Title: “Put the Fucking Salary in the Job Ad”: An Analysis of an Anonymous Corpus of Tweets

DOI: https://doi.org/10.4324/9781003159155-8

Author

1. Tim Ribaric (ORCID ID: 0000-0001-9229-8569)

Abstract

In February of 2016, I activated the @lis_grievances Twitter bot. The dynamics of the bot are straightforward and can be described in three steps: first, a person sends a direct message to the account; second, the message is stripped of all identifying information; and third, upon passing a minimal list of posting criteria, the message is tweeted. Five plus years on, the bot has collected a corpus of thousands of tweets, some safe to publish on Twitter and some not, ranging from benign takes on the library establishment to profanity laden tirades. Quite often, the tweets invoke feelings that range from pathos to disgust, and sometimes even situational irony and humor as evidenced, for example, in this tweet from June 1, 2018: “How can we innovate when we don’t have permissions to install software?” This chapter examines tweeted content through the online disinhibition effect (ODE), a theory that attempts to explain how anonymity tends to push sentiment into more extreme directions. According to ODE, users of @lis_grievances experience a lack of restraint due to their anonymity and, thus, feel comfortable venting and otherwise offering observations of and comments on perceived flaws in their individual workplaces and in the LIS profession at large. Using text analysis and a new customized metric called the grief index, a qualitative and quantitative examination of the corpus of tweets is presented and explored as evidence of systemically dysfunctional library states.
Keywords

Twitter, Text Analysis, Anonymity, Online Disinhibition Effect, Grief Index

Introduction

On February 26, 2016, the Twitter account @lis_grievances tweeted for the first time. Unlike other Twitter accounts the structure and content of this account was crowdsourced and provided via an anonymous mechanism. At the heart of the account was a simple premise: Users submit a direct message to the account and, if the message passes a very minimal set of criteria, the message is tweeted by @lis_grievances without any attribution. The process of submitting the direct message is automated and no one, except for the original message sender, knows the author of the tweet; this includes myself, the creator and maintainer of the project. The code used by LIS Grievances is available on GitHub (Ribaric, 2016/2021) and anyone is free to audit the code to ensure that it functions as specified, including the assurance of complete anonymity. The software package used to build the anonymous functionality was written in Python using Tweepy (https://www.tweepy.org/). This automated Twitter mechanism is often referred to as a ‘bot’ (Ferrara et al., 2016). This chapter examines the corpus of tweets generated by this bot over a period of five years, from February 2016 to February 2021. During this time, the @lis_grievances bot tweeted approximately 4,100 times. The analysis that follows is made using both qualitative and quantitative methods, some established and others novel. Quantitatively, basic characteristics of the corpus is explored and the VADER sentiment scoring system is applied. In addition, a novel metric called the ‘grief index’ is applied to the corpus. Qualitative analysis is then applied by examining text of individual tweets that quantitatively score as outliers. The corpus generated by @lis_grievances is fascinating because it represents the output
of a presumably wide range of individuals working within or adjacent to the library environment, unencumbered by possible retaliation for the comments they choose to make.

**The library workplace and anonymity**

In psychology, the ‘online disinhibition effect’ (ODE) first found prevalence during the 1990s when communication via the internet began proliferating (McGlynn, 2020). Put simply, when people interact in an online environment, they are free to adopt different personas from those they actually live with in their daily lives. Such a departure from their realities allows participants to express themselves in ways — and to extremes — that they would not do normally in real life, day-to-day interactions. Several interesting studies provide evidence of how online discourse can polarize and become extreme. For example, Jordan (2019) discussed a case study of how Chelsea Manning sought anonymity during the leak of the Pentagon Papers in an attempt to avoid inevitable backlash. Omernick and Sood (2013) presented a case study that looked at a collection of anonymous comments made on a technology website called TechCrunch.com. Of the results reported by Omernick and Sood, anonymous comments were found to be more extreme and less restrained compared to a control group of comments that were attributed to users and their Facebook accounts.

The ability to interact via the internet in an anonymous manner is a concept that is much debated and heralded as a necessity of modern life. The Electronic Frontier Foundation (n.d.) provided a comprehensive description of why the option of anonymity should be preserved for online communication and what dangers may await average users of the internet if anonymity is not respected. The supposition here is that because a person can act under an alias or assumed identity, their actions will be attributed to this identity instead of themselves. In this way, users
of @lis_grievances may be influenced by ODE and a few factors help accelerate this effect. First, users do not need to create an additional persona, a carefully constructed artifice, to first pre-stage their comments; they merely need to send a direct message to the @lis_grievances account and await a confirmation message letting them know their submission was added to an approval queue. The user is then encouraged to delete their direct message. The bot also deletes its copy of the original message and subsequent replies. As an anti-harassment mechanism, Twitter requires both parties in a direct message to delete the message in question before it is finally removed. With this mechanism built into the process of checking and responding to messages, the @lis_grievances account retains no information about those who submit to it.

