Decomposing the Story of Stranger Things with Natural Language Processing

Sarah Grobe[1, *]

[1] *University of Vermont*

(Dated: May 10, 2023)

Using ousiometrics, we were able to break down the story of the Netflix show *Stranger Things* into scores of power, danger, and structure. By comparing these values over time, both within the scope of individual seasons as well as the series as a whole, we were able to further evaluate how these scores balance to create a cohesive story. Further, using word shift diagrams, we were able to compare seasons to each other in terms of danger and sentiment in order to better understand the tone shifts that occur throughout the series, and the language which drives them.

I. Introduction

The corpus utilized for this project is the subtitle transcripts for all four seasons of the Netflix show *Stranger Things*. Created and largely written by brothers Matt and Ross Duffer, *Stranger Things* is one of the most popular programs on the streaming platform, first releasing in 2016, with the most recent season releasing in July of 2022. Since this corpus is comprised of only subtitles, it is restricted to only dialogue and audio descriptions, and does not have other features that would be present in a script or a book, such as descriptions of the setting, or any non-verbal interactions between characters.

Stranger Things tells the story of four young boys who become entangled in a battle with an alternate world, called the "Upside Down," and the monsters that inhabit it. Heavily based on the fantasy game *Dungeons & Dragons*, the show has a wide range of monsters and battle sequences, as well as light-hearted moments between its many characters. This makes *Stranger Things* an interesting candidate for decomposition into its different story elements. Some similar projects appear to have been done in the past[1, 2], although there do not appear to be any yet which are this thorough, or which are officially published.

II. Description of data sets

Transcripts were gathered from an online source which provided them in a per-episode format. These transcripts were stored as text files, both for each episode individually, and as a single file containing the transcript for the entire series.

The corpus is comprised of 150,265 1-grams, which boils down to 8,072 unique words (case insensitive). The show itself features four seasons, each made up of 8 or 9 episodes following roughly the same cast of characters across a season-long story arc, which fits into larger story spanning the entire series. For this reason, it was assumed that, despite some episode-length ebbs and flows within the story, sliding windows spanning between both episodes and seasons were permitted in order to help capture these lengthier story lines.

Some basic preprocessing was implemented to the data sets as well, before any analyses were applied. In particular, when decomposing into 1-grams, all non-alphabetic characters (i.e., punctuation and numbers) were removed. Additionally, a copy of the data set was created in which some nicknames or other alternate names for major characters were replaced with a single, standardized name for that character in order to assist with more consistent naming. For example, one character goes by the name "Mike" for the majority of the story, but is occasionally referred to as "Michael." This portion of the preprocessing replaced all instances of "Michael" with the name "Mike" instead, in order to have more consistent and accurate results when evaluating the story around the appearance or presence of specific characters.

All code and data files for this project can be found here on GitHub.

III. Results

To begin gaining a sense of power and danger across the series, a heatmap was created for each of these values, shown in Figures 1 and 2. These plots were created by calculating the average of power and danger respectively for all 1-grams in a given episode.

When looking specifically at danger in Figure 1, we can see that in general, danger seems to be at its highest

* sgrobe@uvm.edu

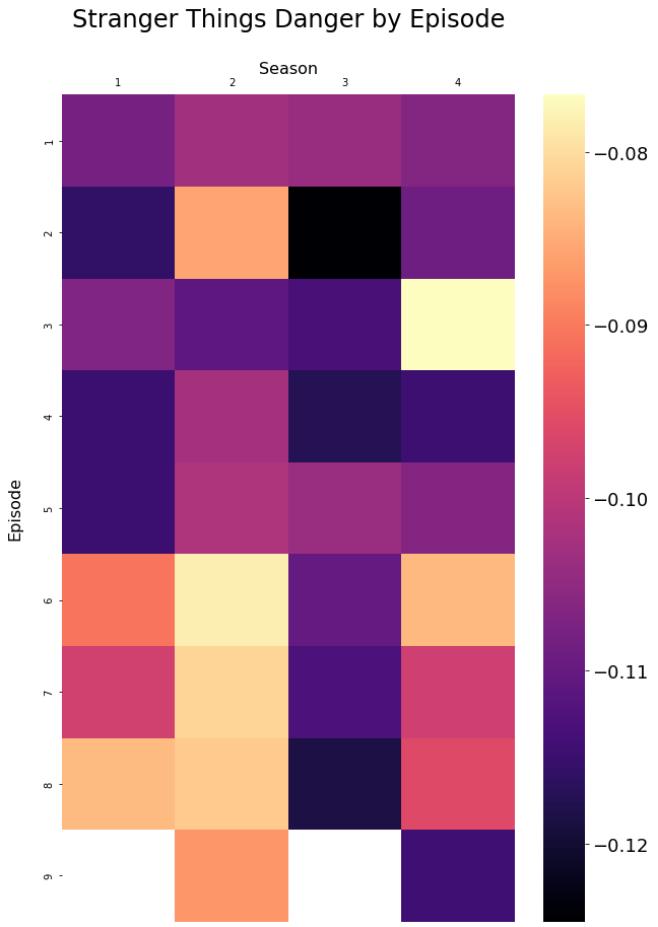

FIG. 1. Danger of each episode, obtained by averaging across all 1-grams appearing in that episode.

