

# INF2209: Final project submission

Audrey Aw

December 15, 2022

## 1 Introduction

American Horror Story (AHS) is an anthology horror series that has aired on the American cable network FX since October 2011. All seasons in the series are set in the same universe, however each season is self-contained with different plots and sets of characters. For example, the first season focused on the life of a family living in a haunted house in present-day Los Angeles, while the second season told the stories of inhabitants in an asylum in 1960s Massachusetts. In this project, we are interested in investigating if there are common themes that can be found amongst all seasons despite there being different stories told each season.

## 2 Dataset

In this project, we will be using a corpus of scripts from the television show AHS. The scripts were scraped from the website Springfield! Springfield!, a database of scripts from television shows and movies (Springfield!, n.d.) The corpus contains a collection of scripts from AHS Seasons 1 to 8, with the number of episodes per season ranging from 10 to 13.

Web scraping was conducted on November 7, 2022 using the BeautifulSoup Python package. In preparation for the construction of topic models, data cleaning was carried

out with the following steps taken:

1. **Word tokenization**

This was done using the `word_tokenize` function from the Natural Language Toolkit (NLTK) library (Loper & Bird, 2002). In this step, all tokens were also converted to lowercase.

2. **Stop word removal**

Five separate sources of stop words / common words were removed from the corpus: (1) NLTK’s stop words corpus, (2) an extended list of stopwords compiled by Github user Sean Bleier (Bleier, 2010) (this was used as it was more extensive than NLTK’s stop words corpus), (3) names of main characters from the AHS wiki site (TV, n.d.), (4) stage directions (e.g. ‘GASPS’) and (5) additional words such as abbreviations of character names and exclamations (e.g. ‘viv’, ‘yeah’) that were identified after generating a frequency distribution from the results of the previous four phases of stop word removal. It was essential that the character names / proper nouns were removed, as without doing so, the topics would have comprised only of character names.

3. **Removal of non-English words** A set of English words was defined using NLTK’s English words list. Words in the dataset that were not found in the English words list were removed.

4. **Lemmatization**

Lemmatization was performed using NLTK’s WordNet lemmatizer (Loper & Bird, 2002). This was done to normalise text by reducing words to their root forms.

After the data cleaning, each episode was saved as a document. There were 94 documents in total and the mean length of each document was 755 words.

Four different topic modelling algorithms were used to computationally analyse the texts. They are 1. Latent Dirichlet allocation (LDA), 2. Biterm topic model (BTM), 3. Non-negative matrix factorisation (NMF) and 4. Correlation explanation (CorEx).

## 3 Results

### 3.1 Latent Dirichlet allocation (LDA)

LDA was implemented using the Gensim Python package (Rehurek & Sojka, 2011). To find out the optimal number of topics, model perplexity for a varying number of topics (1 - 50) was plotted against model coherence, as seen in Figure 1a.

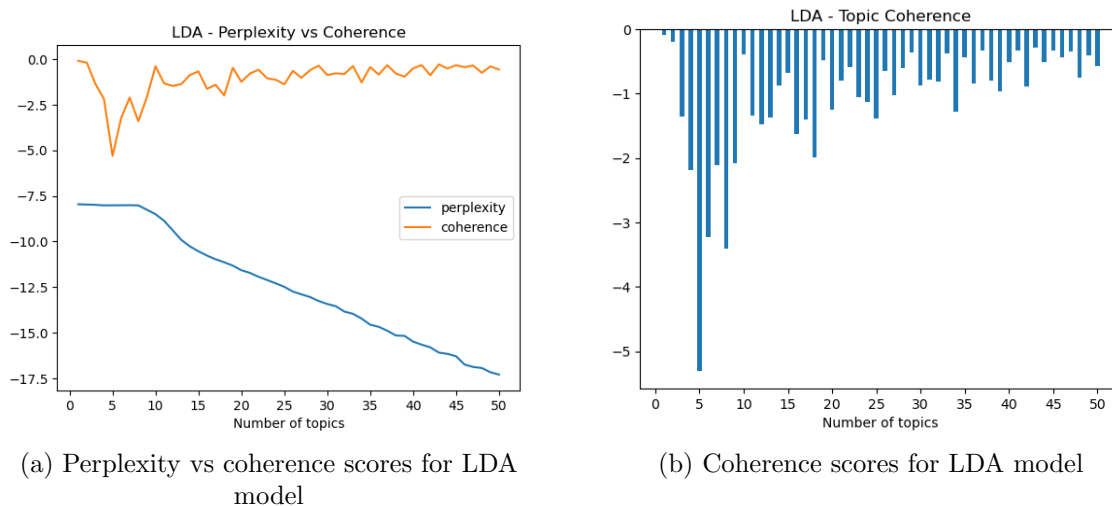


Figure 1: Graphs for LDA evaluation metrics

There was no convergence of model perplexity and coherence, as perplexity scores were unusually low while coherence scores remained somewhat constant across all numbers of topics. It is difficult to estimate the optimal number of topics as ideally, the ideal number is where perplexity and coherence converge. Although lower perplexity scores generally indicate a better-fitting model, the negative perplexity scores are indeed perplexing. We see however that when there are 5 topics, perplexity and