With the ubiquity of portable devices such as cell phones and tablets in the hands of potential users, the ability to interact with the bot is frictionless. Submissions can be made on a whim, and once the bot was known to the LIS Twitter community, as measured by a high follower count, potential users know they can quickly air their grievances to a wide audience. As of August 2021, the bot had approximately 2,600 followers. Unfortunately, knowing the exact number of unique users of the bot is not possible, so it is difficult to assess how many individual voices are represented.

The bot software itself has been modified and repurposed by others. The @timetotalkaboutIF Twitter account was created in January 2020 by Jane Schmidt to create a dialog about intellectual freedom in the library (The Incidental Academic Librarian, 2020). This bot was created by copying and configuring the source code of the original project, a process more commonly known as forking. The term ‘fork’ is used here because creating a new version introduces a split or fork in the tree representation of the software linage. In addition, the presence and impact of @lis_grievances has been documented by several practitioners in the LIS
field. A sizable collection of blog posts exists that highlight and analyze discourse released by the bot (e.g., Burns, 2018; Erin, 2016; Lib_idol, 2018; Popowich, 2017; Sarwark, 2018). A particularly astute analysis of the role of @lis_grievances was seen in the following, translated from the original German in Schuldt (2020, para. 8), regarding the lack of public dialogue and criticism about the library profession:

The anonymous Twitter bot LIS Grievances exists as an outlet for such complaints in the English-speaking world. There are some comments here, especially (probably) from colleagues who are forced by their library management either against their own convictions to keep their respective library open or in addition to their actual work, which they already do in the current situation hardly manage, also have to take on special tasks.

Schuldt concluded this observation by linking to a specific @lis_grievance post: “Admins keeping libraries open are 100% on board with their staff and patrons dying. That’s the tweet” (LIS Grievances, 2020), implicitly evoking ‘vocational awe’ (Ettarh, 2018), a divisive notion amongst practicing librarians. Additionally, the bot was written about somewhat positively as a venue for new professionals to communicate with others as they begin their careers in the field (Skyrme & Levesque, 2019) and was featured as a way to address library burnout as a safe way to vent (McHone-Chase, 2020).

Divergent opinions, both positive and negative, exist on the bot’s role and analyzing the dialog that people have with and about the bot by way of Twitter mentions and other social media posts would, indeed, be an interesting endeavor. However, the current chapter is limited to examining only the tweets posted by the bot, not what others say or think about them. This
narrowed scope is necessary to provide insights in the trends and patterns evidenced in the collection of analyzed tweets and those tweets only.

While a collection of tweets generated by the bot provides an interesting corpus to examine, it is worth situating this analysis against a theme. Thus, the corpus is investigated against the supposition of the library as dysfunctional workplace. There exist many prevailing notions and anecdotes that work life in the current library workplace is becoming increasingly problematic. A variety of factors might create this dysfunctional library environment such as the neoconservative tactic of austerity; decreasing funds for public services; lack of meaningful long-term employment, job precarity, and de-professionalization; a general change in work responsibilities as technology automates more processes; poor leadership; and more. Moreover, quantification in specific terms of the effects these trends have is possible (e.g., counting the reduction in permanent full-time positions over a time period); however, it is difficult to exactly describe the level of dysfunction in the profession in a similar way as no instruments or methods have consistently been applied to library workplaces.

Following Allcorn and Stein’s (2015) strategy of capturing stories of abnormal situations in the workplace and applying analytical techniques to the perpetrators of these actions, it is possible that @lis_grievances can be used as a way to measure the dysfunctional state of libraries. Through this method, the dysfunction of the library workplace is measured by the abnormal behavior of those who participate in it. This proposed method is the motivation for the analysis laid out in this chapter. Every single tweet, or a consecutive series of tweets representing a thread, is an anecdote about the current or past experiences of an information professional working in a library environment. By striving to understand tweets in the corpus that display abnormal or otherwise dysfunctional characteristics, the opinions that bot users have of their
workplaces can be constructed and assessed to determine if those workplaces are indeed dysfunctional.

**Corpus generation and VADER**

The corpus itself was generated directly from Twitter via the ‘export my Twitter data’ function which allows an account owner to request the complete authoritative archive of posts and other actions for that account. The archive includes text and metadata of tweets and media objects used in those tweets among various other pieces of information. This information is provided in JSON format. The data for the current investigation was exported on the fifth anniversary of the first tweet made by @lis_grievances, thus comprising the complete archive of the first five years of the life of the bot. The analysis was performed using the Python programming language in a Jupyter notebook environment created by the Anaconda scientific package (https://www.anaconda.com/).