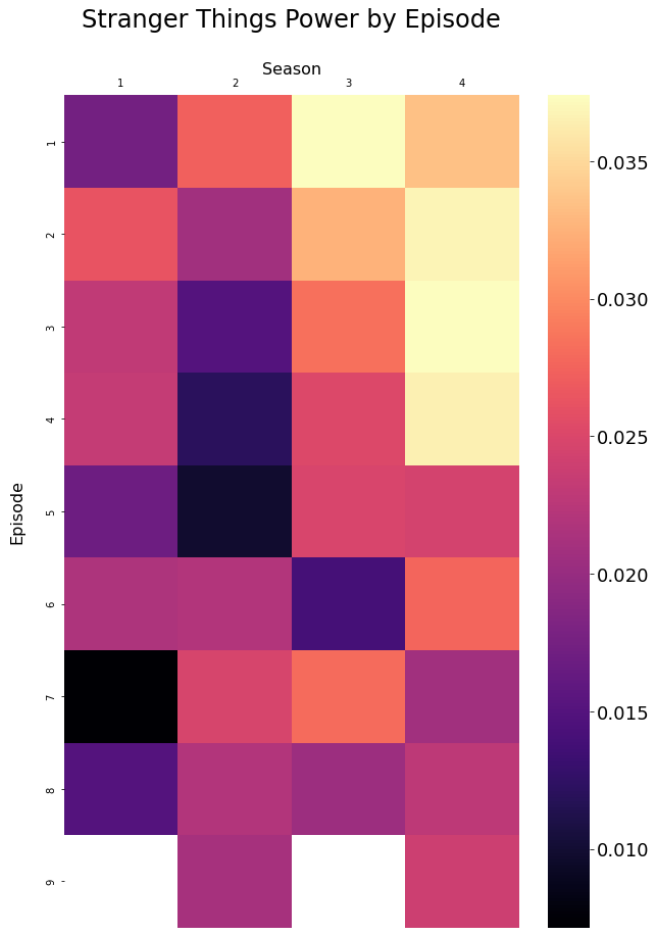

FIG. 2. Power of each episode, obtained by averaging across all 1-grams appearing in that episode.

towards teh end of the season, and the lowest in the early episodes. Season 4 is a bit of an exception to this, peaking danger in episode 3, and having a second, but slightly lower peak, towards the end as we see in the others. We can also gather from this that episode 3 of season 4 appears to be the most dangerous episode, while episode 2 of season 3 is the least dangerous, on average. Additionally, it is apparent from this plot that season 3 is the least dangerous of the four seasons.

The heat map of power in Figure 2 shows almost the opposite of the danger heat map; we instead see, in general, higher power earlier in the season than later on. Additionally, we can see that season 4 has the highest power in general compared to other seasons, with episodes 2, 3, and 4 having some of the highest average power values across the entire series. The lowest power scores appear to happen more in the first half of the series as well, while higher averages tend to happen in seasons 3 and 4.

Now that we have a baseline understanding of the distribution of danger and power across episodes, we can move to understanding time series plots of these values, built using a sliding window.

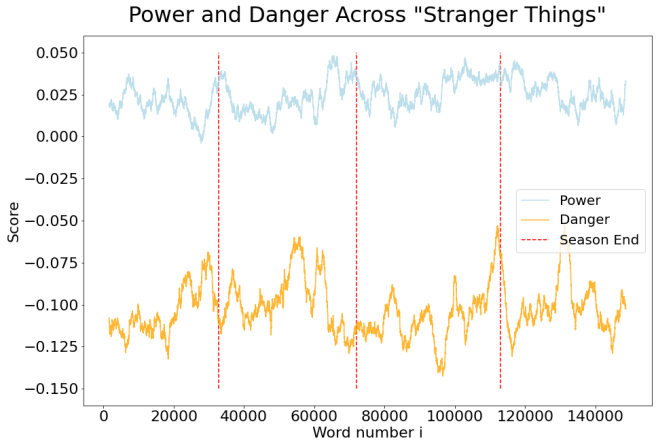

FIG. 3. A time series plot of power and danger scores across the series using a sliding window of $10^{3.5}$ words. A window of this size is optimal for capturing episodic fluctuations in scores for this data set.

FIG. 4. A time series plot of power and danger scores across the series using a sliding window of 10^4 words. A window of this size is optimal for capturing season-length trends in scores, burying more of the noise from individual episode arcs.

The first of these plots is shown in Figure 3, built using sliding windows of $10^{3.5}$ 1-grams. This size was determined to be best for showing the overall trend in power and danger, while still retaining some episodic fluctuations. Using the vertical lines as a guide denoting the ends of each season, we see much of the same fluctuations as were shown in the heat maps, this time in greater detail. Specifically, we can see that each season has multiple smaller story arcs, taking place over the course of an episode or two, culminating in multiple peaks within the course of the season plotline. We can also see that in terms of danger, seasons 1 and 2 appear to follow a similar trend, beginning with relatively low danger which then rapidly increases into two peaks, separated by an episode. This is reinforced by the heat map in Figure 1, although it is more clearly shown in the time series plot.

The same plot was then replicated using a window size of 10^4, as this window size filters out much of the noise from individual episodes, allowing for a better understanding of the progression of power and danger across seasons and the series as a whole. We see more clearly the clear peaks that occur towards the end of each season, present in the values of both power and danger. We can also see here more clearly the gradual increase in power as the series progresses. From a story telling point of view, this is likely due to early seasons being more focused on introducing the large cast of characters and the setting, allowing for later seasons to dive into more of the fantasy elements of the show, which likely relates to higher values of power.

One difficulty presented by these time series plots in terms of comparing seasons is the varying lengths of these seasons in terms of the number of 1-grams. To remedy this, Figure 5 shows the four seasons stacked