coherence are the Thus we look at just the coherence scores in Figure 1b to see if we can identify a good number of topics. Coherence scores were negative for all numbers of topics. This is normal for UMass coherence scores, which typically decrease when the number of topics  $k$  increase. For UMass, the ideal number of topics is when coherence is at its smallest (closest to 0). According to Figure 1b, coherence is closest to zero when  $k = 10$ . Topic words for  $k = 10$  are shown in Table 1. We see that the top 10 words per topic were almost the same for all topics, with the exception of topics 3 & 4 and were not very informative in contextualising the different topics. Moreover, based on background knowledge of the themes present in the AHS series, the topics do not appear to make much sense.

Topic	Topic words
1	good, time, god, thought, leave, told, woman, girl, family, love
2	time, good, god, love, night, thought, people, doctor, work, talk
3	snoopy, fluffer, dragon, carving, granny, bobbing, pat, decorate, elle, viol
4	sen, tout, compliment, godliness, salve, inversion, decadent, harmless, shave, oversight
5	god, good, time, doctor, love, thing, place, wife, work, girl
6	god, time, good, girl, love, night, place, thought, doctor, leave
7	time, god, people, good, love, life, thing, place, feel, kill
8	time, good, god, love, night, told, family, talk, thing, feel
9	good, time, god, leave, girl, family, thought, love, night, place
10	good, leave, god, time, love, doctor, thought, feel, talk, girl

Table 1: Top 10 words per topic for LDA model with 10 topics

Out of curiosity, we look at when coherence is at its minimum ( $k = 5$ ) to see the topic words that can be obtained. Topic words for  $k = 5$  are shown in Table 2. Somehow, these topics are slightly more understandable and differentiable as compared to topics when  $k = 10$ . Topic 2 potentially relates to Season 3 - where themes of witchcraft and struggles for power are prominent. Topic 4 is likely associated with Season 7, where the story of Valerie Solanas' radical feminist SCUM Manifesto was told. Lastly, topic 5 is potentially associated with Season 2, which follows the journey of a journalist 'Lana Banana' who found herself committed in a church-run asylum.

Topic	Topic words
1	storm, sprite, therapy, butler, stepmother, surgical, ant, collateral, harvester, searing
2	time, people, life, supreme, kill, place, good, power, witch, die
3	time, god, love, good, people, thing, feel, thought, night, life
4	people, men, turd, hail, hate, scum, whoa, ruler, pussy, pastor
5	narrator, cardinal, banana, interview, bo, freedom, verdict, weeping, wail, ye

Table 2: Top 10 words per topic for LDA model with 5 topics

### 3.2 Biterm topic model (BTM)

BTM was implemented using the Bitermplus Python package (Terpilowski, 2022). Similar to the LDA model, perplexity and coherence did not converge for all numbers of topics. Topic perplexity was extremely high, indicating that the model fared poorly when predicting words that appear in new documents. This was likely due to the fact that the ‘documents’ used for the BTM model were sentences and not the episodes, which were used when building the LDA model. This was done so that the model could actually work, which did not happen when attempting to use documents in their entirety as they were quite long at around 1000 words each.

Topic coherence followed a similar decline when number of topics was increased, as observed in Figure 1b for the LDA model. However, topic coherence was significantly lower for BTM as compared to LDA.

As it was not possible to arrive at the optimal number of topics based on perplexity and coherence score, we look at the top 10 words per topic for a 5-topic BTM model, as it was the optimal number that was used for LDA. The topics and words are listed in Table 3. These words were also rather similar to the results of the LDA model, with words like ‘time’, ‘god’, ‘life’, ‘woman’, ‘love’ recurring. Again, as the words per topic are too similar for all the topics, it is not possible to have a meaningful interpretation of these topics.

