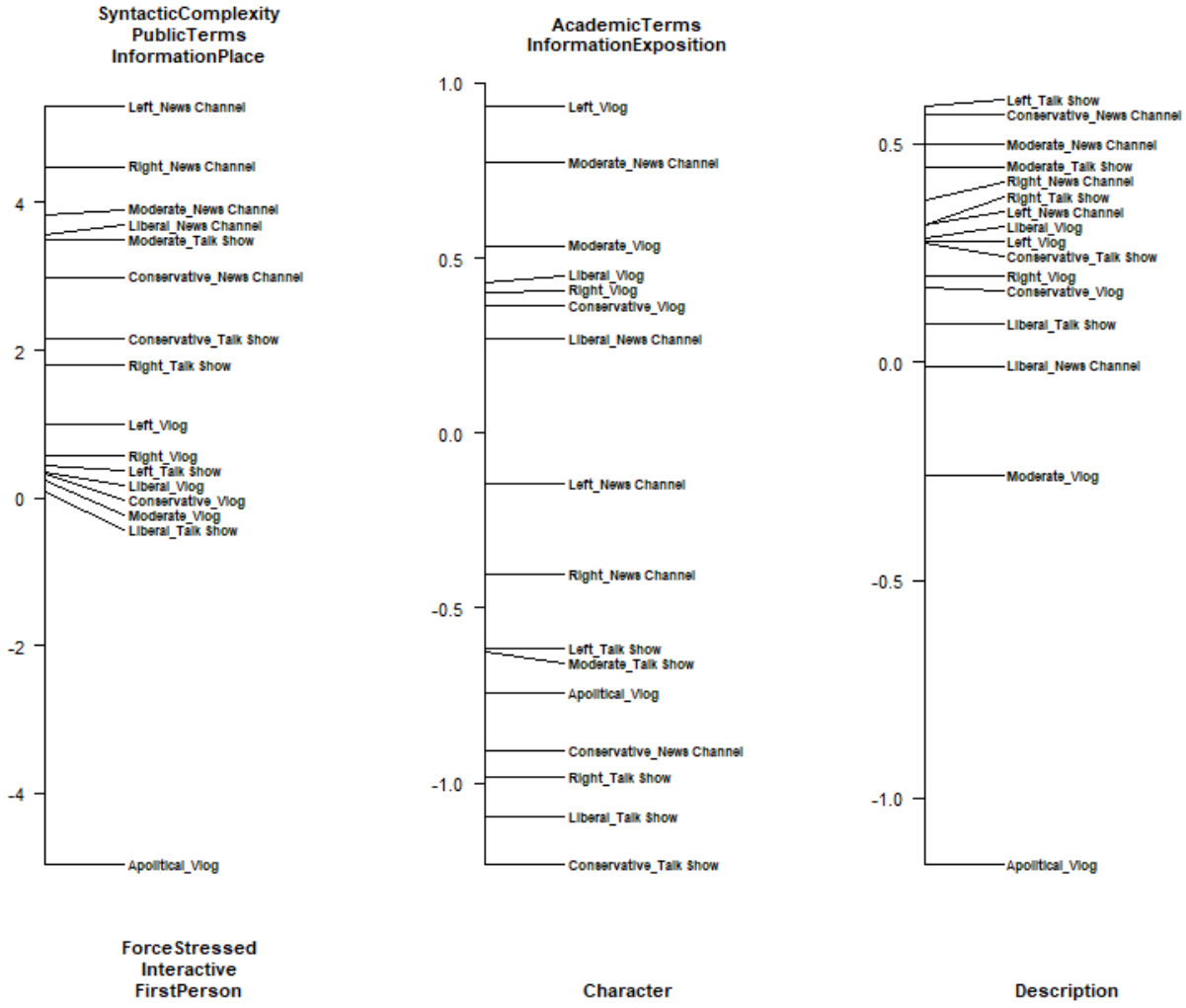

Figure 2: Factor loadings

The second factor roughly maps to a polarity between explanations of complex topics, and characterization. On its own it can explain account 11.6% of the variation in the channel level data. It is interesting that left Vlogs and Moderate News Channels are more towards exposition, against talk shows and conservative news. Again polticial vlogs are quite near to one another, though Left vlogs score a bith higher.