Existing literature is replete with examples of Twitter use and tweet analysis from many different disciplines that employ qualitative and quantitative approaches. One study worthwhile to note showcased the creation of a tool for future researchers to use in the analysis of sentiment found in social media texts (Hutto & Gilbert, 2014). In that study, the authors created a sentiment analysis rubric, known as VADER, that is purposefully trained against a corpus of social media content. The VADER system is robust and able to accurately score the sentiment of posts made on Twitter due to scoring rules that include emoji comprehension and grammar constructs based on the colloquial style of speech found in social media posts. VADER was built through a comprehensive supervised machine learning model which strove to compute a sentiment score. Consequently, the VADER score is represented by three quantities: a score, known as positive
sentiment, from 0 to 1 indicating how strongly positive the sentiment is in a piece of text; the converse measure, known as negative sentiment, from 0 to 1 indicating the negative sentiment of the text; and a composite score ranging from -1 to 1 that maps the sentiment of the considered text on a single spectrum from negative (-1) to positive (+1). This composite score is useful because it allows for a quick calculation to determine if the sentiment of a selected piece of text is neutral or scored at 0. The Natural Language Toolkit available in Python, first described by Bird et al. (2009), included support for calculating VADER scores of text and was the platform with which the analysis in this chapter was conducted.

**Facets of investigation**

To add dimension and differentiation to items in the corpus, two metrics were developed and are presented in the following descriptions.

*Engagement score*

The first measure, ‘engagement score’ (ES), is calculated simply as the sum of the number of retweets and the number of favorites the tweet received during the five-year period. In calculating ES, weighting retweets vs. favorites is not done because there is no information on which is of these is ‘better’ than the other. A retweet broadcasts the content to more users, but in the case of @lis_grievances which relies on anonymity, readers of the bot’s content might not want to engage with it in the same way they do when they know the author (e.g., friends, family, celebrities, news sources, etc.). Favoriting a tweet, versus retweeting it, is more clandestine as it indicates support, amusement, or endorsement of a tweet in a way that does not broadcast to followers quite to the same extent as a retweet. With these considerations, ES treats a ‘favorite’
and retweet as the same. For this analysis, tweets with a non-zero ES are examined more closely than those with an ES of zero. In total, 92% of the corpus has a non-zero ES (n = 3,806). A summary of ES components is shown in Table 1. Identification of ES outliers against other metrics is a recurring theme of analysis in this chapter. A boxplot shows the distributions of means and standard deviations of retweet and favorite counts in Figure 1. The ‘number of favorites’ metric has a higher magnitude of values and demonstrates a larger number of outliers, contributing more to ES.

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Engagement Score</td>
<td>16.7</td>
</tr>
<tr>
<td>Standard Deviation, ES</td>
<td>12.5</td>
</tr>
<tr>
<td>Maximum Engagement Score</td>
<td>737.0</td>
</tr>
<tr>
<td>Average number of retweets</td>
<td>1.2</td>
</tr>
<tr>
<td>Standard Deviation, RT</td>
<td>3.5</td>
</tr>
<tr>
<td>Maximum number of retweets</td>
<td>91.0</td>
</tr>
<tr>
<td>Average number of favorites</td>
<td>15.5</td>
</tr>
<tr>
<td>Standard Deviation, Fav</td>
<td>27.1</td>
</tr>
<tr>
<td>Maximum number of favorites</td>
<td>672.0</td>
</tr>
</tbody>
</table>

Table 1 – Overview of quantitative measures of corpus of tweets (N=3,806)
Figure 1 – Boxplot showing the two components of engagement score (ES) – ‘favorite count’ and ‘retweet count’ – and their distribution over the full corpus of tweets.

With ES defined, categorization of all entries in the corpus against it is possible. Here, a histogram representing the frequencies of ES demonstrates useful trends. In particular, outlier tweets that have scored very high is shown in Figure 2. As the distribution is vast, plotting it along a logarithmic scale is the only way to see the outliers with any clarity.
Figure 2 – Histogram showing the distribution of engagement score (ES) across the full corpus of tweets. The y-axis is logarithmic here to ensure outliers are detectable.