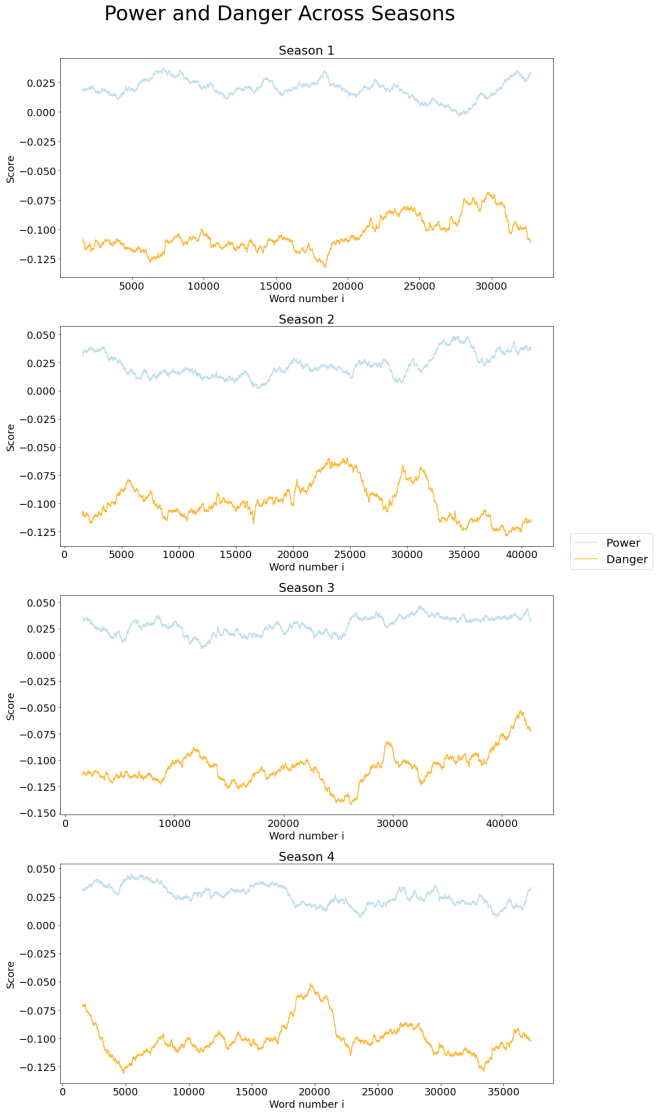

FIG. 5. A stacked time series plot of power and danger scores for the four seasons, using sliding windows of size $10^{3.5}$.

and scaled to the same size regardless of the number of 1-grams, in order help show how power and danger vary in regards to the location within the season. One key finding which jumps out is that danger does not appear to peak at consistent times within seasons. Season 4 has the earliest peak, occurring near the middle of the season followed by some smaller fluctuations. Season 2 is the next earliest with two peaks, the first of which appears to occur just after the midway point of the season. Season 1 appears to follow a more complete story arc, building in danger to a peak over the last couple episodes, before declining over the last part of the story. Season 3, by contrast, has relatively low danger throughout before finally peaking right at the very end of the season. We can then see this season 3 cliffhanger leaking into the beginning of season 4.

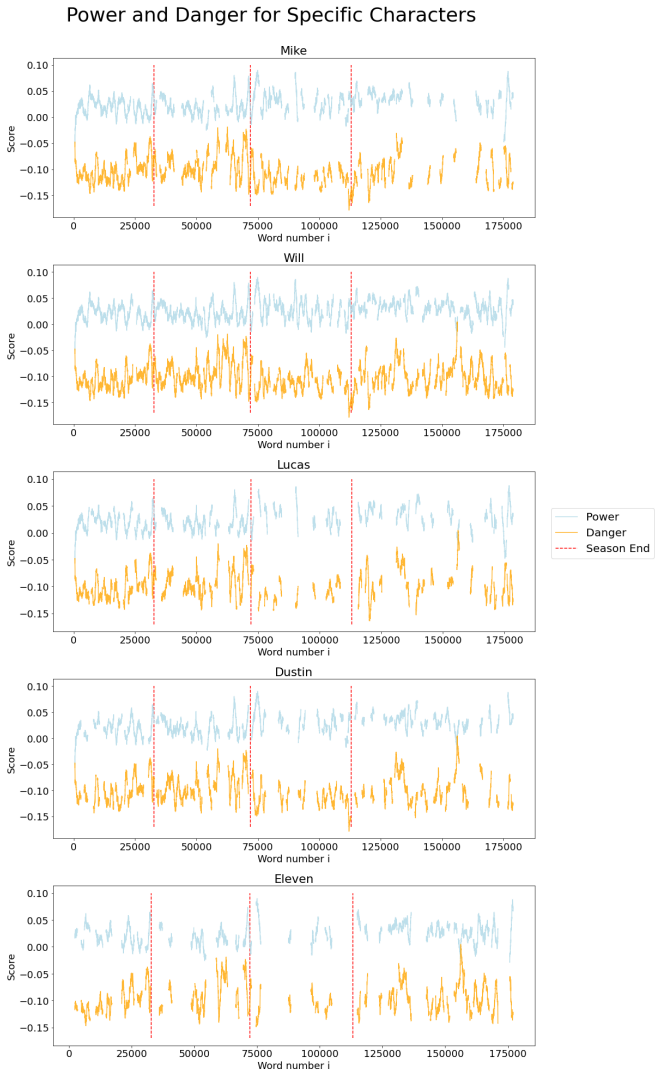

FIG. 6. Power and danger time series for main characters, calculated using a sliding window of 100 words, for all windows in which that character's name appears. Main characters were defined as characters who appeared regularly throughout the entire course of the show, and who played a major role in the story throughout.

Power scores, however, tell a bit of a different story from danger. Specifically, we don't see many fluctuations in power across seasons, with the exception of what appears to be a slight upward trend through the first three seasons, with potentially a decrease in season 4. However, since these trends are not incredibly obvious, further analysis is needed to determine whether or not these trends are significant.

Next, we chose to investigate power and danger for the appearance of certain characters. These were calculated by calculating the average power and danger over every 100 word window in which the character's name appears (i.e., 99 words on either side every time they are mentioned). The preprocessing conducted on the corpus was designed to reduce the number of different names a given character could go by throughout the series in order to help improve accuracy of these methods.

Figure 6 shows these character plots for the five main characters, Mike, Will, Lucas, Dustin, and Eleven. As expected, these characters are all present for the majority of the series, and as a result, their respective plots show the power and danger of nearly the entire series. In the latter half of the series, we do start to see some separation in the characters' plots, and it is at this point in the show when some of the characters begin breaking off into smaller groups. This is especially apparent in season 4, when groups of characters were in entirely different parts of the world at certain parts of the story.

