SAS® GLOBAL FORUM 2018

USERS PROGRAM

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF

Impact of leaks on TV Shows

Presenter

Prathap Maniyur, Engagement Manager, Fractal Analytics

Prathap has more than nine years of experience of using strategic frameworks and analytic methods in diverse industries such as consumer goods, consumer durables, information technology and lighting with successful track record of leading high level initiatives in multiple channels. Prathap has not only designed the marketing strategies for organizations of different sizes and have executed them successfully in chains such as Walmart, Big Bazaar and Croma. He is also an analytics consulting leader building solutions for CPG and Retail clients.

https://www.linkedin.com/in/prathaprao/

USERS PROGRAM

SAS® **GLOBAL FORUM** 2018

Impact of Leaks on TV Shows

Impact of Leaks on TV Shows

Prathap Maniyur, Prashant Nayak, Mansi Bhat

ABSTRACT

- Typically, TV show and movie leaks have a crippling effect on the studios and might put the seller at a financial risk. Apart from this, there is a risk of endangering the corporate image and reputation of the company. However Game of Thrones managed to generate a positive word of mouth thanks to the leaks and got nonviewers to start their HBO viewership.
- This study helps studios and entertainment industries understand the impact on the viewers and the key concerns the viewers have about the leaks especially from the study of Game of Thrones Leaks.

METHODS

Research was done to select forums, websites, and social media groups where people post reviews about GOT episodes. The below forums and websites were chosen because of the high popularity amongst users.

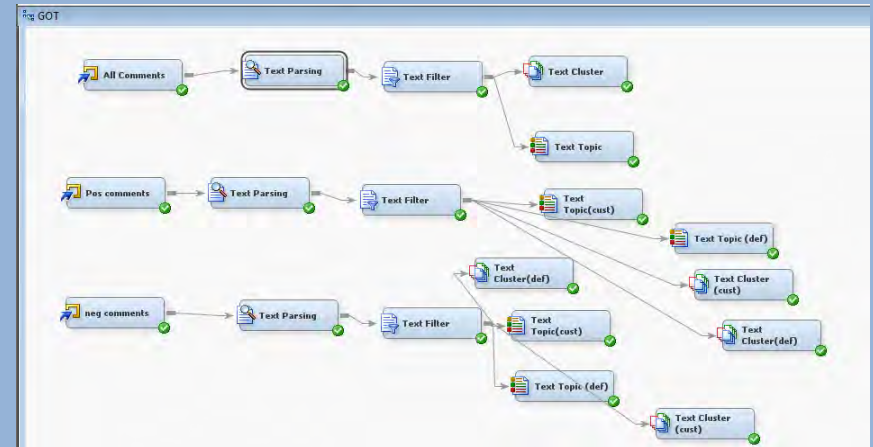Forums: http://asoiaf.westeros.org/index.php?/forum/31-game-of-thrones-the-hbo-tv-series/

https://winteriscoming.net/

https://www.facebook.com/: Gave insights on the big influencers and how the network of consumers is formed.

https://twitter.com/: Used to understand the sentiments of the people live as the news breaks out.

https://reddit.com/: Used to understand the sentiments mainly post airing of the episodes

Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms
1	0.055	0.019	+episode,+leak,last,+watch,series	57
2	0.054	0.020	tyrion,+betray,daenerys,varys,cersei	84
3	0.059	0.019	sansa,arya,littlefinger,lf,brienne	80
4	0.056	0.018	gt,+day,drake,+wall,+story	23
5	0.056	0.019	+know,+fake,+point,+talk,+implication	41
6	0.058	0.020	jon,dany,aegon,snow,+feel	78
7	0.053	0.020	+dragon,ice,fire,+wall,+ice dragon	96
8	0.050	0.020	+show,+book,+point,+watch,+character	110
9	0.053	0.020	cersei,+point,dany,+army,jaime	118
10	0.044	0.018	oc,cretu,aris,cyoa,charlie	32

RESULTS

Impact of Leaks on TV Shows

Prathap Maniyur, Prashant Nayak, Mansi Bhat

RESULTS CONTINUED (CLICK TO EDIT)

Text Cluster	Percentage	Interpretation
+leak +early +happen +watch +hack +gots7e4 +post wow	26%	This cluster is a group of terms which indicates happiness with the leaks and ability get to watch early the episode
+look +spoiler +know +happen +big help +hacker +gonna read link want	20%	This cluster is a group of terms thanking the leak and stating it was good to watch it early
+bad +ruin +start bad hbo gameofthrones gots7e7 +wait	16%	This cluster is a group of terms with negative review cluster with people expressing disappointment
+dragon +arya mother holy Sunday epic +god +throne +cersei	15%	This cluster is a group of terms about the characters

CONCLUSIONS

The social media gives us a good glimpse on why the viewership increased overall in spite of the leaks. Concept linkage shows a strong association between the leaks and the characters being discussed and people wishing that they can't wait to see the series. Overall there was a lot of positive sentiment around leaked content characters especially by the GOT fans and those who have paid for the HBO channel. It was one of the most pirated content in terms of discussing it on social media before the episode is aired on HBO. Thankfully the Game of Thrones brand was strong and the leaks had rolled over and a lot of people might have rewatched by subscribing to the network.