Grief index

A small set of restrictions on content was introduced for anti-harassment purposes as noted on the About page of the LIS Grievances website (https://lisgrievances.com/about.html). However, not all users abided by these stated guidelines. The ‘grief index’ (GI) is a custom metric developed to provide a measure of the variance of illegitimate tweets submitted versus valid tweets published. The calculation of the index is completed monthly using the formula below. This resolution provides a compromise to show motion in the change of value in the index versus frequency of tweets made by the bot. For instance, applying the index at a weekly interval would provide too granular of a data set, while yearly would be too infrequent.
In the formula, GI is an adjusted ratio of how many tweets do not get posted. A higher GI value indicates a ‘worse’ month. The fluctuation of GI over the five-year period investigated is plotted out in Figure 3. In months where GI is zero, all tweets submitted were posted.

\[
\text{GI} = \frac{\text{All Possible Submissions}}{\text{(All Postable Submissions+Non Postable Submissions)}}
\]

(1)

**Figure 3** – Line chart of grief index (GI) as it varies each month over the five years of the full corpus of tweets. The straight line represents the overall average GI score, while the fluctuating line shows the pattern of monthly scores across the entire five-year period.

The overall GI for the five-year period is 8.8. This index is another dynamic metric, along with ES, with which the corpus may be explored. By looking at a set of tweets made during a time period with an above average GI, a snapshot of a controversial segment in the corpus is revealed. By examining the discussion being carried out during the time-period we see the zeitgeist unfold directly. Over the five-year time period, 4,436 total submissions were made to the bot and 4,096 of those were actually tweeted. This ratio is demonstrated in Figure 4, showing that most submissions to the bot were posted; not many needed to be disqualified. Those
submissions that were disqualified, as calculated by GI, occurred in all but 10 months, mostly in 2016. In other words, in the first five years of life for @lig_grievances, a total of 60 months in operation, 50 of those months had submissions made that were unsuitable for posting.

![Pie chart showing the ratio of submissions to the bot that were posted versus those that were not posted](image)

**Figure 4** – Pie chart showing the ratio of submissions to the bot that were posted versus those that were not posted

With ES and GI explained, quantitative analysis proceeds with a basic assumption that any identifiable piece of the corpus that exceeds the ES and GI of similar pieces of the corpus is evidence of an outlier. These outliers indicate which parts of the corpus stand out as unique components deserving of more attention than others. Methodology to identify outliers is well developed and can be computed through vigorous statistical calculation. For the purposes of this investigation, interquartile range (IQR) is used as the general rule of thumb to assess if corpus entries are considered outliers. IQR computes the difference between the third quartile and first
quartile, which is where most values will land, to find the collection of values that extend beyond this interval. Because tweets that stand out from others are of particular interest, looking for high score values to see what those corpus entries possess is the objective.

**Initial impressions of the corpus**

As is custom in most text analysis, an initial review of the corpus can be summarized in the form of a word cloud. Figure 5 represents a word cloud based on the total corpus of text; this is the body of all tweets with a few common stop words removed for clarity. This complete corpus word cloud demonstrates some clear trends in the tweets but also shows a heavy reliance on words such as library, librarian, libraries, librarians, and so on. Because the main motivation of the bot is to provide a venue for library staff to talk about the situations and places they encounter every day, it is understandable that these libra* words would appear. Figure 6 presents an amended word cloud that purposely excludes all words with the libra* stem. Excluding these words from analysis is sensible as their inclusion would impede meaningful analysis due to their high frequency, yielding little additional insight into the corpus.
Figure 5 – Word cloud rendered using all text from the full corpus of tweets

Figure 6 – Word cloud rendered similarly as in Figure 5, but with all forms of libra* words removed

The heavy font size and typeset of the words ‘job,’ ‘work,’ ‘people,’ ‘time,’ and ‘staff’ indicate a high frequency of use for these words in the corpus. This finding can be interpreted as a signal that a prevalent theme in the corpus is ‘library as workplace.’ This preoccupation with the library as a workplace is also reinforced with an examination of word frequency seen in the corpus of tweets. In preparation of analysis along this dimension, the corpus, excluding words that match the aforementioned libra* pattern, was prepared through a process of normalization that included the removal of common stop words (i.e., articles and conjunctions that do not add to meaning of text in which they occur) and removal of punctuation. This normalization process removed approximately 46% of the text from the original corpus. Once normalized, a frequency distribution was calculated. The top 20 most occurring words in the normalized corpus are
provided in Figure 7. The y-axis provides the word and the x-axis shows each time that word appears in the corpus if it were sorted chronologically.

![Frequency distribution graph showing the top 20 words occurring in the full corpus of tweets](image)

**Figure 7** – Frequency distribution graph showing the top 20 words occurring in the full corpus of tweets

Clearly, three of the top four words identified in this frequency analysis (i.e., ‘work,’ ‘people,’ and ‘job’) can be thought of as pertaining to the library as a workplace.