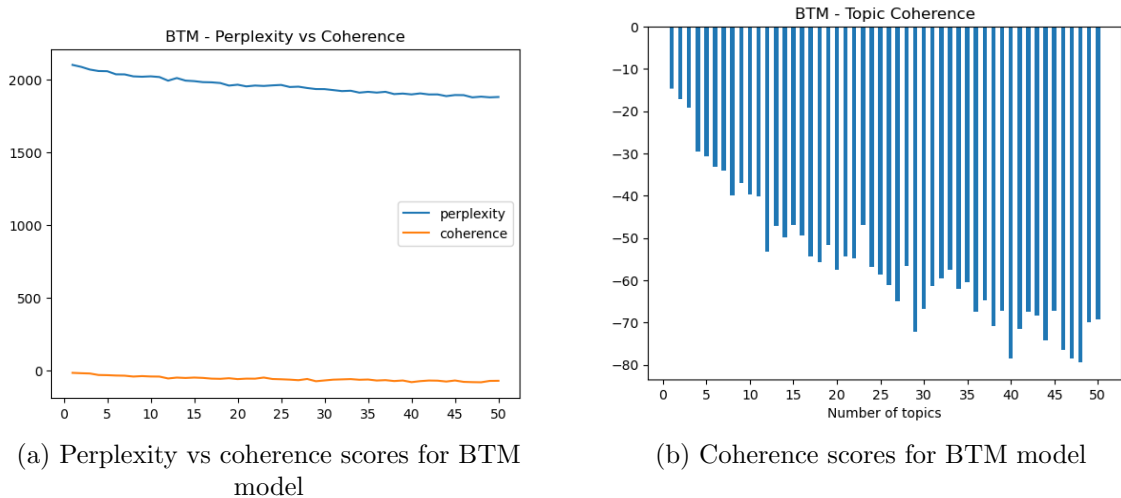


Figure 2: Graphs for BTM evaluation metrics

Topic number	Topic words
1	time, good, god, love, people, life, thing, place, night, leave
2	time, god, good, love, kill, people, thought, place, life, girl
3	time, god, people, love, good, life, place, thing, night, thought
4	time, god, good, people, love, life, thing, feel, told, night
5	time, love, good, people, thought, life, place, thing, work, real

Table 3: Top 10 words per topic for BTM model with 5 topics

### 3.3 Non-negative matrix factorisation (NMF)

NMF was implemented using the Gensim Python package (Rehurek & Sojka, 2011). Topic coherence measure that was used was the UMass coherence that was used for LDA. Perplexity was not calculated as this metric was not available to NMF models in the Gensim package, and this will be discussed further in the subsequent section on next steps for this project.

Coherence steadily decreases when the number of topics  $k$  increases. Based on 3, there is no indication on the ideal number of topics, therefore we again look at the topics when  $k = 5$  as this gives us a way to compare the quality of topic words with that of other models. The top 10 words per topic for number of topics = 5 can be observed in Table 4 below. The NMF model seems to be performing better than the