Another similar plot was created for secondary figures in Figure 7. For these purposes, secondary characters are considered characters who were present in all four seasons of the show, but did not have the same importance to the story as the five main characters. The secondary characters included here are Steve, Nancy, Jonathan, Hopper, and Joyce. Since these characters are mentioned a bit less throughout the story than the main characters, they have slightly more segmented time series than shown in Figure 6. In particular, we can see better which characters spent more time together throughout. For example, Steve and Nancy have very similar plots, especially in season 4 when many of the characters were separated, indicating that they spent a large part of the series together. We can also see at the very end of the series when all of these characters are reunited.

A final time series plot was then created for tertiary characters, defined here as any character which played a minor role and/or was not present for all four seasons. By this definition, the tertiary characters were considered to be Max, Robin, Murray, the demogorgon, and Dr. Brenner. It should be noted that although Max plays an important role in the story, she was considered a tertiary character to allow for better comparisons between characters, since she wasn't in the show until the second season.

In these plots shown in Figure 8, we can see the specific points when these characters make an appearance. The demogorgon, for instance plays a much bigger role in earlier seasons before taking more a backseat, serving as a minion for more sinister villains in the second half of the series. We can also see the significant role that Dr. Brenner plays in season 4, despite being a very minor character in the earlier seasons, especially season 3.

The next step to understanding the story of *Stranger Things* was to utilize word shifts to get a better understanding of what is fueling the differences in scores between different corpora within the series. Since

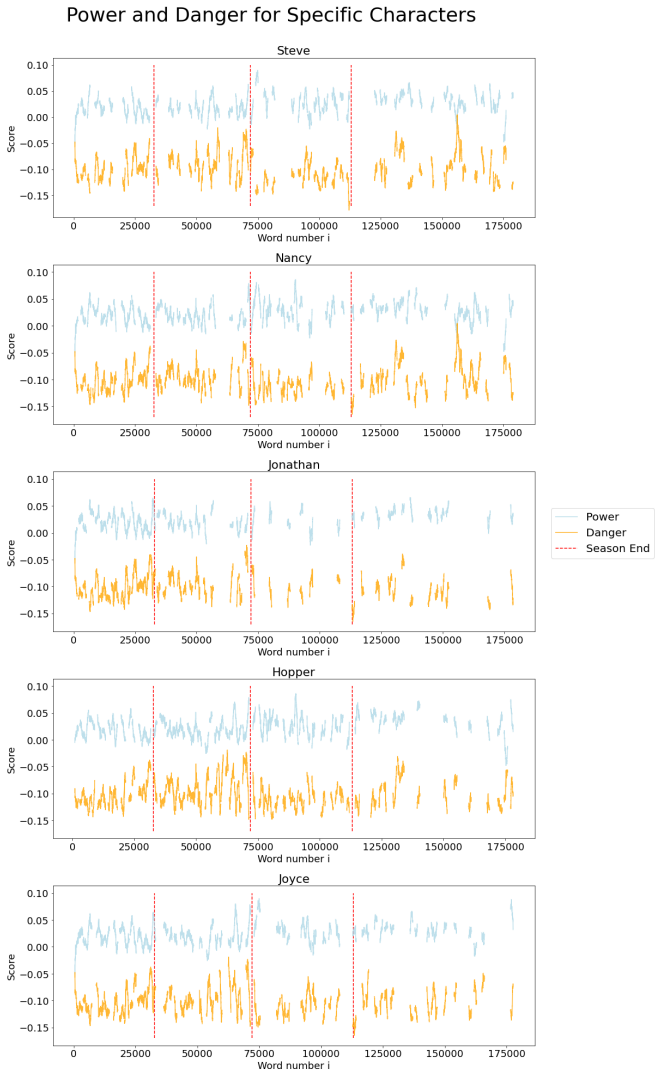

FIG. 7. Power and danger time series for secondary characters, calculated using a sliding window of 100 words, for all windows in which that character's name appears. Secondary characters were defined as characters who were present throughout all four seasons of the show and who often participated and played an important role in the story, but not to the same magnitude as the main characters.