Another interesting phenomenon is the high usage of ellipses. Normalizing the text resulted in the removal of punctuation such as commas, periods, semicolons, and so on. However, a conscientious decision was made to retain all ellipses because of the function that the ellipsis plays in the English language. The *Cambridge English Dictionary* (n.d.) entry of the
word ‘ellipsis’ defines this function: “three dots in a printed text that show where one or more words have been intentionally left out” [italics added]. Users of @lis_grievance are clever with the word count available to them and also with their use of allusions to common scenarios readers of the bot would be able to empathize with or implicitly understand. For instance, the top five ES tweets that contain ellipses are shown in Table 2; all of these score much higher than the average ES of the corpus in total.

<table>
<thead>
<tr>
<th>Full Text of Tweet</th>
<th>Total Engagement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>My coworker asked if other libraries were as dysfunctional as ours. Based on my observations...yes</td>
<td>73.0</td>
</tr>
<tr>
<td>&quot;Second master's degree required...&quot; Fuck you</td>
<td>61.0</td>
</tr>
<tr>
<td>Patron: “WHAT ARE YOU DOING? By quoting the CBC you’re just spreading LIES! EVERYONE KNOWS THAT THE CBC only promotes SOCIALISM!!!” Me: “You’re in a public library...”</td>
<td>50.0</td>
</tr>
<tr>
<td>Why is everything in library-land a &quot;toolkit&quot;? Asking for a friend...</td>
<td>46.0</td>
</tr>
<tr>
<td>People who habitually remain completely silent in all work meetings... Why? How? Do you have no thoughts on what we do and how we do it?</td>
<td>43.0</td>
</tr>
</tbody>
</table>

Table 2 – Top five tweets based on (ES) for tweets containing ellipsis

While frequency is a good indicator of topic seen in the corpus, it does not adequately indicate how ‘even’ words are used in the corpus (i.e., if high use of a word is distributed across
the entire body of the corpus). Figure 8 shows a dispersion plot of selected words in the corpus. Words taken from the frequency distribution is represented by the y-axis and the x-axis indicates every time the word is used as an offset from the first character of the corpus, or said another way, it marks the use of each time the identified word appears. The dispersion of punctuation, however, is not normally calculated which explains the lack of the ellipsis in Figure 8. As shown, users of @lisg_grievances are enamored with talking about the profession and ‘library as workplace,’ and this fascination is consistently seen in the five years’ worth of tweets.

![Lexical Dispersion Plot](image)

**Figure 8** – Lexical dispersion plot of the four most common words in the full corpus of tweets. Plot shows each time the individual word appears if all tweets are ordered chronologically.

A calculation of sentiment was also performed on the corpus of tweets. Sentiment analysis attempts to assign a numerical score to judge the magnitude of how positive, negative,
or neutral the sentiment of a tweet is. In this study, VADER was the scoring mechanism used to analyze social media posts. Figure 9 shows a distribution of sentiment seen in the whole corpus of tweets along three dimensions mentioned earlier that are calculated by VADER: positive sentiment, negative sentiment, and neutral sentiment. This sentiment distribution can be effectively compared with Figure 10 that shows the sentiment distribution of tweets with an above average ES. This comparison indicates a high correspondence between ES and very strong sentiment expression. In other words, tweets that show conviction through a decisive expression of sentiment, either negatively or positively, tended to be engaged with more; that is, followers of @lis_grievances gave more attention to tweets that avoided a neutral stance.

![Pie chart showing the ratio of sentiment for the full corpus of tweets](image_url)

**Figure 9** – Pie chart showing the ratio of sentiment for the full corpus of tweets
Figure 10 – Pie chart showing the ratio of sentiment for tweets with an average engagement score (ES)

Closer examinations of the corpus

To enable a more granular analysis of tweets themselves, several different facets are proposed below.

Hashtag tweets

A mechanism available to draw additional attention to a particular tweet is by including hashtags, allowing a tweet to be included in a broader collocation of tweets that also contain the same hashtag. Tweets with hashtags represented approximately 3.7% of the corpus. A frequency distribution of hashtags that appear more than once in the corpus is presented in Table 3. The average ES of tweets with hashtags is 9.53, well below the general ES average. A histogram plotting the engagement scores of hashtag tweets is presented in Figure 11. Table 4 shows the top three hashtag tweets based on ES.
<table>
<thead>
<tr>
<th>Hashtag</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>#critlib</td>
<td>23</td>
</tr>
<tr>
<td>#lismentalhealth</td>
<td>7</td>
</tr>
<tr>
<td>#poopin</td>
<td>4</td>
</tr>
<tr>
<td>#librarylife</td>
<td>3</td>
</tr>
<tr>
<td>#alaac18</td>
<td>2</td>
</tr>
<tr>
<td>#FridayFeeling</td>
<td>2</td>
</tr>
<tr>
<td>#GTFO</td>
<td>2</td>
</tr>
<tr>
<td>#NationalLibraryWeek</td>
<td>2</td>
</tr>
<tr>
<td>#subtweet</td>
<td>2</td>
</tr>
<tr>
<td>#TFW</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3 – Frequency distribution of top 10 hashtags would in corpus
Figure 11 – Histogram showing the distribution of engagement score (ES) for tweets with hashtags