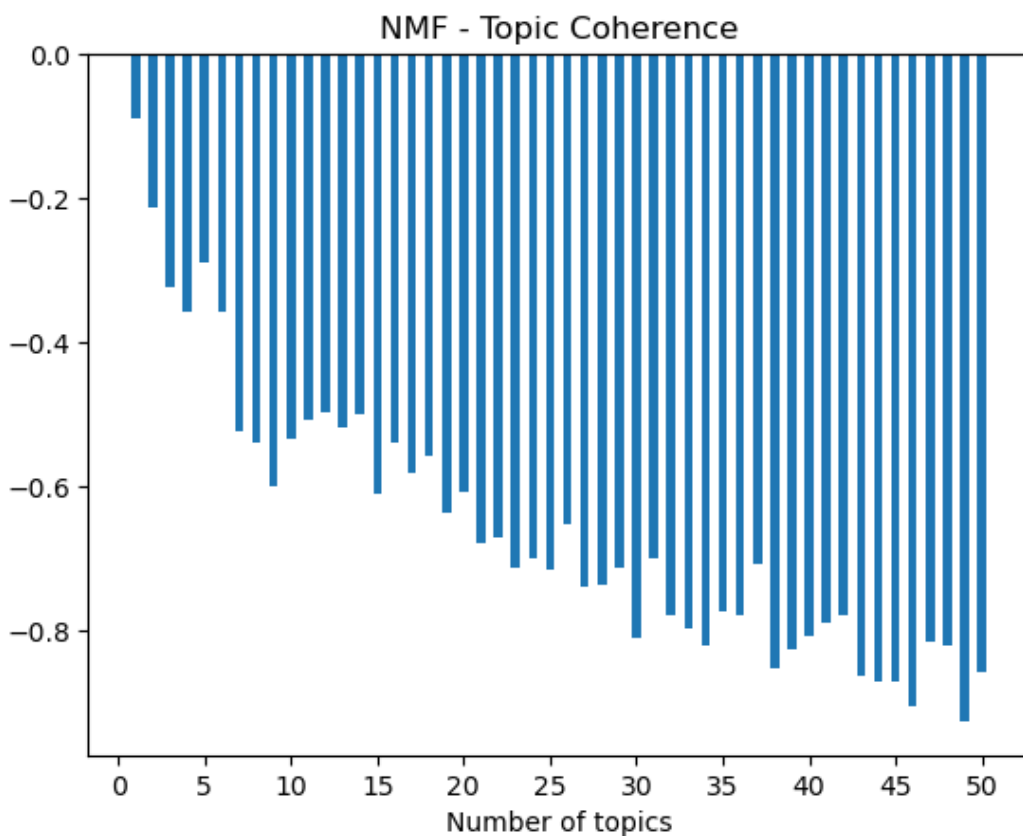


Figure 3: Coherence scores for NMF model

other models so far as the topic words are different for each topic. This is unlike what we saw from the LDA and BTM models, where almost all words for all topics were the same. Based on contextual knowledge and the words for each topic, we know that each of the 5 topics clearly corresponds to one of the AHS seasons. For example, topic 2 references season 3 (a witch coven with a leader ‘the Supreme’ that has a feud with voodoo practitioners), while topic 3 references season 7 (a cult that was formed by a self-appointed ‘divine ruler’ following the election of Donald Trump in 2016).

Topic number	Topic words	Corresponding season
1	love, time, good, place, life, feel, leave, thing, countess, room	S5 (Hotel)
2	supreme, witch, time, power, life, voodoo, magic, council, bitch, dissect	S3 (Coven)
3	ruler, people, trump, divine, time, election, turd, men, vote, kill	S7 (Cult)
4	god, moon, camera, land, blood, real, time, wait, people, truck	S6 (Roanoke)
5	monsignor, time, nun, good, doctor, god, night, thought, story, church	S2 (Asylum)

Table 4: Top 10 words per topic for NMF model with 5 topics

### 3.4 Correlation explanation (CorEx)

An unsupervised CorEx model was implemented using the CorEx Python package (Gallagher, Reing, Kale, & Ver Steeg, 2017). Overall total correlation (TC) scores were used as a measure of topic words similarity, as this was a metric that was provided as part of the CorEx package (coherence was not provided). The TC scores for the CorEx models are reflected in Figure 4. There is a positive correlation between number of topics and TC, indicating that for a greater number of topics, there will be better semantic similarity for topic words. Typically, the ideal topic number is the number where adding more topics do not contribute significantly to TC i.e. TC scores plateau. However as observed in Figure 4, TC scores steadily increase for an increasing number of topics, hence presenting difficulty in selecting the ideal number of topics.

As a result, we default to the arbitrary number of topics (5) that was used in the evaluation of the previous 3 topic models. The topic words for each topic are presented in Table 5. Although the words for each topic are different, a visual assessment tells us that the words do not share much similarity, hence it is difficult to interpret the topics.