The Third factor is substantially marked by a strong component of Descriptive terms. On it's own it can explain 30.0% of the variation in the channel level data. Apolitical Vlogs again are the outlier, favorign descriptive terms against the the other combinations. The rest of the political media groups ar erelatively clustered, eschewing Descriptive terms, though with Moderate Vlogs a smaller negative outlier with relative preference for description. Again that Apolitical Vlog is such an outlier may be explaining the high R^2 terms.

3.2 Moral Signaling

3.2.1 Hierarchical Agglomerative Clustering

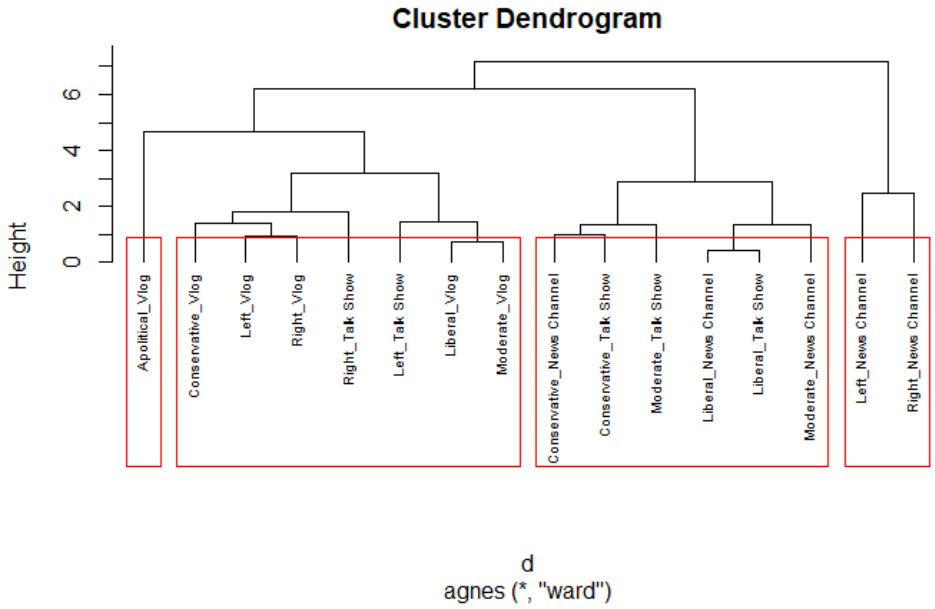

Figure 3: Dendrogram Moral Signaling

In evaluating the appropriateness of clustering methods on our Moral data, the silhouette and gap method suggest 3 clusters. See the appendix for the sihouette plot.

Our hierarchical clustering does not appear to conform to our presupposed notions on first look. There is also a less strong but still notable cluster among the Right/Left Talk Shows and Right/Left/Conservative Vlogs. as these are relatively partisan charged channels of communication we would expect more moral signaling. See appendix for dendrogram distance plot.

3.2.2 Moral Keyness

Our keyness table reveals differences between the Left-Right and LeftLib-RightCons comparisons. Considering the Vlog category we see that the Left vlogs evoke Sanctity significantly more so than those on the Right. However, when also considering Liberal/Conservative positions, we find the moral arugments made in these partisan Vlogs do not significantly differ.

When only considering non-vlogs of the Left v Right, we find substantial differences in the Left's preference for words denoting Loyalty and Care, while using Authority and Sanctity much less When also considering our more traditional domains of News Channels and Talk Shows, we find Left/Liberals to be more apt to use words related to Loyalty and Care, but not so many words related to Authority or fairness.

Media	Comparison	Statistic	sanctity	loyalty	fairness	care	authority
Vlog	Left-Right	LogRatio	0.24	-0.01	-0.05	-0.08	-0.12
Vlog	Left-Right	P-Val	0.000	0.841	0.566	0.129	0.064
Vlog	LeftLib-RightCons	LogRatio	0.00	0.06	-0.10	0.05	-0.08
Vlog	LeftLib-RightCons	P-Val	1.000	0.221	0.103	0.152	0.080
NonVlog	Left-Right	LogRatio	-0.45	0.37	0.09	0.32	-0.37
NonVlog	Left-Right	P-Val	0.010	0.011	0.655	0.008	0.002
NonVlog	LeftLib-RightCons	LogRatio	-0.18	0.20	-0.32	0.29	-0.16
NonVlog	LeftLib-RightCons	P-Val	0.077	0.022	0.003	0.000	0.016
	Vlog-NonVlog	LogRatio	1.02	-0.17	-0.11	0.43	-1.03
	Vlog-NonVlog	P-Val	0.000	0.000	0.040	0.000	0.000

Overall, comparing Vlogs and Non-vlogs we see a substantial preference for Sanctity, and Care, and fewer tokens related to Authority.

