Bandwagoners and diehard fans, are they different? Linguistic sentiment and team identification differences between fans of varying loyalty levels in online sports discussions

Qi Pan
College of Information and Computer Sciences
University of Massachusetts Amherst
Amherst, MA 01002, United States of America
qipan@umass.edu

Abstract

From fanatics to casuals, traditional sports fans show varying levels of loyalty and devotion to their team through large signs, face paint, and loud cheers. But after the game has settled down, does their level of loyalty manifest itself in the linguistic characteristics of their online discussions? This paper shows that high loyalty level fans use more possessive language than low loyalty level fans (e.g. "we won" or "I hope we can recover from this loss") whereas low loyalty fans use more positive language across all contexts. Furthermore, the differences in these linguistic measurements are exaggerated within groups of fans who follow a lower ranked team or when the comments come directly after a game loss. Lastly, sentiment differences between loyalty groups are the largest under contexts that mentioned both the posters own team and/or other teams.

1 Introduction

Traditional sports fans display varying levels of loyalty to their team, ranging from a bandwagoner (someone who only supports the team when they succeed) to diehards (someone who absolutely supports their team regardless of exterior circumstances). **This paper will investigate whether or not there are significant linguistic differences exhibited in online sports discussions between fans of varying loyalty levels.** Specifically, this paper focuses on Reddit user comments from the `/r/nfl` subreddit, an online discussion board where users can talk about the National Football League (NFL - the largest professional league of American Football in America). These questions are important in understanding whether or not ones loyalty to a sports team can manifest as specific linguistic characteristics in their everyday online discussions. Besides scientific curiosity, understanding this phenomena can aid the million dollar sports organizations of the NFL in marketing to fans of varying loyalty levels by using tailored language in marketing campaigns.

The effect of various factors unique to traditional sports will be controlled for to isolate the effect of fan loyalty, but also to add another dimension to the analysis of possible reasons for linguistic differences (e.g. team performance). The first linguistic marker investigated is how much a fan associates themselves with their team (**team identification**). This paper will use methodology found in (Cialdini et al., 1976), where team identification was quantified by measuring how often a fan would say the word "we" when discussing football games. The second linguistic marker is positive/negative **sentiment** which will be measured using VADER, a rule-based sentiment analyzer designed for social media text (such as our Reddit dataset) (Gilbert, 2014). Loyalty will be measured by using the fact that each user on the `/r/nfl` may assign a flair to themselves, which is an icon/avatar of one of the 32 NFL teams. Assuming that having a teams flair at time t means you are a fan of that team at time t, then loyalty can be measured as the consecutive amount of time one has had that flair up to t (no flair switch in-between). A variety of hypothesis will be tested to see if there are potential differences between fans of varying loyalty levels in terms of these two linguistic markers.

2 Data Preprocessing

2.1 Aligning comments with game results

At a high level, this paper will limit analysis to comments made during the 2016-2017 NFL regular season **between games**. Therefore, for any fan of a given team, any of their comments which occur on the day of their teams' games will be discarded. Comments are then associated with the last game result (win, loss, tie) of their team so that later analysis can take into account whether the comment was made by a user who just "experienced" a win or loss.

More formally, let t_{ij} be the continuous chunk of time that starts from the day after team j's ith game to the end of the day before team j's $(i+1)$st game in the PST timezone (the timezone in the USA which "lags" behind the most and therefore the last American fans who enter a new day). The one day buffer between t_{ij} and t_{i+1j} ensures that comments made during games are not considered (see fig. 1). Semantically, this means that comments in the time range t_{ij} are created by fans who only have knowledge about the prior i games and are not posted with partial game state information about the $(i+1)$st game. The buffer is important because fan emotions can fluctuate wildly during a game (Yu and Wang, 2015) affecting measures of sentiment and team identification. Thus, the buffer prevents a blending of comments where the fan has partial game state information (e.g. the score at half time of game $i+1$) about potential future results and comments where the fans don't. Each team plays 16 games in the NFL regular season and there are 32 teams. In the case of $t_{i=16,j}$ where there is no upper bounding game, let that interval last 6 days from the day after team j's 16th game (a typical NFL break between games is a week).