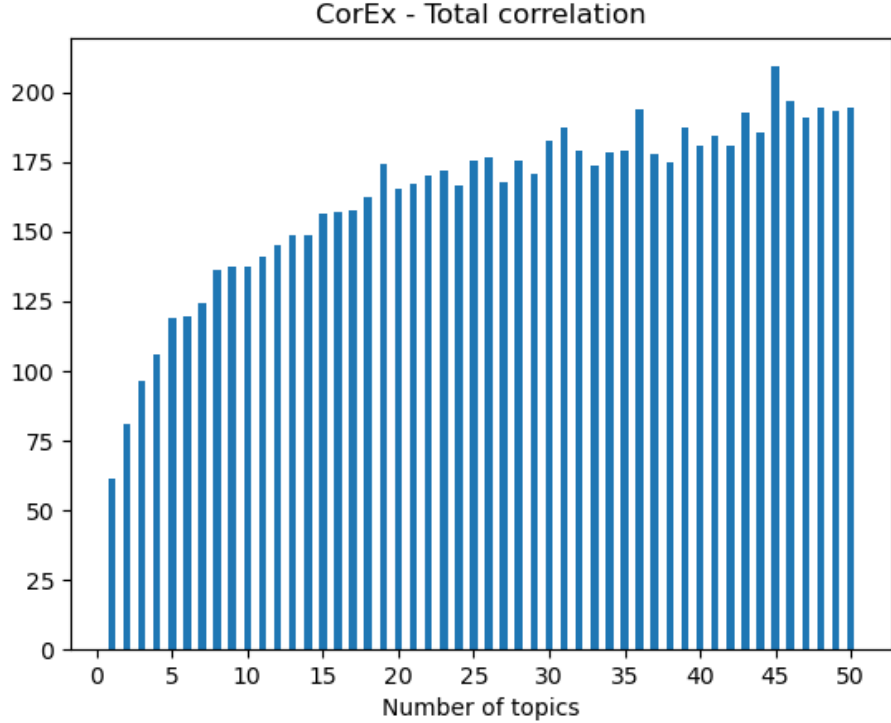


Figure 4: Total correlation scores for CorEx model

Topic number	Topic words
1	ignore, shocking, patience, check, wide, draw, filled, sealed, outrage, pick
2	champagne, gross, balance, reward, threat, marvelous, everlasting, celebrity, cake, deal
3	lead, cheating, sneak, sloppy, guide, checked, cute, hold, evil, responsibility
4	park, careful, beach, unfortunate, screen, training, interested, lesson, therapist, lawn
5	mortal, perfectly, cat, vulnerable, hand, long, leading, slipped, mask, switch

Table 5: Top 10 words per topic for CorEx model with 5 topics

### 3.5 Discussion

The results from all 4 models were poor, with NMF faring slightly better than the others. The topics from the LDA and BTM models were not sufficiently distinct as

the topic words for all topics were almost identical. NMF performed slightly better - the topic words per topic were quite logical and understandable as they corresponded to themes that were present in each of the AHS seasons. Preliminary findings from NMF suggests that it works best for the corpus on hand. Lastly, while we did not face the problem of topics not being sufficiently distinct for CorEx. Topic words were too different and did not suggest logical topics for the model. Perplexity was also not computed for NMF and CorEx, thus we are not able to fairly compare the performance of all 4 models.

One of the reasons for the poor results obtained could be due to the small dataset size. While limiting the number of passes / iterations through the corpus during training helped with improving word distinction between topics for the LDA and NMF models, there is only so much that this can help with. Additionally, as the data used in this project are transcripts from the TV show, the script includes the lyrics of songs that were aired during the episode as well. As it was not feasible to manually check the scripts individually and remove all occurrences of song lyrics, this could have affected the results of this study too.

## 4 Thematic analysis

In the second part of the study, we conducted thematic analysis to interpret and understand the topics that were generated by the topic models. In this section, we will look at the results of the NMF model, which performed the best amongst the four topics models. In the previous section of the report, we looked at an NMF model for the number of topics  $k = 5$ , which yielded 5 topics that each corresponded to a distinct season in the AHS franchise. Here, as our goal is to identify themes across seasons and not within seasons (as this is already known to us), we increased the number of topics and examined the topic words per topic. Based on inspection, 12 topics yielded a model with minimal overlap in topic words between topics, while still

maintaining a fair semantic similarity between topic words.

Table 6 below details the 5 topics produced by the NMF model, along with the top 20 topic words and top 5 documents. Documents are labeled through their season and episode numbers, as the script from each episode is a single document.

Topic number	Topic	Top 20 topic words	Top 5 documents (season & episode number)
1	Mortality & vampirism	countess, love, room, time, blood, feel, place, good, drink, son, thing, kill, shalt, long, life, nick, leave, thought, boy, die	S2E11, S3E7, S5E6, S5E11, S7E9
2	Occult & (feminine) power	supreme, witch, time, madame, life, love, girl, voodoo, bitch, power, council, god, daughter, slave, shake, kill, stake, die, good, cora	S3E4, S3E8, S3E13, S7E11, S8E7
3	Politics	trump, election, people, vote, fear, police, city, pedro, ave, pussy, president, time, council, drop, wall, pinky, god, work, feel, gun	S7E1, S7E2, S7E4, S7E5, S7E6
4	Territory	god, camera, moon, real, people, blood, wait, whoa, teeth, season, stay, rub, truck, land, good, time, listen, crowbar, open, slow	S6E5, S6E6, S6E7, S6E8, S6E9
5	Ecclesiastical & medical	monsignor, time, nun, doctor, good, god, night, thought, church, therapy, story, storm, patient, place, hear, work, told, cardinal, betty, book	S2E2, S2E4, S2E7, S2E12, S2E13
6	Family	time, good, love, pregnant, door, boston, leave, crazy, family, talk, wife, thought, god, child, dad, honey, girl, thing, hurt, feel	S1E5, S1E8, S2E11, S4E11, S6E8