<table>
<thead>
<tr>
<th>Full Text of Tweet</th>
<th>Total Engagement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>As a person of colour struggling to find a stable job in this field, it gets pretty exhausting watching white folks with permanent librarian jobs talk about how #woke they are and how this field needs to be more inclusive from their cushy perch. It feels like empty words.</td>
<td>70.0</td>
</tr>
<tr>
<td>Can we start normalizing providing feedback to unsuccessful interview candidates? #deathbyathousandrejectionemails</td>
<td>69.0</td>
</tr>
<tr>
<td>Title of my #alaac18 presentation: I’m sorry for this, I just wanted to go to NOLA.</td>
<td>61.0</td>
</tr>
<tr>
<td>Disclosed to a senior leader that I was struggling with WFH full-time and parenting. They said they totally 'get it' and then disclosed that they have hired a nanny and are having an in-ground pool installed to get through these 'difficult times'. #relatableAF 😏</td>
<td>52.0</td>
</tr>
<tr>
<td>Swapping war stories with another person who survived a “culture of fear” workplace is surprisingly comforting and validating. Maybe we should start a support group. #LISsurvivors #recovery</td>
<td>44.0</td>
</tr>
</tbody>
</table>

Table 4 - Top five tweets based on (ES) for tweets that contain hashtags
Hashtags provide an interesting dynamic because they allow people tweeting to have their submission potentially scooped up by a wider audience. For example, if a user includes ‘#critlib,’ a hashtag used by the critical librarianship community, in a @lis_grievances submission, all followers of @lis_grievances would receive the tweet as well as those following the #critlib hashtag. Hashtags, however, are demonstrably underutilized in the corpus; only ten hashtags occur more than once, and, as demonstrated, more than half of these only appear twice, equating to a value of only .2% of the entire corpus. The high use – relatively speaking – of #critlib is the only hashtag with any real presence in the corpus, but with a < 1% total presence in the corpus, it is negligible.

Swear word tweets

Another dimension worth extracting from the corpus are tweets that contain swear words. Swears such as ‘fuck,’ ‘shit,’ and others provide evidence of intense sentiment. Swear word tweets represent approximately 4.7% of the corpus. Average ES of tweets with swear words is calculated at 20.04, including swear word tweets with no ES whatsoever. A histogram of the ES of tweets with swear words is shown in Figure 12. The top five ES swear word tweets are summarized in Table 5. An examination of sentiment of swear word tweets is presented as Figure 13. This figure shows a high proportion of tweets classified as negative, in this case around 70%. Contrast this finding to the distribution of sentiment amongst all tweets shown in Figure 9, where tweets considered negative were only represented by about 37% of the corpus. Tweets with swear words seemingly produced strong sentiment, which is consistent with the supposition that high ES is likely aided by strong sentiment.
Figure 12 – Histogram showing the distribution of engagement score (ES) for tweets with swear words

<table>
<thead>
<tr>
<th>Tweet Full Text</th>
<th>Total Engagement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUT THE FUCKING SALARY IN THE JOB AD!!</td>
<td>737.0</td>
</tr>
<tr>
<td>I want an out of office email for the whole pandemic that just says &quot;shits all fucked up right now, might take me a bit to respond to whatever this is&quot;</td>
<td>249.0</td>
</tr>
<tr>
<td>Library fines are classist bullshit &amp; should be abolished. If you return a book then the fee should be waived.</td>
<td>247.0</td>
</tr>
<tr>
<td>Tweet Full Text</td>
<td>Total Engagement Score</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Conference attendees: USE THE MIC. I don’t give one shit if you don’t think you need it. Presenters: if this doesn’t happen, REPEAT THE QUESTION. This should be non-fucking-negotiable.</td>
<td>136.0</td>
</tr>
<tr>
<td>The first thing they should teach you in higher education to prepare for the workforce is that HR gives absolutely zero fucks about any you and will lie to your face</td>
<td>113.0</td>
</tr>
</tbody>
</table>

**Table 5** – Top five tweets based on (ES) for tweets that contain swear words

**Figure 13** – Pie chart showing the ratio of sentiment for tweets with swear words

*Tweets with media*

Another subset of tweets worth investigating are those with attached media. Media attachment was a feature enabled on the bot on January 10, 2018, just about two years after it
began accepting submissions. To include media in a post, some additional steps for the user were required due to a limit on how information can be acquired through the Twitter API. To use the media feature, a potential poster must upload their media to a hosting site such as Imgur and a link to the hosted media must be included in the direct message sent to the bot. Of the total corpus, only 49 tweets contained associated media, representing only 1.2% of the total tweet volume. These added steps may have discouraged users to pursue media use with the bot. Of those tweets with media, 92% were engaged with in some respect. This value is consistent with ES of the corpus more generally and indicates that the addition of media does not impact ES.