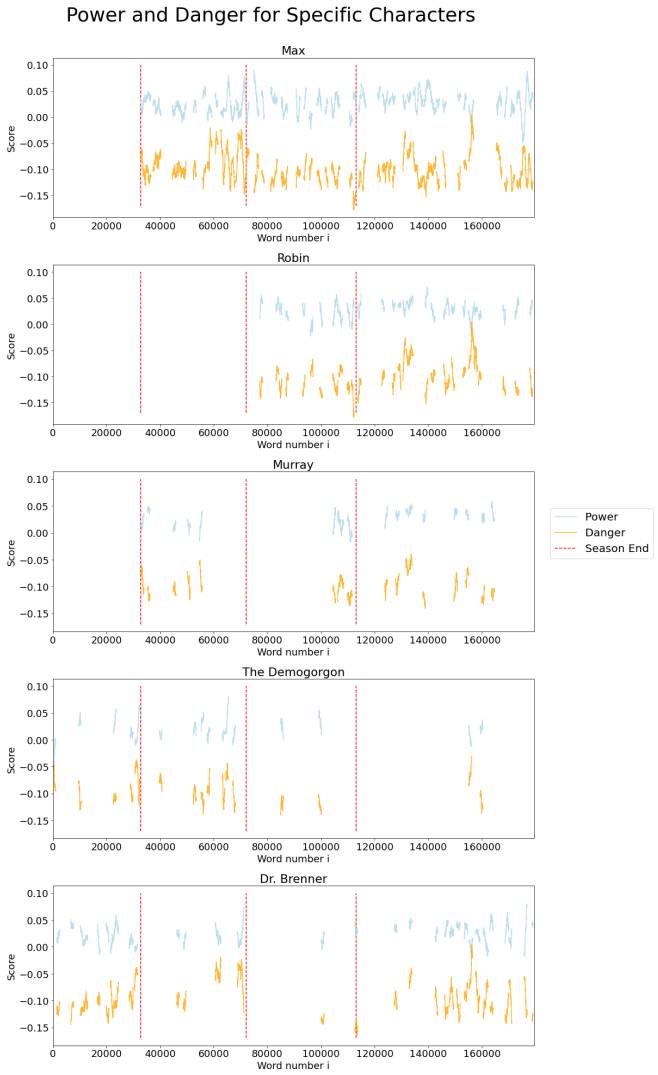

FIG. 8. Power and danger time series for tertiary characters, calculated using a sliding window of 100 words, for all windows in which that character's name appears. Tertiary characters were defined as characters who played a minor role in the story and/or were not present for large portions of it. Although the character Max plays a key role in some aspects of the story, particularly in season 4, she is considered a tertiary character due to her absence in season 1, in order to allow for better comparisons within the stacked time series plots.

there was considerably more variation in the danger scores compared to power, danger was the primary focus of the subsequent word shifts.

Conducting a brief analysis of the average danger scores across each of the seasons, it was found that season 3 had the lowest average danger across each of the seasons. Therefore, the first word shift, in Figure 9, aims to show what makes season 3 less dangerous compared to the other three seasons. The first thing that pops out from these word shifts is the character names, specifically "will," showing high up on the season 1, 2, and 4 side of the graphic. While these results could be useful, they also have the potential to artificially skew the results slightly, since "will" is both the name of a central character in the show, and a regular word with associated semantic meanings. To avoid this, a second word shift was created for the same two corpora, but in this case, all character names were removed from both. The resulting word shift is shown in Figure 10.

On the right side, representing the slightly more dangerous seasons 1, 2, and 4 combined, we see very harsh, negative words, including words like "murderer,"

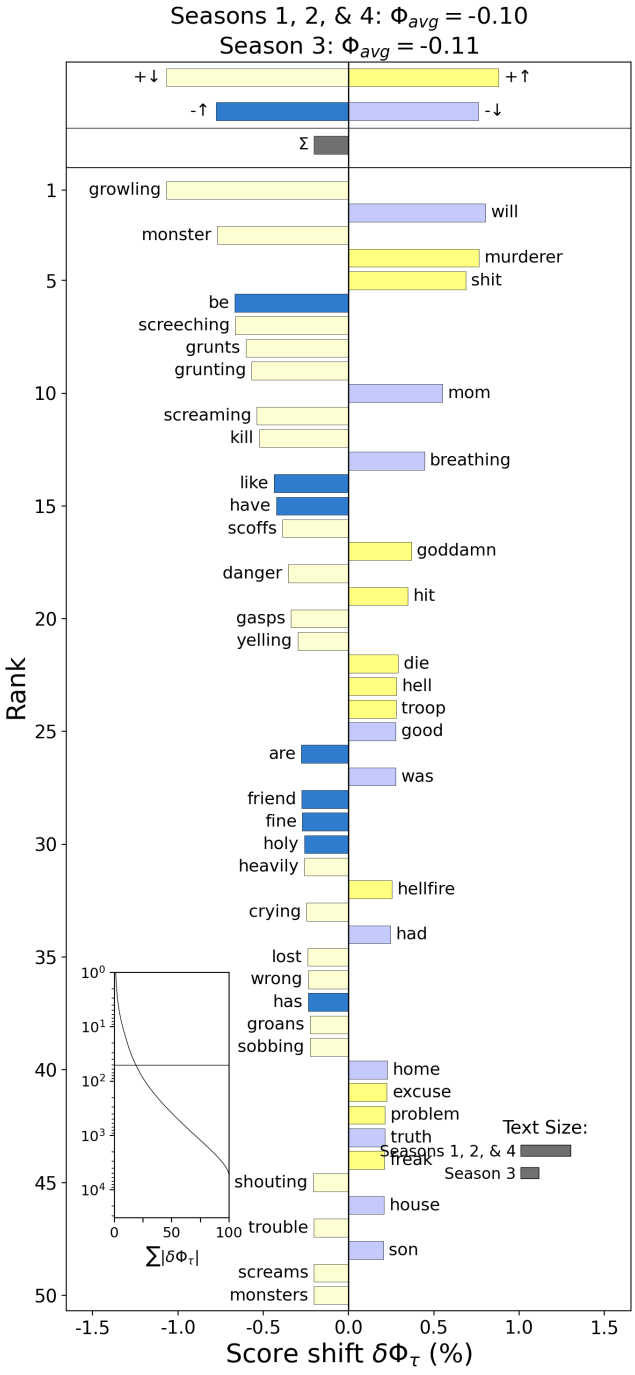

FIG. 9. Words contributing the most to danger scores for Seasons 1, 2, and 4 combined compared with Season 3.