7	Show business	time, good, love, clown, life, tent, fame, lady, penny, trick, hate, night, star, boy, leave, place, people, pay, true, hear	S4E3, S4E4, S4E9, S4E13, S5E7
8	Satanism & occult	hail, spell, antichrist, time, dude, people, magic, witch, conductus, place, powerful, bring, memoria, outpost, soul, ready, nuclear, voodoo, evil, energy	S8E3, S8E6, S8E8, S8E9, S8E10
9	Settlement & absence	cricket, land, thy, time, knew, colony, blood, moon, thee, hath, amy, daughter, lost, inland, understand, wait, consecrate, missing, sir, save	S6E1, S6E3, S6E4, S6E5, S6E9
10	Masculinity, power & conflict	ruler, divine, turd, pastor, men, senator, son, people, councilman, kill, brother, movement, time, scum, army, loyal, woman, family, hate, told	S7E7, S7E8, S7E9, S7E11, S8E3
11	Sorcery & science	supreme, dissect, gypsy, frog, test, power, warlock, live, dear, time, alpha, vision, level, child, magic, brought, identify, lightning, tag, school	S3E8, SEE13, S8E4, S8E5, S8E7
12	Shelter & food	sanctuary, nana, people, time, survive, shelter, overrun, wait, outpost, remember, blast, eat, fruit, room, nuclear, humanity, woman, interview, bisque, eating	S5E8, S5E9, S8E1, S8E2, S8E3

Table 6: Top 20 topic words and top 5 documents for NMF model with 5 topics

Despite the small number of topics obtained from the NMF model, we can still observe some common themes amongst the topics. The next sections examine the themes and topics in greater detail.

## 4.1 Supernatural

The first major theme in AHS is that of the supernatural. Topics 1 (Mortality & vampirism), 2 (Occult & power) and 11 (Sorcery & science) fall under this theme. We can also further group topics 2 and 11 into a subtheme of ‘Magic’.

The top 5 documents for topic 1 come from a few different seasons - seasons 2, 3, 5 and 7. Topic 1 discusses the topic of mortality that that is likely to relate to vampirism, as indicated by words such as ‘blood’, ‘drink’, ‘kill’ and ‘die’. Vampirism was discussed in S5E6 and S5E11, However, this topic appears to be rather broad as it also includes other words such as ‘love’ and ‘good’, which appears in many topics as observed in previous sections of the report. Although the major concept of AHS is that it is based around horror, it is interesting that there are topic words with positive connotations.

Topic 2 on the occult & (feminine) power are primarily associated with documents from seasons 3, 7 and 8. Words such as ‘council’, ‘slave’ and ‘supreme’ hint to a hierarchy within the world of the occult. This was confirmed by reading the top documents for this topic - the most powerful witch is addressed as the ‘Supreme’, and consults a ‘council’ to run a witch coven. On the other hand, ‘slave’ relates to a subplot within the season, which tells the story of a real-life socialite Delphine LaLaurie, who owned, abused and killed slaves in New Orleans during the 1830s (History.com, 2020). This topic may also be interpreted as associated with femininity and this is evidenced through words such as ‘madame’, ‘girl’, ‘daughter’. When reviewing the documents from seasons 3 (which comprise 3 out of the top 5 documents for this topic), a subject that recurs is the relationships between women e.g. mother and daughter, the ‘madame’ and her slaves, and the power disparity / struggle between the two parties.

Although we remain within the realm of magic for topic 11, the scope of this topic expanded to science and its relations to sorcery, observed from words such as ‘school’, ‘dissect’, ‘frog’ and ‘lightning’. The top documents from this season are episodes from

seasons 3 and 8. This is unsurprising as season 8 is considered to be a sequel of season 3 but with the addition of masculine subjects. For example, words like ‘warlock’ and ‘alpha’ point towards a hierarchy amongst men and echoes the presence of a social order.

## 4.2 Real-world

At times, horror is rooted in themes that draw inspiration from real life, which makes it all the more scarier. While the previous theme discussed the supernatural, this theme is more generalized and contains topics that relate to happenings in real life such as politics (topic 3), family (topic 6), show business (topic 7) and masculinity, power & conflict (topic 10).

All the top documents from topic 3 on politics are from season 7, which centers around the formation of a cult and its association with politics. Donald Trump was referenced heavily in the first few episodes of the season as he inspired the formation of the cult that was the center of the plot. A chant ‘We are the wall!’ was used by cult supporters during a political rally, which was highly reminiscent of Trump’s infamous wall as they both seek to divide groups of citizens in America. Words such as ‘gun’ were found in this topic, which is another contentious issue in American politics.