4 Discussion

YouTube content creators have stepped into a role which has substantial rhetorical and moral signaling characteristics than tranditional political news media. The Vlog format is an innovation and clearly an engaging given the growth of the platform, format and corresponding creator community.

From our hierarchical agglomerative clustering, we find that indeed there is rhetorical similarity among partisan vlogs, against non-partisan vlogs and non-vlog media.

Dimensionality reduction via factorization of the primary variables demonstrates that a large amount of the variation of the rhetorical usage can be found in three composite factors.

The factors, analgous to a preference for (1) complex evaluation against engaged interpersonal discourse, (2) complex explanations against characterization, and (3) description do reflect substantial intra medium similiarity, and inter-group difference. Apolitical Vlogs are outliers in two of these major dimensions, but the political ones remain fairly clustered.

The three media appear to be near internally consistent on these factor loadings, but within media the patterns are scattered. The primary question, seeking whether Right and Left Vlogs have similar rhetorical structure seems confirmed, but it is more a factor of the forum, the medium, than necessarily the appeals to the same latent masses.

Regarding moral signaling, we do not find significant support for the Moral Foundations theory hypothesis. The findings were inconsistent between Left vs Right and Left/Liberal vs Right/Conservative Vlogs and non-vlogs. However, Vlogs vs Non-Vlogs do seem to have different moral tokens, with vlogs reflecting on Sanctiy, and not Authority, compared to the Non-Vlog media. The reflection on Authority likely reflects the democractic and low barriers to entry to YouTube, so the less apparent need to appeal to Authority.

Our study has issues of sampling, where the YouTube channels are not necessarily a sample of the left right political spectrum, the left-right political spectrum is not nuanced enought to capture the range of political expression on YouTube, the channel labeling process was manual and subject to the author's personal bias and poor judgement, and the contemporary climate (eg. Trump, Brexit, the democractic primary, etc) may bias our results. In addition, the YouTube transcription algorithm leaves much to be desired, especially in consideration with conversation and accented speakers. Active conversation is blended together, while those with accents are frequently mis-transcribed, which likely is also biasing the sample.

A better executed study may employ third party transcription, and multiple raters of channel. However, the evaluation of the Vlog format is not well understood and well worth systematic evaluation.

References

[1] Kenneth Benoit, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. quanteda: An r package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30):774, 2018.

[2] .bshawwwwwww. Master list of left-wing youtube and podcast channels. `https://www.reddit.com/r/BreadTube/comments/9kwkxa/master_list_of_leftwing_youtube_and_podcast`, Oct 2018.

[3] Simon Choat. 'horseshoe theory' is nonsense – the far right and far left have little in common. `https://theconversation.com/horseshoe-theory-is-nonsense-the-far-right-and-far-left-have-little-in-common-77588`, Mar 2019.

[4] Conor Friedersdorf. Youtube extremism and the long tail. `https://www.theatlantic.com/politics/archive/2018/03/youtube-extremism-and-the-long-tail/555350`, Mar 2018.

[5] Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P Wojcik, and Peter H Ditto. Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology*, volume 47, pages 55–130. Elsevier, 2013.

[6] Jesse Graham, Jonathan Haidt, and Brian A Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029, 2009.

[7] David Kaufer, Cheryl Geisler, Pantelis Vlachos, Suguru Ishizaki, L Van Waes, M Leijten, and CM Neuwirth. Mining textual knowledge for writing education and research: The docuscope project. *Writing and Digital Media. Amsterdam: Elsevier*, pages 115–129, 2006.

[8] Adam Nossiter. Marine le pen may get a lift from an unlikely source: The far left. https://www.nytimes.com/2017/04/25/world/europe/france-melenchon-macron-le-pen.html, Apr 2017.

[9] University of Minnesota Duluth. Improve accessibility. http://www.d.umn.edu/itss/classroom/captioning/youtube_autocap.html.

[10] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira. Auditing radicalization pathways on youtube. *arXiv preprint arXiv:1908.08313*, 2019.