Test Result | Graphical Result |

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [No Feature Ranking]:
Overall precision: 85.42%
Positive precision: 63.16%
Negative precision: 100.00%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Risk Ratio]:
Overall precision: 77.08%
Positive precision: 57.89%
Negative precision: 89.56%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Chi Square]:
Overall precision: 89.58%
Positive precision: 73.68%
Negative precision: 100.00%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Information Gain]:
Overall precision: 89.58%
Positive precision: 73.68%
Negative precision: 100.00%

BEST MODEL is Smoothed Relative Frequency and Chi Square

REFERENCES

http://ew.com/tv/2017/07/16/game-of-thrones-cast-slam-leakers/
https://moviepilot.com/p/do-game-of-thrones-leaks-cost-hbo-money/4352251
http://wikiofthrones.com/11342/game-thrones-season-7-episode-6-leaked-online/
https://www.unilad.co.uk/tv/the-entire-plot-of-game-of-thrones-season-7-has-just-been-leaked/
http://wikiofthrones.com/11342/game-thrones-season-7-episode-6-leaked-online/

#SASGF

SAS®
GLOBAL
FORUM
2018

April 8 – 11 | Denver, CO
Colorado Convention Center

An Analytics Approach for Impact of Leaks on TV Shows

Prathap Maniyur, Engagement Manager, CPG & Retail, Fractal Analytics

SAS.**GLOBAL**FORUM

Typically, TV show and movie leaks have a crippling effect on the studios and might put the seller at a financial risk. Apart from this, there is a risk of endangering the corporate image and reputation of the company. However Game of Thrones managed to generate a positive word of mouth thanks to the leaks and got nonviewers to start their HBO viewership.

2014 cyber attack on Sony Pictures entertainment nearly pulled the studio under. HBO has had a good share of leaks headaches in 2017. The HBO hackers have been notoriously leaking the unaired episodes and are continuously releasing content such as Game of Thrones season 7 episodes, Curb your Enthusiasm comedy series and Insecure's Sunday night episode.
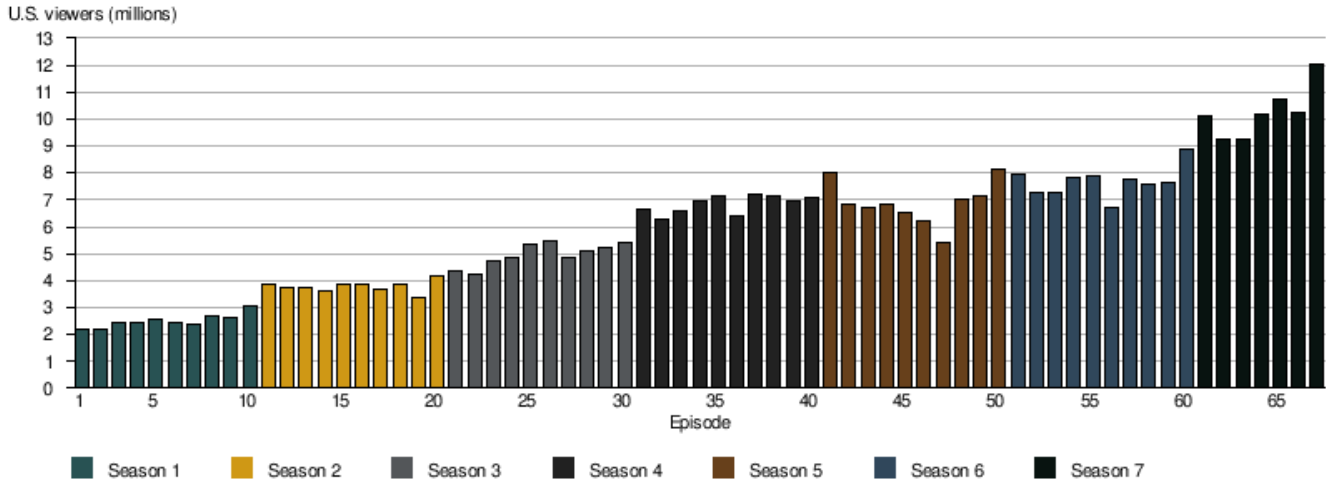
The recent ones in our memories are the Game of Thrones leaks. Two Game of Thrones episodes has been pirated before the premiere in season 7. Interestingly it might be the case where the low-resolution leaks have helped HBO to generate buzz and generated discussion about the characters which in fact aided non Game of Thornes fans to subscribe to HBO which has resulted in high viewership probably thanks to the leaks. The leaks have only fueled the TV shows unprecedented star power.

Thus rather than reduced viewership Episode 4 which was leaked managed to break the records with 10.2 M viewership.

This study helps studios and entertainment industries understand the impact on the viewers and the key concerns the viewers have about the leaks especially from the study of Game of Thrones Leaks.