Figure 14 is a histogram of engagement score across all tweets with media. A closer examination of the few outliers in the Figure 14 distribution is revealing. One of these outliers (Figure 15) boasts an ES of 202. Plus, an even higher scored outlier exists (ES = 220) but its tweet could not be shown here because it contained the *Is this a Pigeon?* meme that uses a copyrighted screenshot from the Japanese anime *The Brave Fighter of Sun Fighbird*. In the meme, the character Yutaro Katori gestures to a butterfly; the text “LIS leadership” appears over Katori, the text “Man who is OK at his job” appears over the butterfly, and the text “Is this a genius?” appears across the bottom.
Figure 14 – Histogram showing the distribution of engagement score (ES) for tweets with embedded media
Figure 15 – Example of a media-embedded tweet. This tweet is an extreme outlier according to its engagement score (ES = 202).
**Question-asking tweets**

Another facet of tweet composition is the asking of questions. A question in a tweet can function as a rhetorical device to make the reader understand the situation at hand. A histogram of distribution of ES on question-asking tweets is presented as Figure 16, while sentiment of question-asking tweets is shown as Figure 17. The top five question-asking tweets by ES are shown in Table 6. These tweets are all examples of rhetorical questions facetiously asked by the original poster to help make sense of the dysfunctional library workplace.

![Histogram showing the distribution of engagement score (ES) for tweets containing questions](image)

**Figure 16** – Histogram showing the distribution of engagement score (ES) for tweets containing questions
### Full Text of Tweet | Total Engagement Score
---|---
How can we innovate when we don't have permissions to install software?? | 338.0
Please stop telling professionals in their 40s that they are too young for leadership. How old do you need us to be? | 287.0
Every course, reading, & teacher in my LIS program suggests that librarianship is best learned through an apprentice model - so why am I paying $24k for a professional certification that doesn't give me the skills I need for the job? | 213.0
Fellow academic librarians, can we please stop looking down on public librarians? They're our colleagues, and we're all working for the same purpose. And some folks find encyclopedias quite useful, btw. | 191.0
Why do so many library directors/admin gaslight employees dealing with toxic supervisors/colleagues, instead of taking steps to ensure that toxic individuals are held accountable? Is employee morale and developing a genuinely collegial work environment not high on priorities?! | 168.0

**Table 6** – Top five tweets based on (ES) for tweets that contain questions

### Summary: Engagement scores

In summary, ES of the different facets of tweets investigated are presented as box plots in Figures 18 and 19. These boxplots show that some components of tweets can be readily identified as contributing to above-average ES.
**Figure 17** – Pie chart showing the ratio of sentiment for tweets containing questions

**Figure 18** – Boxplot showing a comparison of engagement scores (ES) of tweets for the categories of all tweets, tweets containing questions, tweets with embedded media, tweets with swear words, and tweets with hashtags
Grief index: A closer investigation

Assessing the impact of GI required a more nuanced and subtle approach than examining means and standard deviations. Assessing the sentiment of tweets in the corpus with an above-average GI versus a below-average GI, as seen in Figures 20 and 21, is further evidence of what has been determined already: Tweets with more definitive sentiment are associated with stronger reactions in the corpus.
**Figure 19** – Similar boxplot as Figure 18 but with two changes: 1) the ‘all tweets’ category is replaced with the category of tweets with engagement by users, and 2) outliers according to engagement score (ES) are shown.

**Figure 20** – Pie chart showing the ratio of sentiment for tweets where the grief index (GI) is lower than the accumulated GI average.
Corpus analysis revealed three occasions where GI spiked considerably with exceptionally high values. Chronologically, these extreme values occurred in August 2018 (GI = 18.75), June 2020 (GI = 43.64), and July 2020 (GI = 20.34). For these three months, the top five tweets according to ES are presented in Tables 7, 8, and 9. Extrapolation to determine characteristics of the corpus during these three tumultuous months is difficult by virtue of the definition of GI, but some observations can be made at least for those two months in 2020. During that summer, activity surrounding @lis_grievances was polarizing due to the COVID-19 global pandemic when library workplaces were in the middle of unprecedented disruption and disarray.
**Full Text of Tweet** | **Total Engagement Score**
---|---
I smuggled a tea kettle into my office today because we are no longer allowed to use a nearby kitchen/break area because now it's for faculty only. Tomorrow: rice cooker. Next week: mini fridge. | 39.0