[11] Kevin Roose. The making of a youtube radical. https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html, Jun 2019.

[12] Christopher L Suhler and Patricia Churchland. Can innate, modular "foundations" explain morality? challenges for haidt's moral foundations theory. *Journal of cognitive neuroscience*, 23(9):2103–2116, 2011.

5 Appendix

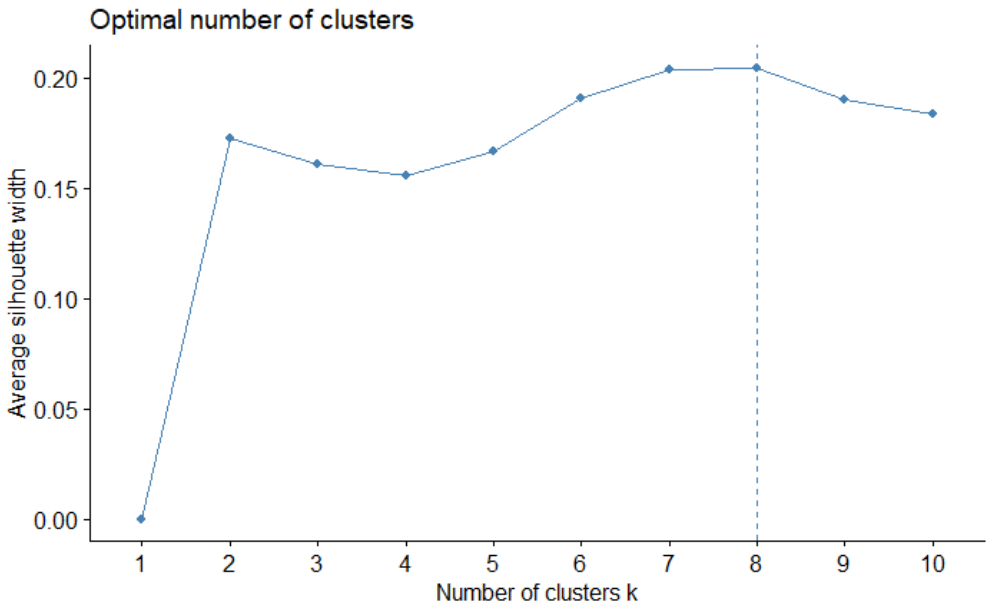

Figure 4: Silhouette Plot for Rhetorical Cateogrization of Political Variables

A distance plot shows the high similarity of the Vlogs, especially the partisan charged ones. The *Moderate_Talk Show*, *Conservative_Talk Show* and *Conservative_News Channel* also have low distance, as seen in the dendrogram. However, no other obvious low distance groups appear.

When we plot the data in a k-means, using fewer clusters to highlight inter-group difference, we see the same clustering pattern - vlogs near to each other, and news channels and talk shows relatively near to each other. The Moderate News Channel stands out in its own cluster, representing rhetorical distance from the more partisan media.

As our tagging metric is zero sum by construction, we can suppose that the rhetorical tags to be correlated among each other. We can use this knowledge to create composite feature factors which can explain some of the variance of the rhetorical usage.

Clustering via K-means demonstrates similar results.