Game of Thrones: U.S. viewers per episode (millions)

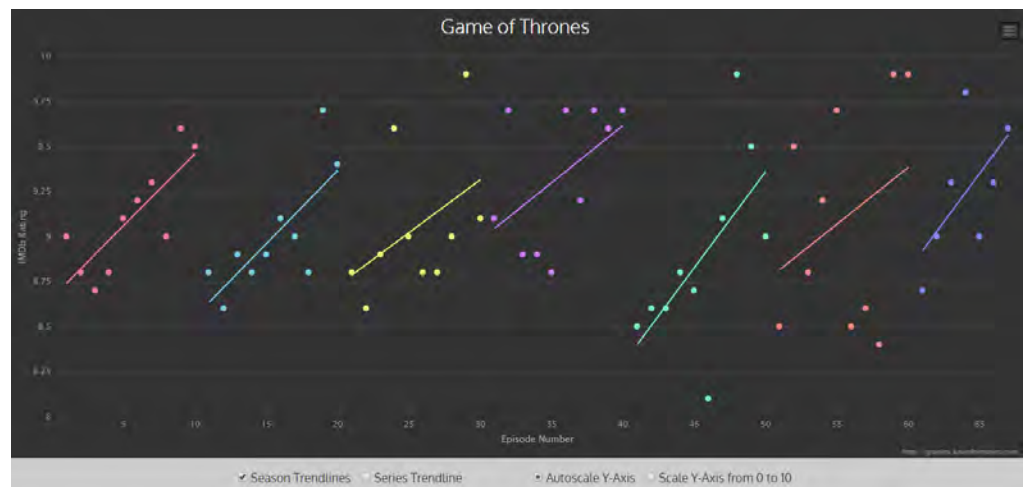| Season	Ep. 1	Ep. 2	Ep. 3	Ep. 4	Ep. 5	Ep. 6	Ep. 7	Ep. 8	Ep. 9	Ep. 10	Average
1	2.22	2.20	2.44	2.45	2.58	2.44	2.40	2.72	2.66	3.04	2.52[1]
2	3.86	3.76	3.77	3.65	3.90	3.88	3.69	3.86	3.38	4.20	3.80[1]
3	4.37	4.27	4.72	4.87	5.35	5.50	4.84	5.13	5.22	5.39	4.97[2]
4	6.64	6.31	6.59	6.95	7.16	6.40	7.20	7.17	6.95	7.09	6.84[3]
5	8.00	6.81	6.71	6.82	6.56	6.24	5.40	7.01	7.14	8.11	6.88[4]
6	7.94	7.29	7.28	7.82	7.89	6.71	7.80	7.60	7.66	8.89	7.69[5]
7	10.11	9.27	9.25	10.17	10.72	10.24	12.07	N/A			10.26[6]

Source: Nielsen Media Research[8]

Game of Thrones Leak

HBO has struggled multiple times with *Game of Thrones* leaks in the past. The first four episodes of season 5 leaked online, thanks to advance screeners that are sent out to members of the press. An HBO subsidiary also released a season 6 episode a day early. HBO decided to scrap screeners for season 7, but that hasn't prevented two major leaks this season. Episode 4 The Spoils of War leak was attributed to the employees of a vendor working for HBO Partner Sky India. Episode 6 Beyond the Wall was accidentally published by HBO's European services Nordic in Spain for an hour before removed.
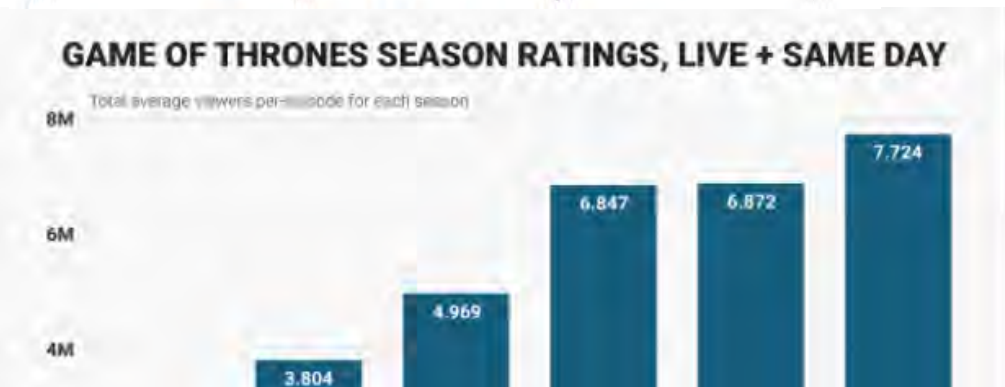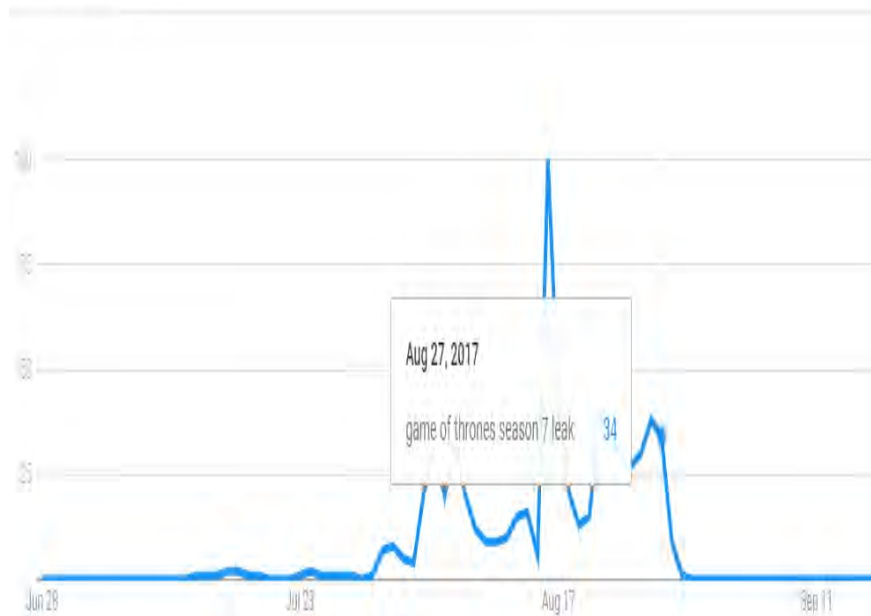
GOT team takes a considerable number of security measures, from coded scripts to ultra-secure sets to policing their airspace from peeping drones. But *GoT* also films outdoors for nearly half a year, often in tourist-drawing locations — and all it takes is one sneak with a camera to potentially ruin a plot-point. The ratings for Game of Thrones are consistently increasing across the season and probably generated conversations leading to the season premiere.