The initial comment data pool consists of all `/r/nfl` subreddit posted during [September 2016, February 2017] inclusive (months of the 2016-2017 NFL regular season). The comments were extracted from `pushshift.io` (the data was made available by Reddit user (Stuck_In_the_Matrix)). These comments were filtered down to only those which fell into one of the previously defined time ranges $t_{ij} \quad \forall i = 1, \ldots, 16 \quad \forall j = 1, \ldots, 32$. Since each comment and game can be associated with a t_{ij}, comment and game data can be inner joined on t_{ij} resulting the data schema shown in table 1.

Before calculating any of our linguistic marker metrics, the alignment of comments relative to the definition of t_{ij} was checked. As defined by t_{ij}, no comment for a given team should be kept if it was posted on the day of one of the team's games (in terms of the PST timezone). This was verified by plotting the game day intervals as thick lines against the time of comments as dots on a timeline to see if they overlapped. This is shown in fig. 2 for the team with the most comments from fans, the New England Patriots.

Field	Description
UTC	The UTC timestamp of the comment
Team (j)	The chosen flair (symbol of an NFL team) of the poster and therefore their associated team
Game Order (i)	The game order (1st through 16th) of team j associtated with this comment
Result	The result of the game associated with this comment (W,L,T)
User	Unique user id of the poster
Body	The comment text itself

Table 1: Data schema after joining comment data with game data

2.2 Calculating loyalty

Loyalty is measured for each unique user as the consecutive amount of time they have had their teams flair up to t_{start} = 00:00am PST of the first game of the season, this field will be called **Years Existed Back**. In the case of the 2016-2017 NFL season, t_{start} would be September 8th 2016 00:00am PST. Due to time and computational constraints, loyalty is estimated with the following procedure:

1. Extract all comments for only the month of September in the `/r/nfl` subreddit for the years 2010-2015 inclusive

2. For each user that exists in our 2016-2017 NFL regular season dataset (described in the previous section), find the year furthest back in which they have commented in the month of September with the same team flair that they have at t_{start}

   (a) If the user has not commented in any September between 2010-2015 with the same team flair, then set their Years Existed Back = 0

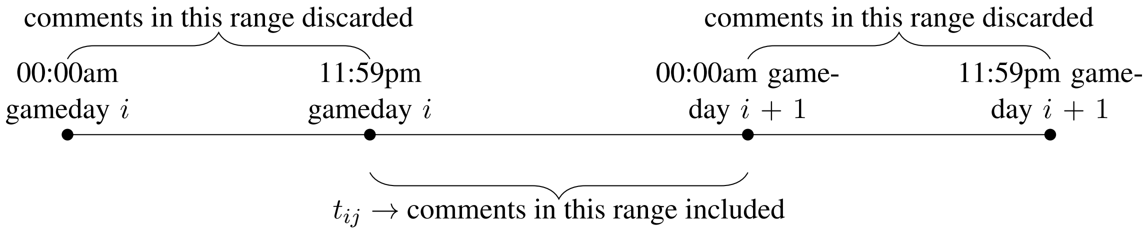

Figure 1: The time between games is precisely the interval t_{ij} shown in this diagram $\forall i = 1, \ldots, 16$ and $\forall j = 1, \ldots, 32$

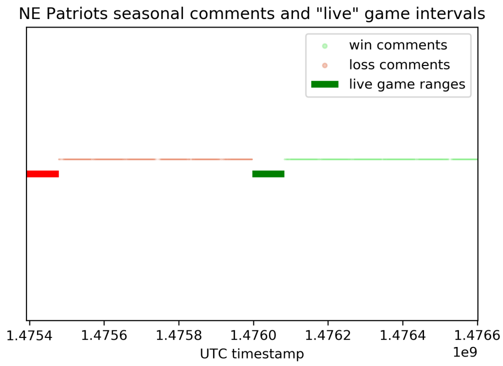

Figure 2: Zoomed in 2 week range of comments from New England Patriot fans: The lower thick lines (color indicates if the game was a win (green) or loss (red)) show game day intervals where comments should not be, the higher dots indicate when comments were actually made (color indicates if the comment was associated with a win (green) or loss (red)) $\Rightarrow$ since they do not overlap and win/loss of comments have been properly assigned to the last game's results the alignment appears to have been correct

(b) Else, set their Years Existed Back as the difference between 2016 and the year furthest back in which a comment was found with the same team flair (e.g. if the last September comment a user made with the New England Patriots flair was during September 2013, and this user also had the New England Patriots flair at t_{start}, then Years Existed Back $= 2016 - 2013 = 3$

2.3 Calculating linguistic metrics

The final preprocessing step is to calculate the relevant linguistic metrics for the body text of each comment. A summary of these metrics can be found in table 2.