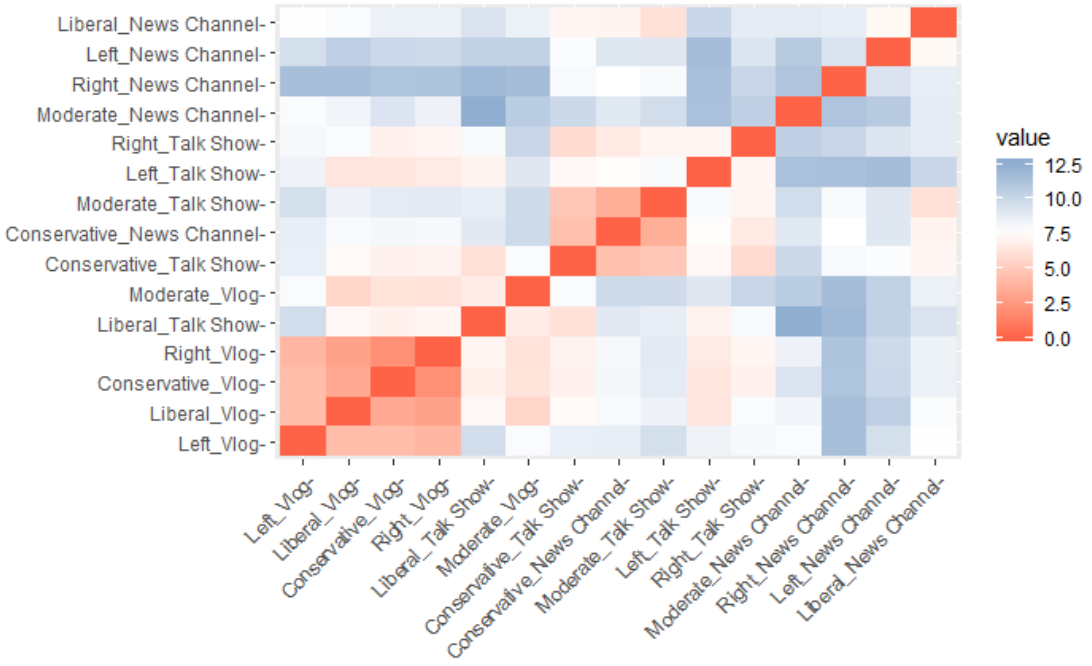

Figure 5: Docuscope Distance Gradient

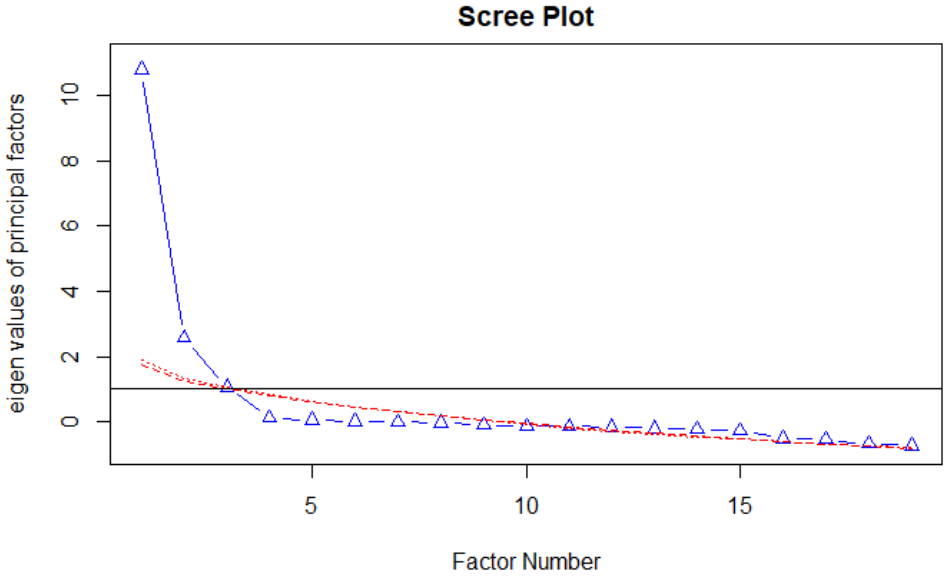

Figure 6: Docuscope Scree Plot

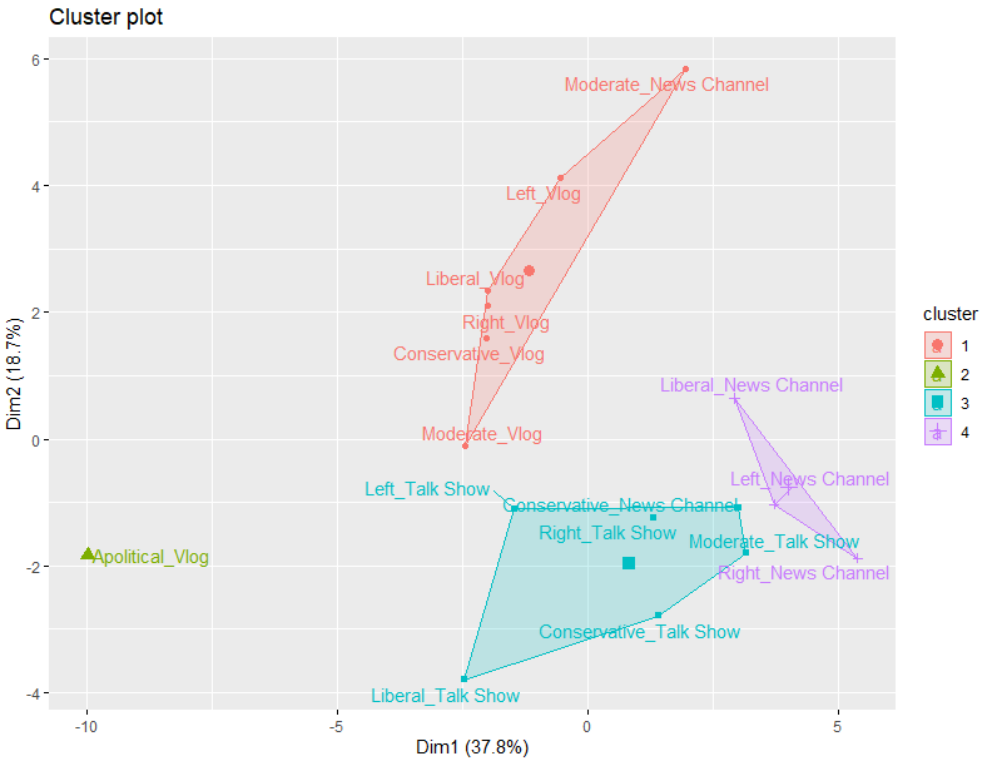