The Games of Thrones was one of the most pirated TV shows of 2017. The piracy collecting firm MUSO revealed that the seventh season of the Game of Thrones was pirated more than a billion times and the viewership is steady or increasing in spite of leaks.

Clearly the Google interest over time peaks during the time of the leaks where there is a lot of buzz around leaks. The ratings continue to go up due to the user feedback and there was a lot of Facebook engagements as seen in buzzsumo.

Interest over time ⓘ

Aug 27, 2017

game of thrones season 7 leak   34

Jun 28   Jul 23   Aug 17   Sep 11

GAME OF THRONES SEASON RATINGS, LIVE + SAME DAY

Total average viewers per-episode for each season

8M

7.724

6M   6.847   6.872

4.969

4M

3.804

Sort by	Total Shares ∨		Facebook Engagements	Linkedin Shares	Twitter Shares	Pinterest Shares	Number of Links	Total Shares ↓
The Entire Plot **Of Game Of Thrones** Season 7 Has Just Been **Leaked** By Francesca Donovan – Oct 23, 2016 unilad.co.uk	Save View Backlinks View Sharers Share		102.6K	0	93	0	-	102.7K
Game of Thrones season 7 episode 6 has been **leaked** online! By Rahul Jadhav – Aug 16, 2017 wikiofthrones.com	Save View Backlinks View Sharers Share		97.6K	0	36	0	-	97.7K
Plotline **Of** The Finale **Of Game Of Thrones** Has Been **Leaked** & Yes It Will Blow Your Mind - DotComStories By Team Dcs – Aug 18, 2017 dotcomstories.com	Save View Backlinks View Sharers Share		62.8K	0	47	1	-	62.9K
Game Of Thrones season 7 plot 'leaked' online By Ross Mcdonagh For Dailyma... – Nov 8, 2016	Save View Backlinks View Sharers Share		57.3K	6	202	9	-	57.5K

5

To understand the problem better, a word cloud was created off data collect from social media and social communities to see if it could explain sentiment of leaks, but nothing concrete came up.

Since the companies might spend huge amounts of time and resources for these TV shows, we wanted to understand if it is possible to build an analytical approach for seeing the impact of the leaks
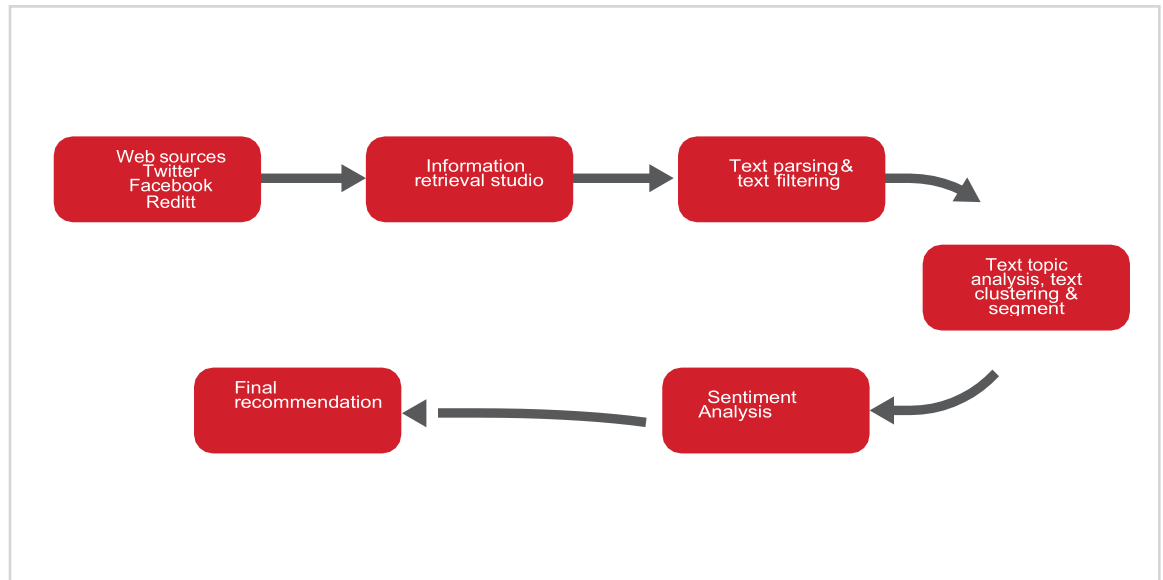
Analytics Workflow

The web was researched to find important websites and forums where people talked about leaks. Three critical sources of information for getting the viewers feedback on the leaks were identified: Facebook, Twitter and Reddit. A total of 13,550 comments were analyzed in these social media under the GOTleaks hashtags on the days when the leaks happened until the episode was aired to understand the moods and the trends of the viewers concerning the leaks.