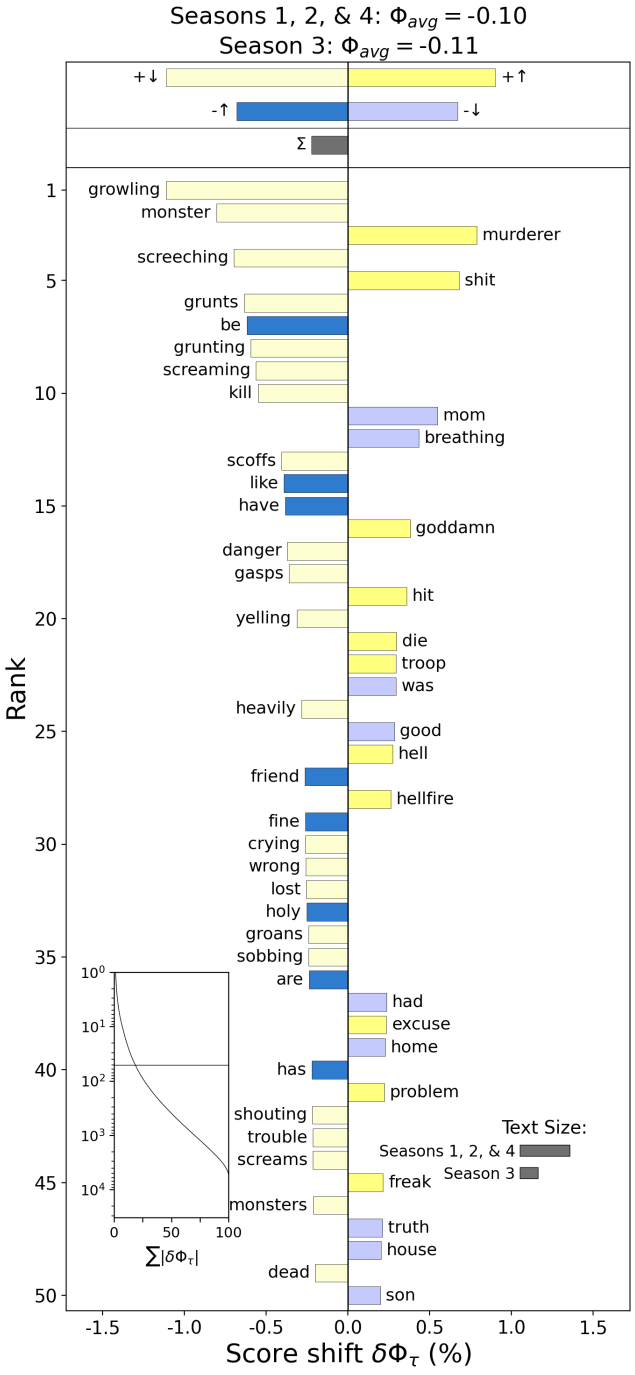

FIG. 10. Words contributing the most to danger scores for Seasons 1, 2, and 4 combined compared with Season 3, ignoring character names in both corpora.

"die," and various curse words. While season 3 on the left contains its fair share of dangerous words, such as "growling," "monster," and "screaming," it doesn't feature dangerous words with the same magnitude as the other seasons. Additionally, there is a clear difference in positive words found in season 3, such as "holy" and "friend," which help result in a much lower danger score, and these types of words do not appear to be present in the other three seasons to the same degree.

To investigate further the difference between more and less dangerous seasons, we then built a word shift for the two seasons with the highest average danger, seasons 2 and 4, compared to the two seasons with the

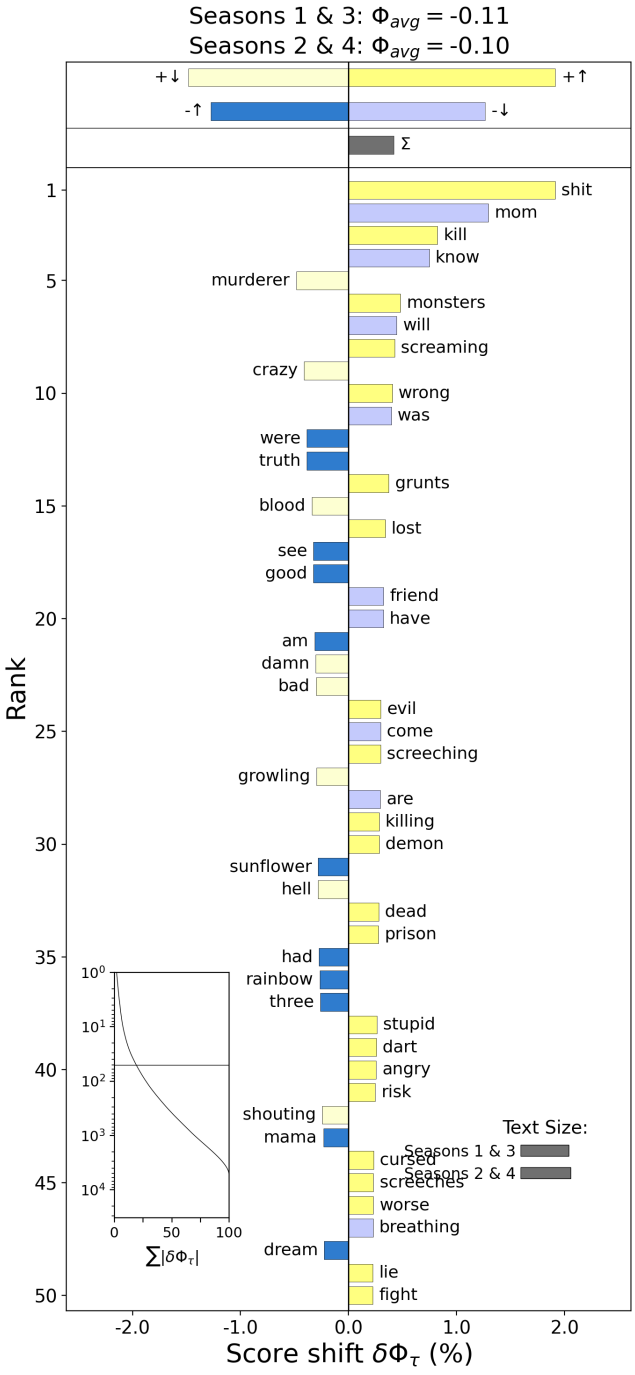

FIG. 11. Words contributing the most to danger scores for the two least dangerous seasons, seasons 1 and 3, compared to the two most dangerous seasons, seasons 2 and 4.

lowest average danger, seasons 1 and 3, the results of which are shown in Figure 11. This word shift also has the added benefit of comparing two corpora with more much more similar sizes than were shown in Figures 9 and 10.

Notably, this word shift shown in Figure 11 is asymmetric near the top. Similar to above, we also see a higher volume of stronger words on the right side, representing the more dangerous seasons 2 and 4. Some of these words include "kill," "evil," "demon," "prison," and of course, curse words. While the less dangerous seasons have powerful dangerous words as well, such as "murderer" and "blood," these words are neutralized by a higher volume of less dangerous words that aren't seen in the more dangerous seasons to the same degree, such as "sunflower," "rainbow," and "dream."