Sentiment (s) is measured with the VADER implementation available on NLTK applied to the comment body text. "We" percentage (w) is measured using simple string matching for the presence of "we" in the body of the comment. To find mentions of the 32 NFL teams, a dictionary is created which will include the official team name, the official city/area of the team, and other slang terms for the team (e.g. Patriots, New England, pats) and match these phrases to comment body text irrespective of capitalization and ending /s/'s. Additional slang terms were sourced from (Wikipedia). This dictionary is used to compute the o and f binary tags described in table 2. Additionally, f was set to 1 if w was 1 since using the word "we" without directly referencing the team name is still talking about ones own team.

2.4 Controlling for overall team performance

To control for the affect team performance has on a user's posting behavior, one could further divide users by how well their team does in the regular season. Teams are ranked by their end of season win percentage, where tied teams are given a mean

Name	Var	Description
Sentiment	s	A positive/negative sentiment score between $[-1, +1]$ of the post based on VADER, a rule based model as described in (Gilbert, 2014)
Percentage "we"	w	For a given comment, 1 if it contains the word "we" and 0 otherwise
Flair Team Mention	f	For a given comment, 1 if it contains a mention of the users flaired team and 0 otherwise
Opposing Team Mention	o	For a given comment, 1 if it contains a mention of a team that is not the users flaired team and 0 otherwise

Table 2: Comment level measurements

rank (e.g. if teams are 4 teams are tied for 3rd, then they all get a rank of mean(3rd, 4th, 5th, 6th) = 4.5th). This rank will be referred to as **Win Rank**.

3 Summary statistics

Before diving into the analysis, it is important to get a feel for the dataset through summary statistics of comments made during the 2016-2017 NFL season as defined by our t_{ij}'s from prior sections. Since the analysis will always be comparing users of varying loyalty levels, it is a good idea to know the sample size in terms of comments and unique users for each loyalty level (as measured by Years Existed Back = YEB). The total number of comments, unique users, and users/comment grouped by YEB is shown in fig. 3.

As expected there are more low loyalty users and comments than high loyalty ones but their commenting activity is roughly uniform on average (comments/user does not seem to be influenced by loyalty as measured by Years Existed Back (YEB)). Notably however, the amount of users and comments decreases quite drastically with increasing YEB; this may reduce the statistical power of hypothesis testing when considering differences between YEB groups. To alleviate this, the comments will be grouped into "low" and "high" loyalty groups as defined by (1).

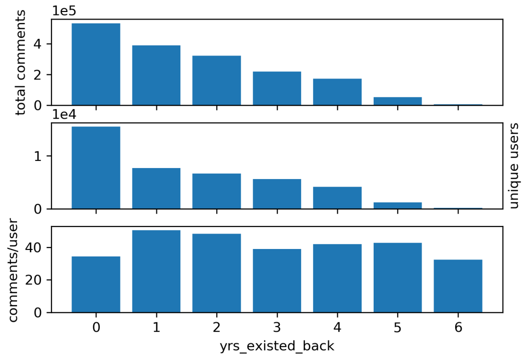

Figure 3: The total number of comments, unique users, and comments/user grouped by YEB on the `/r/nfl` subreddit during the 2016-2017 NFL regular season