The top 10 TV shows on Reddit

Subreddit Subscribers

Show	Subscribers
Game of Thrones	427,000
Pokemon	369,000
Breaking Bad	218,000
Doctor Who	218,000
The Walking Dead	168,000
Community	138,000
Adventure Time	137,000
Futurama	112,000
Arrested Development	109,000
Avatar: The Last Airbender	105,000

Two key analysis was identified: Sentiment Analysis and Text analysis.

Below is the sentiment analysis and text builder workflow, which are used for the analysis.

To conduct the exercise for the data preparation and analysis, the following steps were considered:

1. Data Collection

2. Text Analysis

3. Text Cluster Analysis

4. Sentiment Analysis

5. Final Recommendation

1. Data Collection Source

Research was done to select forums, websites, and social media groups where people post reviews about GOT episodes. The below forums and websites were chosen because of the high popularity amongst users.

Forums: http://asoiaf.westeros.org/index.php?/forum/31-game-of-thrones-the-hbo-tv-series/

https://winteriscoming.net/

https://www.facebook.com/: Gave insights on the big influencers and how the network of consumers is formed.
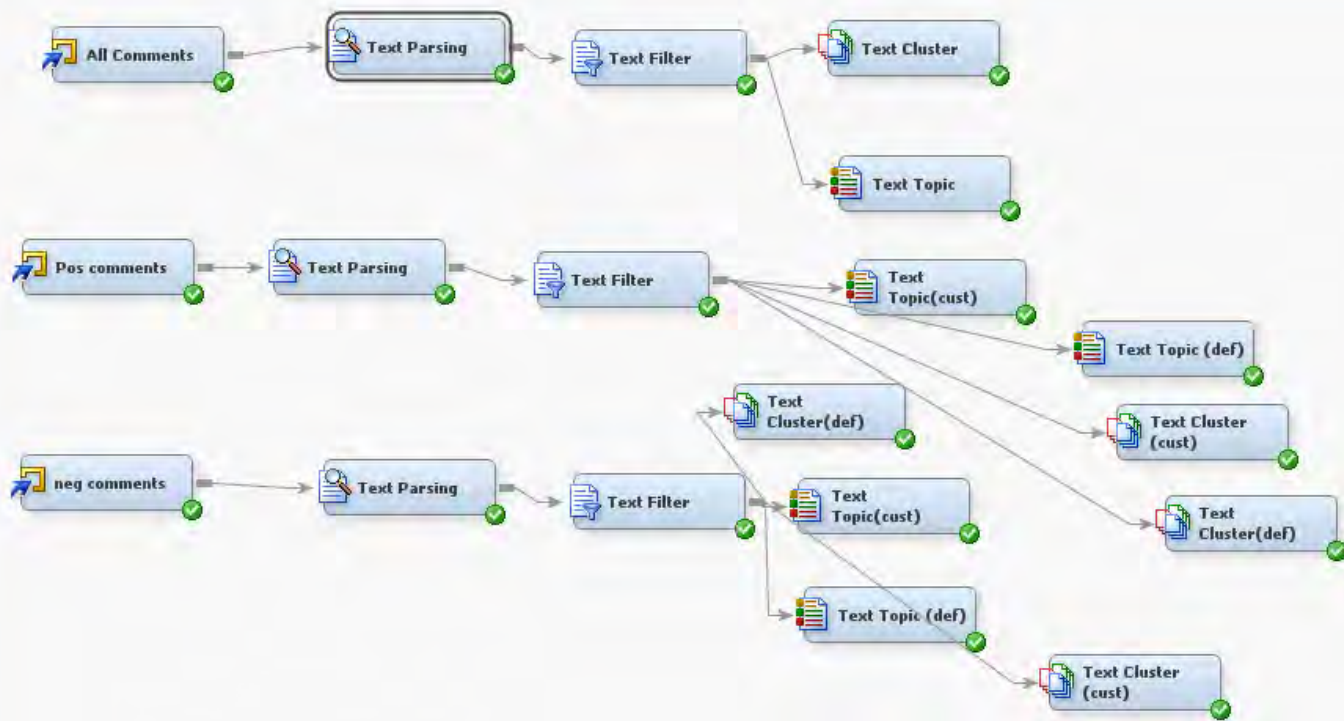
https://twitter.com/: Used to understand the sentiments of the people live as the news breaks out.

https://reddit.com/: Used to understand the sentiments mainly post airing of the episodes

Web Crawlers – Information Retrieval Studio:

SAS Information Retrieval Studio is a framework and graphical administration interface for crawling, normalizing, analysing, indexing, and searching text documents. It serves as the default user interface to SAS Web Crawler and SAS Search and Indexing, and provides integration with other products including SAS Content Categorization and

SAS Sentiment Analysis.

2. Text Parsing/Text Filter

First, Text Parsing and Text Filter nodes were used for preparing the data. To prepare the data text parsing and text filter nodes were used. The Text Parsing Node takes the raw text from the data source and can parse different languages and different parts of speech so that the unstructured data can be cleaned up and the most frequent terms can be analyzed. It is seen that we have parsed the unnecessary parts of the speech. The English dictionary is used and other languages are parsed out.