Figure 7: Docuscope k-means

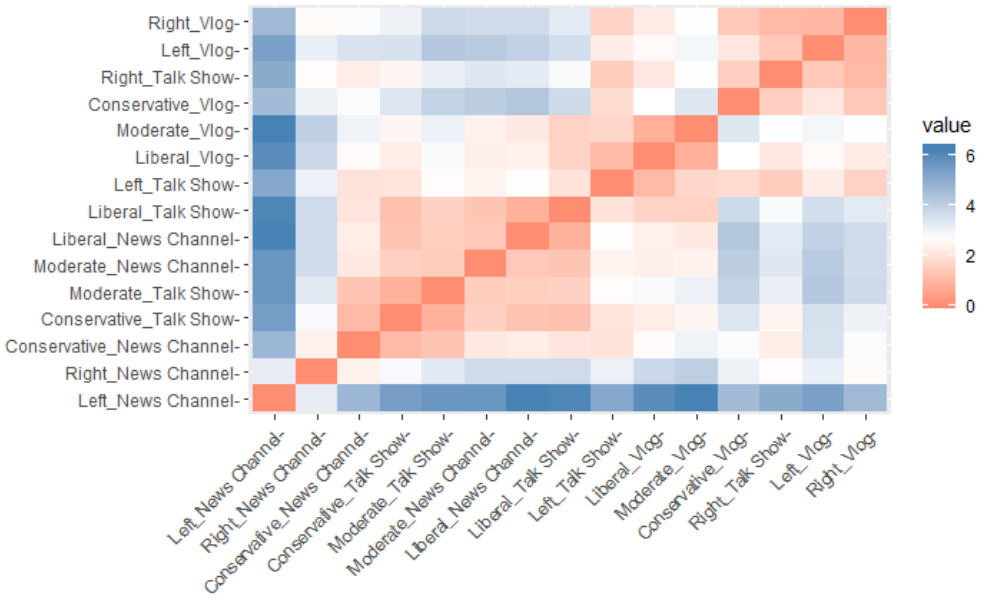

Figure 8: Factor loadings

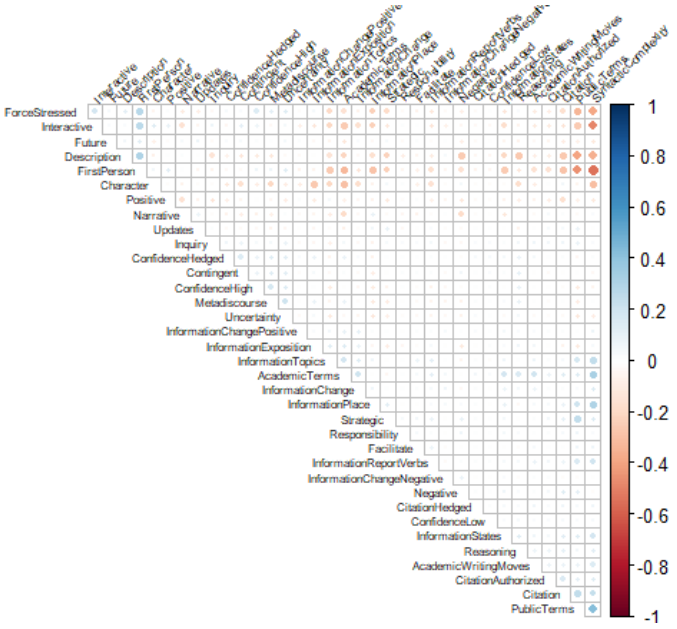

Figure 9: Docuscope Corrplot