Additionally, topic 10 (masculinity, power & conflict) is a topic with strong links to season 7, with 4 out of 5 top documents belonging to this season. While topic 11 above discussed masculinity in the magical realm, topic 10 discusses masculinity in the context of conflict and power. The words in this topic mostly carry negative valence and suggest an association between masculinity (e.g. ‘brother’, ‘men’, ‘son’, ‘loyal’) and violent conflict (e.g. ‘kill’, ‘scum’, ‘hate’).

Topic 6 discusses the theme of family as suggested by kinship and endearment terms such as ‘dad’, ‘child’, ‘wife’, ‘girl’ and ‘honey’. This topic is unusual as most of the topic words carry either neutral or positive connotations, unlike the discussed previously that were linked to the macabre. Documents from this topic can be traced

to seasons 1, 2, 4 and 6, indicating that the topic of family is integral in the AHS series.

Majority of top documents from topic 7 (show business) can be attributed to season 4 ‘Freak Show’ - which tells the stories of freak show performers during the 1950s in Florida. It was not surprising that 4 out of 5 top documents were traced back to this season, as the plots from this season were very different from the other seasons. Similar to topic 6, most words from this season were neutral / not negative apart from ‘clown’, which has long been associated with the genre of horror.

### 4.3 Religion

The theme of religion may be broad and can be found in many topics - the word ‘god’ is present in many topics, however the two topics that were categorized under these theme had significantly greater associations with the church (topic 5) or anti-church (topic 8).

All of topic 5’s (ecclesiastical & medical) top documents originate from season 2, which was set in an asylum operated by a church - explaining the presence of words relating to positions in the church (e.g. ‘monsignor’, ‘cardinal’, ‘nun’). In the asylum, ‘patients’ were often subjected to psychiatric treatments such as the lobotomy, hydrotherapy, conversion therapy and shock treatment, all of which are unthinkable and horrifying to us in the 21st century.

In contrast, topic 8 (satanism & occult) directly contrasts topic 5, similar to how the ‘Antichrist’ being the symbol of opposition to Christ, a venerated figure in the church. Similar to Satanism, the occult has also been strongly denounced and feared by the church over the years - corroborated words such as ‘evil’. Documents from this topic all originate from season 8 (‘Apocalypse’), where there was a nuclear incident that brought about the apocalypse and was initiated by a character representing the Antichrist.

It is curious how although there are two polarizing views to religion, horror may

be found in both camps, and prompts the following thought - there are numerous ways that religion can be interpreted and they cannot be strictly categorized into the black and white or ‘good’ and ‘evil’, rather, it is one’s interpretation that gives religion its meaning and virtue.

## 4.4 Survival

The last major theme comprises topics that are related to survival, be it happenings on a particular territory (topic 4), settling in a place (topic 9) or finding shelter and food (topic 12).

On first glance, topic 4 does not immediately appear to be linked to survival on a particular territory. However, all of the top documents from this topic can be traced to season 6, which was set in North Carolina and on the alleged site of disappearance of the Lost Colony of Roanoke (History, 2022). There were impending signs of doom for characters in this season that were conveyed through objects such as ‘teeth’ and a phenomenon known as the ‘blood’ ‘moon’. Furthermore, characters attempted to survive attempts on their lives using tools such as a ‘crowbar’ and a ‘truck’ to escape from the ‘land’ that they were trapped on. Words from topic 9 directly reference the Roanoke colony’s settlement on a territory ‘inland’ and attempt to convey the colony’s experiences through flashbacks and the use of archaic language such as the word ‘thy’. The plot for the season centers around the investigation behind the disappearance of the lost colony and parallels the disappearance of the ‘daughter’ of one of the characters, who went ‘missing’.

The last topic of survival is topic 12 (shelter & food). Top documents from this topic come from seasons 5 and 8. This topic is somewhat related to topic 8 which approached the apocalypse from the angle of the harbinger, however topic 12 is centered around the survivors of the apocalypse. These people were sheltered in a ‘sanctuary’ following a ‘nuclear’ ‘blast’, and they ate food such as ‘fruit’ and occasionally ‘bisque’. This is another reminder that there are always two sides to a



coin - where there is destruction, there is also survival.