Further comparing the two most dangerous seasons, a word shift was created for season 2 vs. season 4, shown in Figure 12. There isn't much of a difference in the overall danger between these two corpora, with a fairly symmetrical diagram which is also fairly balanced when we consider the magnitude of some of the words on both sides. Both sides of the diagram contain curse words, as well as words related to the monsters of the series, such as "screeching" and "monsters" on the left side and "cursed" and "demon" on the right side. This word shift, unlike the others prior, helps show balance between these two corpora, and how they've achieved very similar overall danger scores.

As a final word shift, seasons 1, 2, and 4 were once again compared against season 3, this time using sentiment of the text to help determine differences, the motivation being that season 3 had not only a lower danger score than the other seasons, but also to many viewers was considered to have a very different, and even lighter atmosphere for the majority of the episodes, compared to the rest of the series. When we consider this word shift using sentiment in Figure 13, we see many of the same results present in the word shift using danger. In particular, we see that season 3 was found to be slightly more positive than the other three seasons, which contained a higher volume of curse words and negatively connotated words, such as "dont," "problem," and "worry." Interestingly, season 3 on the right also contains a high volume of negative words, as shown by the large amounts of blue in the plot. However, since this corpus has a higher sentiment than the remaining three seasons, it stands to reason that it must have a lower volume of more positively sentiment words, which result in the more positive sentiment than is found in the other corpus.

Finally, an ousiogram was created on the power-danger axes, shown in Figure 14. There appears to be the most variation in this plot along the dangerous-weak to safe-powerful axis, with a majority of words somewhat surprisingly falling on the "safe" end of the plot, specifically in the region of safe-powerful. Despite the majority of words falling within this region, however, we can see from the histograms on the top and right side of the plot that the distributions are skewed with large tails, as shown on the end of the "dangerous" and

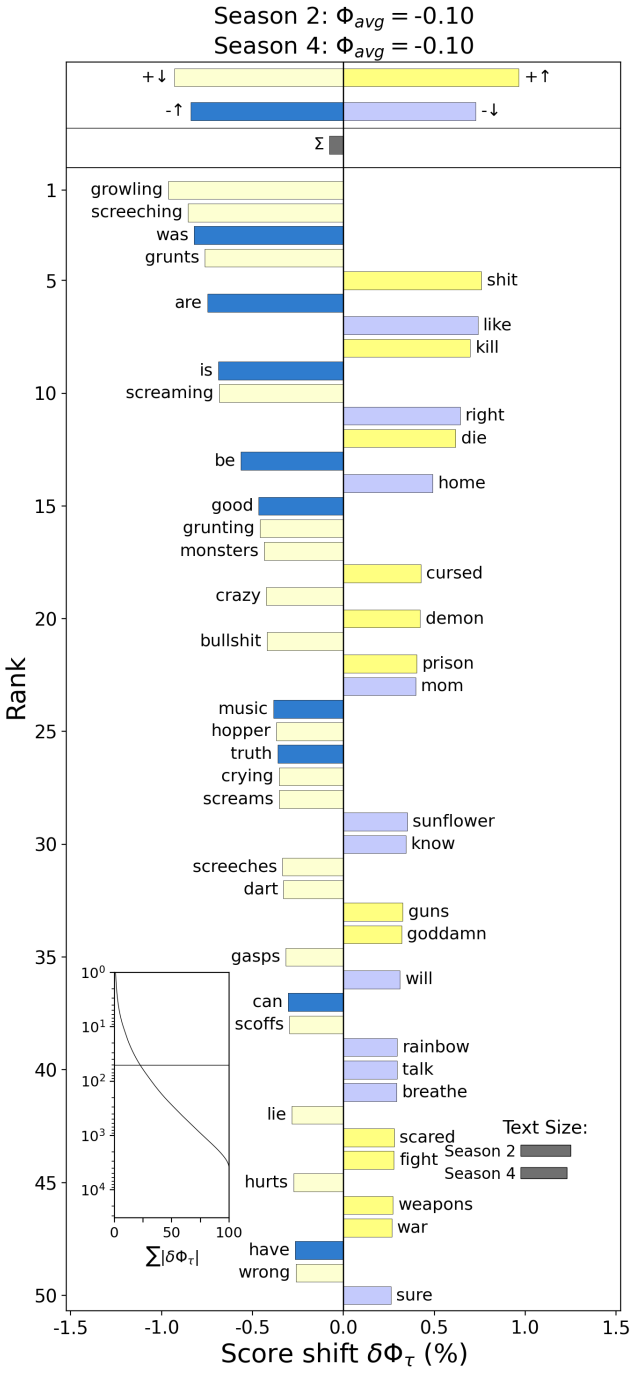

FIG. 12. Words contributing the most to danger scores for the two most dangerous seasons, season 2 and season 4.