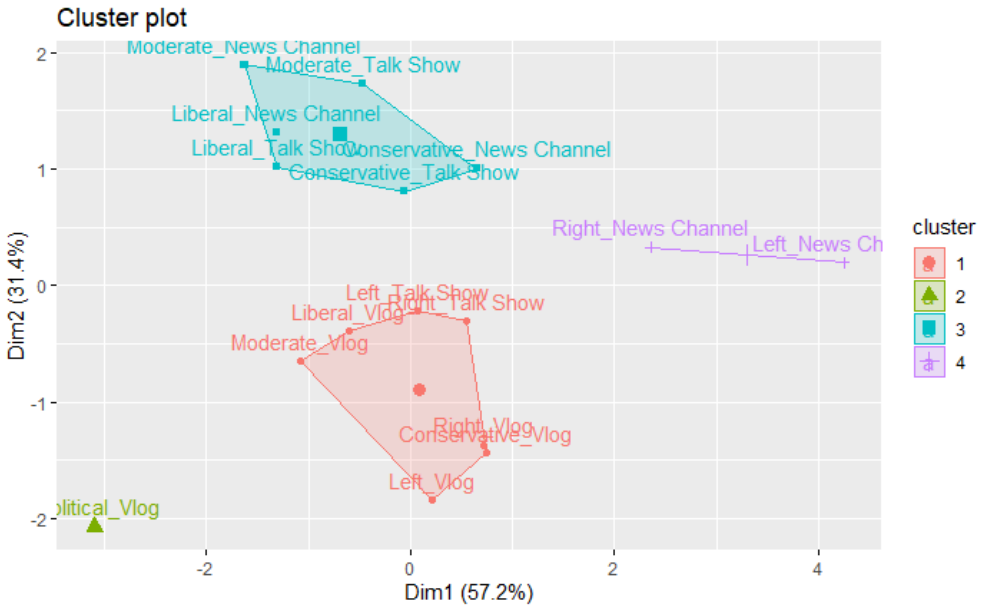

Figure 10: Factor loadings

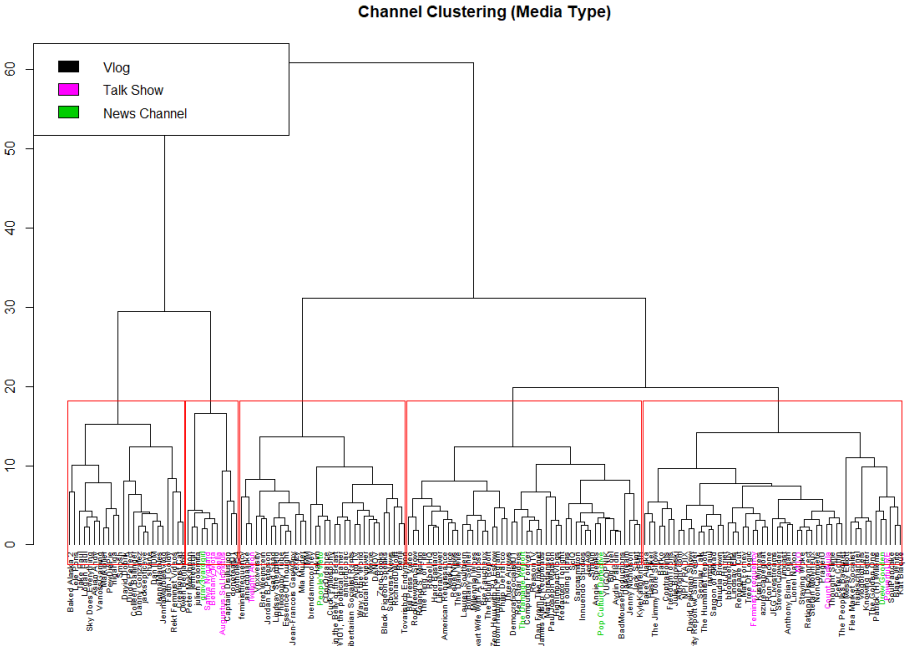

Figure 11: Dendrogram of Channels, Colored by Media Type

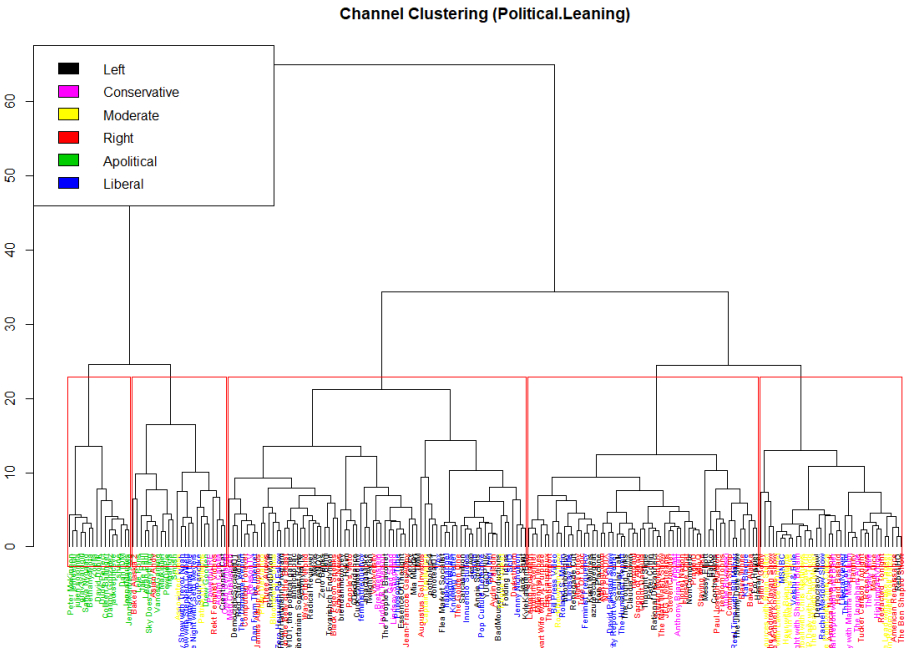