## 4.5 Schematic diagram

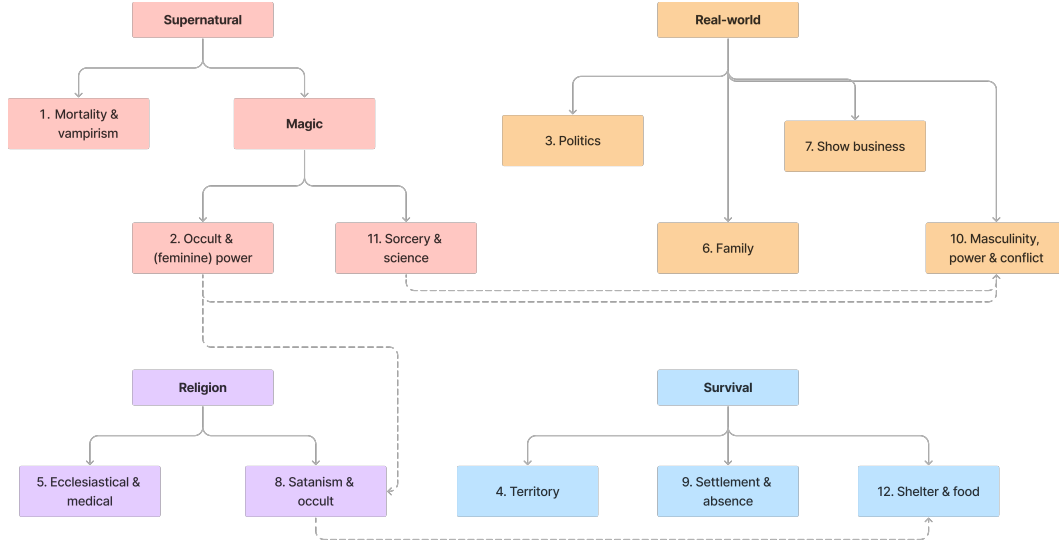


Figure 5: Schematic diagram of themes and topics in AHS

Figure 5 is a schematic diagram that models the relationships between themes and topics. Dotted lines connecting two topics indicate that the topics are loosely related.

## 5 Conclusion & next steps

In this paper, several topic models were utilized to help us identify latent topics in the AHS TV series. Although the results of the topic models were not ideal due to several challenges discussed in earlier sections, a qualitative thematic analysis complemented the results of the topic modelling and provided us with a richer understanding of topic modelling results. For example, we uncovered how some topics opposed each other, such as masculinity vs. femininity and how power is interpreted for both sides, as well as religion and anti-religion.

There are several steps that may be taken next in order to improve on the work from this study. The results from this preliminary application of topic models on the AHS scripts suggests that manual data cleaning to remove song lyrics might be required in order to achieve a better performance for the topic models.

Additionally, a perplexity metric (or another metric measuring dissimilarity) should be calculated for the NMF and CorEx models so that the performance of all models may be assessed fairly. CorEx should be assessed for coherence as well, as this is the similarity measure that was computed for the other 3 models.

Lastly, we may also look into anchoring to support semi-supervised learning for the CorEx model. The use of anchor words as guidance for the topic model is what differentiates CorEx from the other unsupervised models (LDA, BTM, NMF). Therefore, anchoring should be explored as this might help us to achieve a better result for CorEx.

## References

- Bleier, S. (2010, August). *NLTK's list of english stopwords*. Retrieved 2022-11-07, from <https://gist.github.com/sebleier/554280>
- Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017, December). Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge. *Transactions of the Association for Computational Linguistics*, 5, 529–542. Retrieved 2022-11-21, from <https://direct.mit.edu/tac1/article/43415> doi: 10.1162/tac1a-00078
- History, S. (2022, August). *What Happened to the 'Lost Colony' of Roanoke?* Retrieved 2022-12-15, from <https://www.history.com/news/what-happened-to-the-lost-colony-of-roanoke>
- History.com, E. (2020, April). *A torture chamber is uncovered by arson*. Retrieved 2022-12-15, from <https://www.history.com/this-day-in-history/a-torture-chamber-is-uncovered-by-arson>
- Loper, E., & Bird, S. (2002). Nltk: The natural language toolkit. *CoRR*, cs.CL/0205028. Retrieved from <http://dblp.uni-trier.de/db/journals/corr/corr0205.html#cs-CL-0205028>
- Rehurek, R., & Sojka, P. (2011). Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).
- Springfield!, S. (n.d.). *American Horror Story Episode Scripts*. Retrieved 2022-11-07, from [https://www.springfieldspringfield.co.uk/episode\\_scripts.php?tv-show=american-horror-story](https://www.springfieldspringfield.co.uk/episode_scripts.php?tv-show=american-horror-story)
- Terpilowski, M. (2022). bitermplus. Retrieved 2022-12-15, from [https://bitermplus.readthedocs.io/\\_/downloads/en/stable/pdf/](https://bitermplus.readthedocs.io/_/downloads/en/stable/pdf/)
- TV, F. (n.d.). *Characters*. Retrieved 2022-11-07, from <https://americanhorrorstory.fandom.com/wiki/Category:Characters>