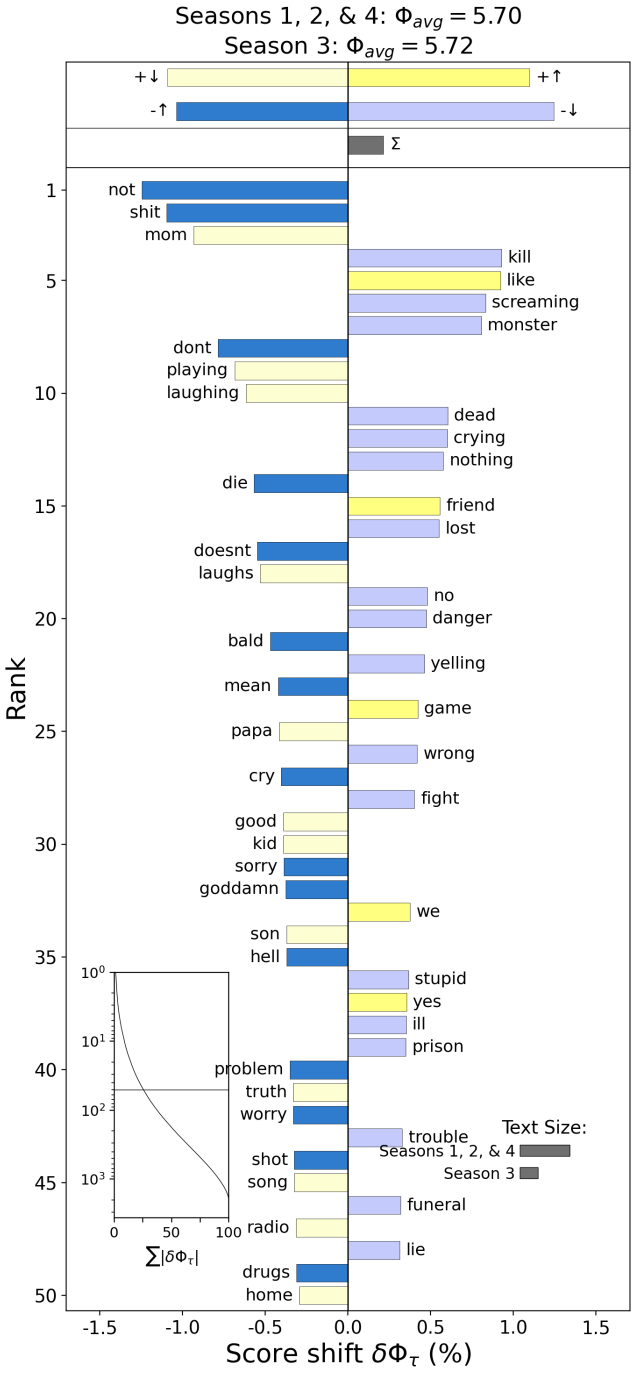

FIG. 13. Words contributing the most to the sentiment of seasons 1, 2, and 4 compared with season 3. Season 3 was isolated for comparison due to having a very different tone in general from the other three seasons.

"powerful" axes. These results indicate that, although these words are not used frequently in comparison to other words in the text, the corpus does contain both highly dangerous words, and highly powerful words at certain points in the story.

IV. Concluding remarks

Through breaking down the show's transcripts into sentiment, power, danger, and structure scores, we are able to build visualizations of these different metrics

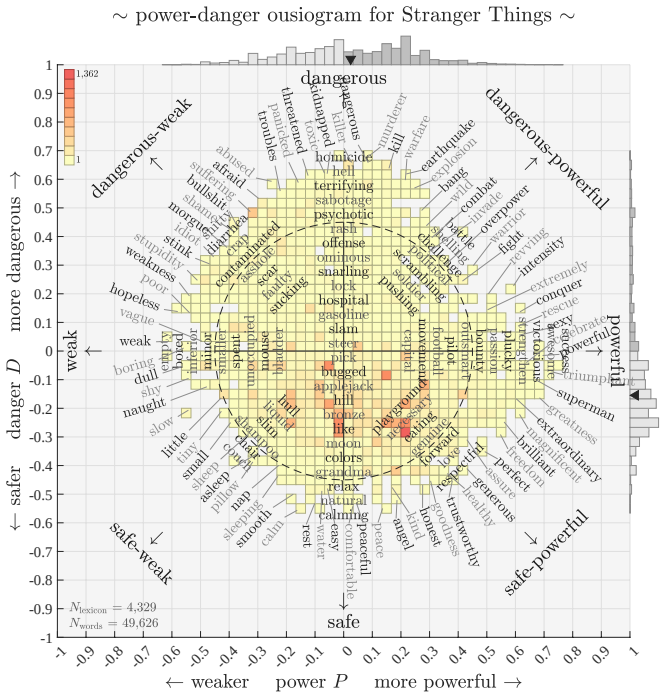

FIG. 14. An ousiogram depicting power and danger throughout the series.

over time, and how they ebb and flow to build a cohesive story. Additionally, by comparing parts of the story against each other, including individual episodes and seasons, we were able to better understand tone shifts throughout the series, as well the words driving those changes throughout the story. With this type of analysis, we are able to unlock and quantify the story-telling elements that make a written work unique.

V. Methods

As mentioned in the data set description, some light preprocessing was performed on the text in order to help facilitate better analyses. To start, the text was broken down into 1-grams, with punctuation removed. These 1-grams were then written to a file for easier access, with one word per line. As described above, some character names were also replaced with alternatives in order to reduce the number of different names associated with a single character, for instance, replacing all instances of "Jane" with the name "Eleven," since that is the name that particular character goes by for the majority of the story.

Sliding windows were also used to create the time series plots by calculating averages over the window of specified length, moving the the window by one 1-gram in order to keep the same size, and recalculating the average. This method was then repeated until the sliding window reached the end of the corpus. Optimal sizes for the sliding windows were determined by evaluating a range of values. It was ultimately determined that for this corpus, $10^{3.5}$ was the optimal size for understanding general flow of the story while also retaining episodic ebbs and flows, while 10^4 was optimal for getting a better sense of the ebbs and flows of a season or the entire series, but drowning out much of the smaller fluctuations occurring as a result of individual story arcs.

Acknowledgments

We are grateful for everyone in POCS Vol. 2 who helped with debugging code and building the plots for this project.

[1] P. Banović, The analysis of monsters and trauma in stranger things, (Undergraduate thesis). University of Zadar. Retrieved from https://urn.nsk.hr/urn:nbn:hr:162:478139.

[2] A. A. Shoaybi and M. Naghashzadeh, The capacities of todorov's narrative theory in the plot analysis of fiction series, (case study: Stranger things series), Quarterly Scientific Journal of Audio-Visual Media **15**, 41 (2021), http://javm.iribu.ac.ir/article$_1$29620$_d$b86ecf10939209c11fcd4a405338339.pdf.