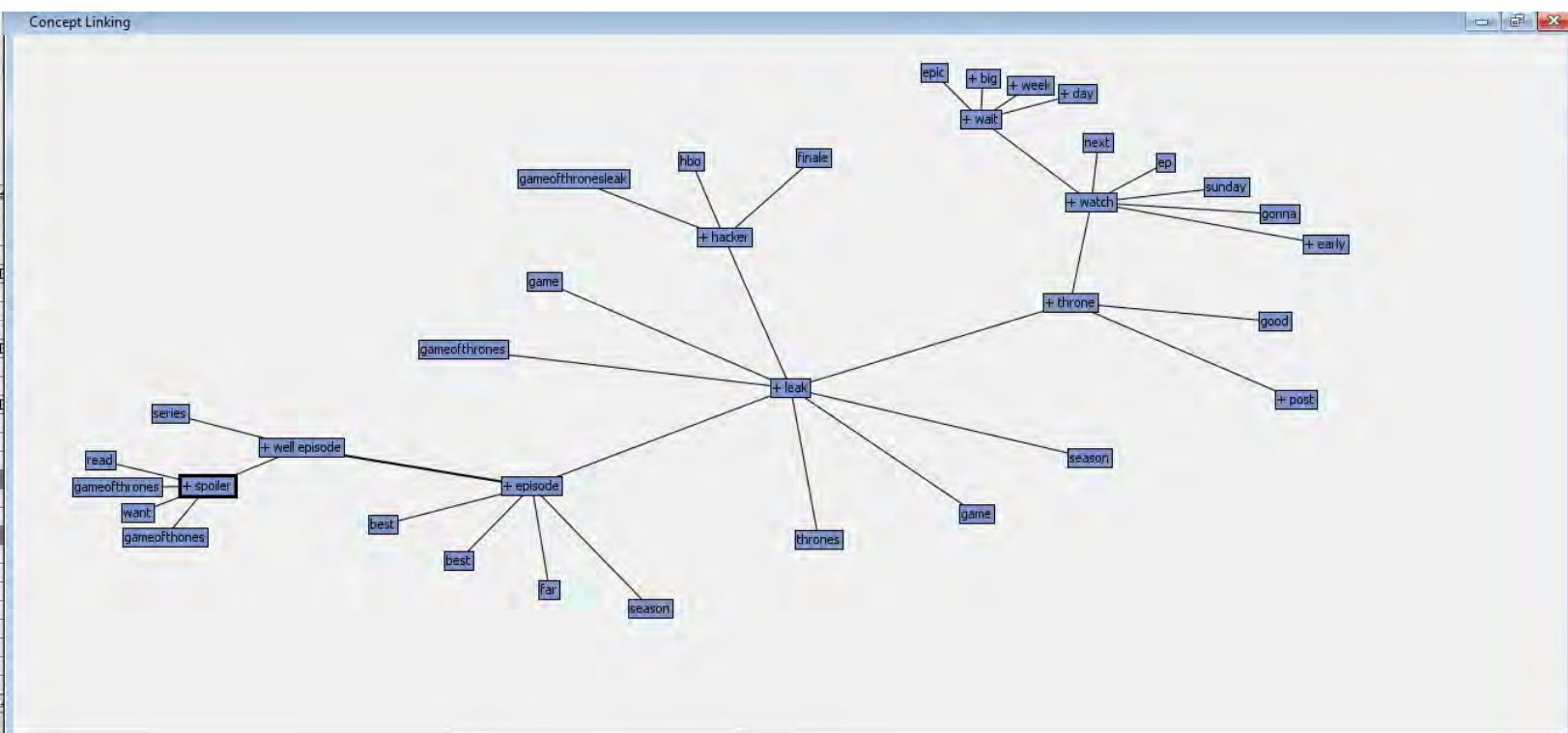Results - Node: Text Parsing (4) Diagram: GOT