Woke white librarians who leverage their wokeness to further their careers and professional development. | 27.0

If someone says you have to leave in 10 minutes and it takes you 10 minutes to pack your shit, the next thing you should do is pack your shit. #closingtime | 26.0

People say some real dumbass shit about the value of libraries and librarians. | 25.0

The amount of bullshit we have to go through because we can't just say, "Hey, stop being a fucking bitch already, everyone is sick of it." | 21.0

**Table 7** – Top five tweets based on (ES) for August 2018, where (GI) was 18.75

---

**Full Text of Tweet** | **Total Engagement Score**
---|---
An employee at another branch tested positive for COVID-19. There are no plans to close the branch and no plans to tell the public. | 151.0

Whole lot of middle-aged white ladies in the staff zoom meetings crying about all the "soul searching" they've done over the weekend. It's so tragic how police brutality was invented in 2020. | 63.0

Anti-grievance: when you say something you're sure someone more senior than you is going to argue with you over and they back you up instead. 😍 | 54.0

During my 15 years in libraries I've found that the lack of defined duties occupies a good 1/4 to 1/2 of people's time. "Who does that now?" | 41.0

all the responsibility but none of the authority: not a good place to be | 38.0

**Table 8** - Top five tweets based on (ES) for June 2020, where (GI) was 43.65

---

**Full Text of Tweet** | **Total Engagement Score**
---|---
Normalize not expecting librarians and library staff to double as social workers 👏 | 463.0

Every course, reading, & teacher in my LIS program suggests that librarianship is best learned through an apprentice model - so why am I | 213.0
<table>
<thead>
<tr>
<th>Full Text of Tweet</th>
<th>Total Engagement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>paying $24k for a professional certification that doesn't give me the skills I need for the job?</td>
<td></td>
</tr>
<tr>
<td>No cops in libraries</td>
<td>100.0</td>
</tr>
<tr>
<td>6 months ago: We don't want there to be hundreds or thousands of LibGuides for everything. Now: We went ahead and made a LibGuide for every class this summer and fall. Then we made a LibGuide about why we made these LibGuides.</td>
<td>77.0</td>
</tr>
<tr>
<td>how does anyone furlough youth services staff when all over the country libraries are about to have hundreds of new homeschooling families who need help getting books and remote access to learning resources like a month from now</td>
<td>67.0</td>
</tr>
</tbody>
</table>

Table 9 - Top five tweets based on (ES) for July 2020, where (GI) was 20.34

Conclusion

Much has been published in social media analysis studies. The preceding investigation, however, is a novel contribution to existing studies due to the inherent limitation in the dataset and how it was constructed. By having totally anonymized author information, attribution of tweets in the corpus is impossible, limiting some avenues of investigation (e.g., comparing the sentiment of certain demographics of authors). To make up for this lack of identifiable markers, an analysis based solely on the corpus itself is presented. By introducing and developing engagement score (ES), grief index (GI), and partitioning the corpus along identifiable facets, insights are gained into what the corpus tells us about @lis_grievances users. In summary, it is fair to say that users are preoccupied with talking about the library as a workplace, invoke strong sentiment when describing situations, and are reluctant to retweet content posted by the bot.

Perhaps a conclusive summary of the dialog produced in the corpus of tweets is best achieved by examining the single tweet with the highest ES of 737. This tweet (Figure 22)
provides quite a salient snapshot of shared dysfunction in the library as a workplace, and does so with a flourish:

Figure 22 – The tweet most demonstrative of the corpus

This chapter is a pre-print; the final version appears in Libraries as Dysfunctional Organizations and Workplaces (publisher: Routledge, editor: Spencer Acadia).
Endnote

The name and other details of the *The Brave Fighter of Sun Fighbird* meme was determined using the Know Your Meme website. The meme could not be published here due to copyright concerns.

References


The Incidental Academic Librarian. (2020, January 13). #timetotalkaboutIF campaign [Blog post]. The incidental academic librarian: Because, as it happens, academic freedom is a terrible thing to waste. https://incidentalacademic.wordpress.com/2020/01/13/timetotalkaboutif-campaign/


LIS Grievances [@lis_grievances]. (2020, March 23). *Admins keeping libraries open are 100% on board with their staff and patrons dying. That’s the tweet.* [Tweet]. Twitter. https://twitter.com/lis_grievances/status/1242058520549142530

McGlynn, J. (2020). Disinhibitory and inhibitory effects. In D. L. Merskin (Ed.), *SAGE international encyclopedia of mass media and society* (pp. 495–496). SAGE.