Figure 12: Dendrogram of Channels, Colored by Political Leaning

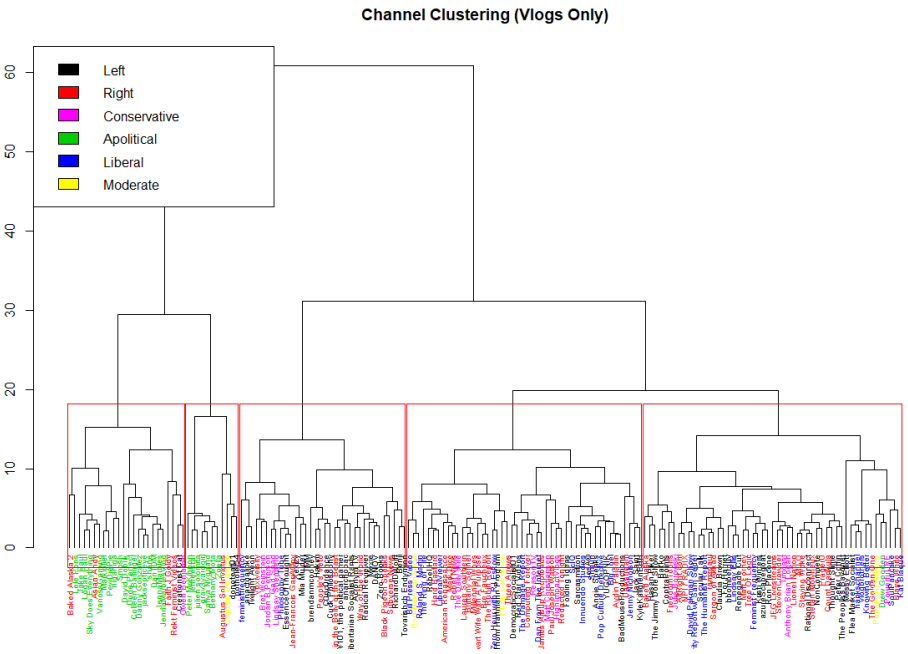

Figure 13: Dendrogram of Only Vlog Channels, Colored by Political Leaning