File Edit View Window

Terms

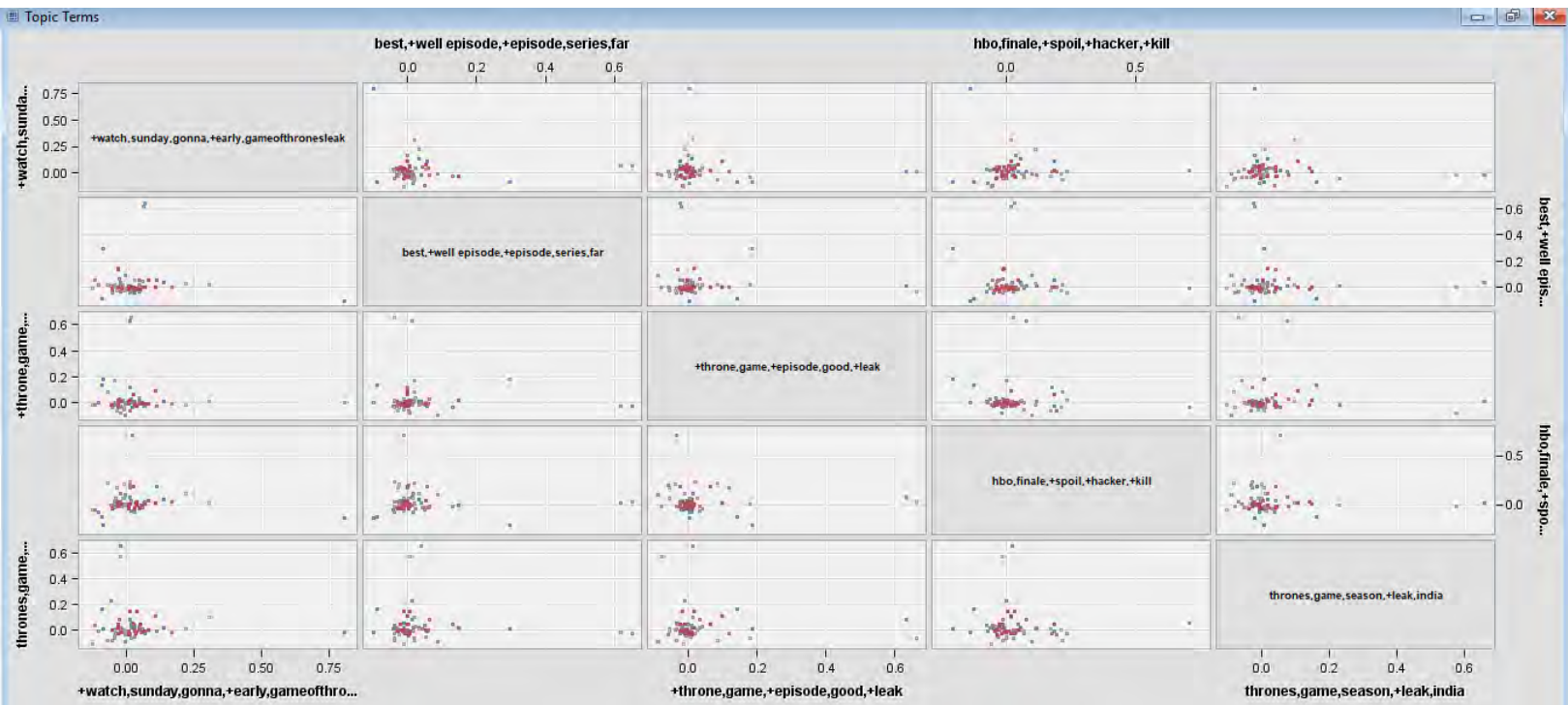| Term	Role	Attribute	Freq
gameofthronesleak	...Prop	Alpha	1
+ episode	...Noun	Alpha	
+ be	...Verb	Alpha	
gameofthrones	...Prop	Alpha	
+ leak	...Verb	Alpha	
+ watch	...Verb	Alpha	
not	...Adv	Alpha	
+ have	...Verb	Alpha	
+ get	...Verb	Alpha	
+ do	...Verb	Alpha	
s	...Noun	Alpha	
now	...Adv	Alpha	
hbo	...Prop	Alpha	
so	...Adv	Alpha	
+ wait	...Verb	Alpha	
thrones	...Prop	Alpha	
gameofthones	...Prop	Alpha	
game	...Noun	Alpha	
+ spoiler	...Noun	Alpha	
best	...Adj	Alpha	
gameofthrones	...Noun	Alpha	
just	...Adv	Alpha	
+ leak	...Noun	Alpha	
season	...Noun	Alpha	
+ well episode	...Noun Group	Alpha	
+ see	...Verb	Alpha	
game	...Prop	Alpha	
+ know	...Verb	Alpha	
+ throne	...Noun	Alpha	
again	...Adv	Alpha	
finale	...Noun	Alpha	
gots7e4	...Prop	Mixed	
new	...Adj	Alpha	
still	...Adv	Alpha	
+ go	...Verb	Alpha	
gots7	...Prop	Mixed	
+ spoil	...Verb	Alpha	

The Text Filter Node applies filters to the text data and can define the own dictionary, term weighting, frequency weighting, and term filters, which can use the defaults provided in Enterprise Miner. The least commonly occurring words are filtered out which are present in minimum number of documents. We also redefined leak which appeared as both noun and verb separately and merged them as one. The text filter also helps to correct some errors in the words. Each topic has certain key words and would describe the central theme of those comments. Clearly there are multiple themes around leaks and the storyline are getting discussed thanks to the leaks. There is a generally more positive rather than negative themes that see around the leaks.

In the Text Topic node it is seen that people are talking about the characters which is leading to additional interest being generated.

Topic ID	Document Cutoff	Term Cutoff	Topic	Number of Terms
1	0.055	0.019 +episode,+leak,last,+watch,series	57
2	0.054	0.020 tyrion,+betray,daenerys,varys,cersei	84
3	0.059	0.019 sansa,arya,littlefinger,lf,brienne	80
4	0.056	0.018 gt,+day,drake,+wall,+story	23
5	0.056	0.019 +know,+fake,+point,+talk,+implication	41
6	0.058	0.020 jon,dany,aegon,snow,+feel	78
7	0.053	0.020 +dragon,ice,fire,+wall,+ice dragon	96
8	0.050	0.020 +show,+book,+point,+watch,+character	110
9	0.053	0.020 cersei,+point,dany,+army,jaime	118
10	0.044	0.018 oc,cretu,aris,cyoa,charlie	32

Concept links show the association analysis between the terms used in the comments. If the length of the links shows the strength of the association. The concept linkage shows the leaks linked to the hacker, the episode, and thrones. Thrones, in turn, is linked to watching and waiting for the next week and how one can watch early thanks to the leaks. The episode is linked to the spoiler and people who have read GOT book also contributing to the spoilers and millions resisted watching the leaked versions.

The above table is shown in the below matrix as distribution shows the common topics where the main topic is that it is the best episode and talking about the killing the enthusiasm.

3. Text Cluster Analysis

SAS Enterprise Miner creates clusters for the terms which are associated with each other.The clusters are well separated from each other. The top four clusters are used for the analysis as it gives an idea of the clusters discussed. Cluster 2 talks about how the people who have got access to the streaming for the first time are lucky. Cluster 3 talks about the settings and music of GOT. Cluster 4 talks about how the characters and events are built in GOT. Cluster 5 talks about how spoilers kill the major event.

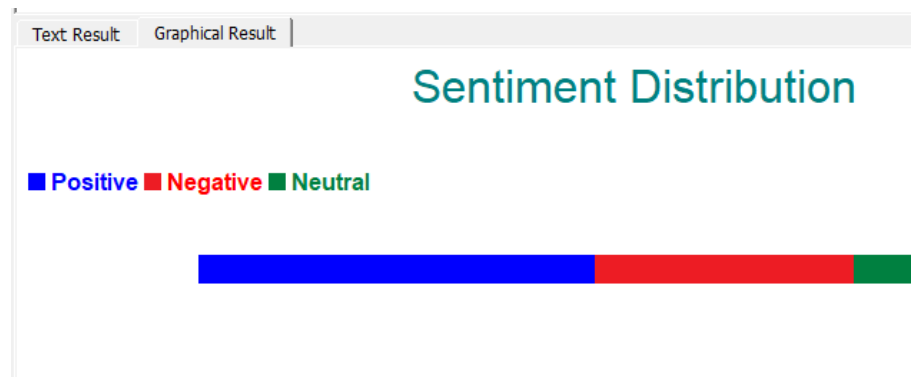| Text Cluster	Percentage	Interpretation
+leak +early +happen +watch +hack +gots7e4 +post wow	26%	This cluster is a group of terms which indicates happiness with the leaks and ability get to watch early the episode
+look +spoiler +know +happen +big help +hacker +gonna read link want	20%	This cluster is a group of terms thanking the leak and stating it was good to watch it early
+bad +ruin +start bad hbo gameofthrones gots7e7 +wait	16%	This cluster is a group of terms with negative review cluster with people expressing disappointment
+dragon +arya mother holy Sunday epic +god +throne +cersei	15%	This cluster is a group of terms about the characters

The biggest two clusters discuss how the leaks were a big help to watching early and are indicating happiness. The third cluster is people expressing disappointment about the spoilers and how disappointed they have been. The next cluster is about the guessing regarding the story line and the characters and what is going to happen next. Thus there is a definite buzz created due to these leaks looking at these topics.

Sentiment Analyzer

From the text topic and text cluster node, leaks are getting discussed and are generating a buzz.

The parsed data is then sent to the SAS sentiment analysis studio which should be able to classify the tweets as positive or negative. 80% of the comments are used for training the dataset. The best model post training is the Smoothed Relative frequency and Chi Square. Chi Square is the feature building algorithm that uses the terms in the document based on the frequency and importance and the Smoothened Relative frequency algorithm performs text normalization.

Text Result | Graphical Result |

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [No Feature Ranking]:
      Overall precision: 85.42%
      Positive precision: 63.16%
      Negative precision: 100.00%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Risk Ratio]:
      Overall precision: 77.08%
      Positive precision: 57.89%
      Negative precision: 89.66%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Chi Square]:
      Overall precision: 89.58%
      Positive precision: 73.68%
      Negative precision: 100.00%

With text normalization algorithm [Smoothed Relative Frequency] and feature ranking algorithm [Information Gain]:
      Overall precision: 89.58%
      Positive precision: 73.68%
      Negative precision: 100.00%

BEST MODEL is Smoothed Relative Frequency and Chi Square

The rest of the data is used for the actual data validation. 56% of the sentiment was positive, 38% of the sentiment was negative and 6% of the sentiment was neutral. Thus the majority of the sentiments were positive where the characters were discussed at length and how leaks have generated more interest among the non-viewers. This sentiment analysis thus shows the views and opinions of the leaked views and is overall positive resulting in lots of buzz for the show.

Text Result | Graphical Result |

Sentiment Distribution

■ **Positive** ■ **Negative** ■ **Neutral**

Conclusion

The social media gives us a good glimpse on why the viewership increased overall in spite of the leaks. Concept linkage shows a strong association between the leaks and the characters being discussed and people wishing that they can't wait to see the series. Overall there was a lot of positive sentiment around leaked content characters especially by the GOT fans and those who have paid for the HBO channel. It was one of the most pirated content in terms of discussing it on social media before the episode is aired on HBO. Thankfully the Game of Thrones brand was strong and the leaks had rolled over and a lot of people might have rewatched by subscribing to the network. Also, the piracy also became a massive opportunity to engage people such as discussing the plotline and bring them to legitimate platforms results in high viewership levels. Thus the Game of Throne leaks didn't harm the HBO network financially because of the resilient income streams in subscriptions and merchandise. If the content was not strong enough the leaks might have killed the TV show with negative word of mouth and but in the GOT case it clearly has helped the franchise grow thanks to the leaks.

References

1) http://ew.com/tv/2017/07/16/game-of-thrones-cast-slam-leakers/

2) https://moviepilot.com/p/do-game-of-thrones-leaks-cost-hbo-money/4352251

3) http://wikiofthrones.com/11342/game-thrones-season-7-episode-6-leaked-online/

4) https://www.unilad.co.uk/tv/the-entire-plot-of-game-of-thrones-season-7-has-just-been-leaked/

5) http://wikiofthrones.com/11342/game-thrones-season-7-episode-6-leaked-online/

6) http://support.sas.com/demosdownloads/001175/readme_all.pdf