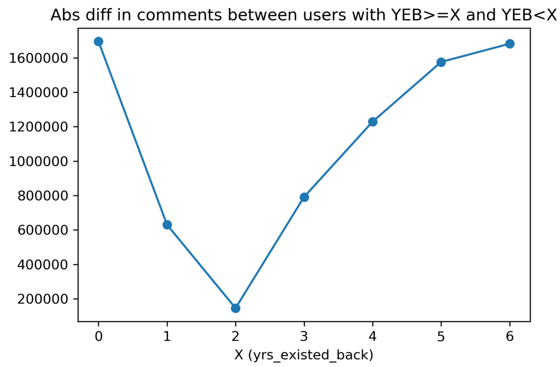

Figure 4: Absolute difference of total comment samples with varying X between $\mathcal{L}_{loyal}$ and $\mathcal{H}_{loyal}$

$$\begin{aligned} \mathcal{C} &= \{\text{set of all comments}\} \\ \mathcal{L}_{loyal} &= \{c \in \mathcal{C} | c[\text{YEB} < X]\} \\ \mathcal{H}_{loyal} &= \{c \in \mathcal{C} | c[\text{YEB} \geq X]\} \end{aligned} \tag{1}$$

X will be set to the integer YEB which minimizes the absolute difference of total comment samples between $\mathcal{L}_{loyal}$ and $\mathcal{H}_{loyal}$. This procedure creates a binary grouping where the comment sample size is the most balanced between the two groups and hence, provides the best statistical stability when estimating population parameters for each group. The difference is plotted in fig. 4, the optimal threshold is $X = 2$. Using this threshold, the statistics from fig. 3 are recalculated but grouped by $\mathcal{L}_{loyal}$ and $\mathcal{H}_{loyal}$ ("low" and "high" **Loyalty Groups** respectively), this is shown in fig. 5.

The same procedure is done but in terms of team performance instead of YEB as dividing samples

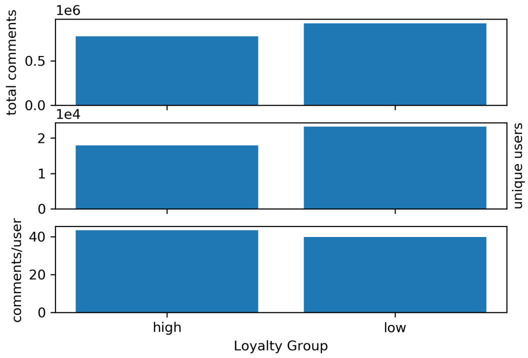

Figure 5: The total number of comments, unique users, and comments/user grouped by Loyalty Group on the `/r/nfl` subreddit during the 2016-2017 NFL regular season

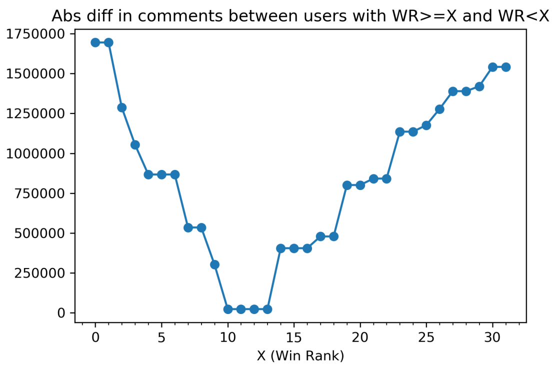

Figure 6: Absolute difference of total comment samples with varying X between $\mathcal{L}_{rank}$ and $\mathcal{H}_{rank}$

up by just team leads to very imbalanced buckets (the largest group of fans is 10 times larger than the smallest). A grouping of comments by **Rank Groups** for an integer Win Rank X is defined by (2). The difference in total comments with a varying X is plotted in fig. 6 and the optimal threshold is 13. Using this threshold, the overall comment and user statistics for the two groups are shown in fig. 7 ("high" and "low" Rank Groups refer to $\mathcal{L}_{rank}$ and $\mathcal{H}_{rank}$ respectively).

$$\begin{aligned} \mathcal{C} &= \{\text{set of all comments}\} \\ L_{rank} &= \{c \in \mathcal{C} | c[\text{Win Rank} \geq X]\} \\ H_{rank} &= \{c \in \mathcal{C} | c[\text{Win Rank} < X]\} \end{aligned} \tag{2}$$

It is also reasonable to look at posting rate throughout the season as it could have an impact on the sample size of comments and users when considering hypothesis that are applied per t_{ij}.

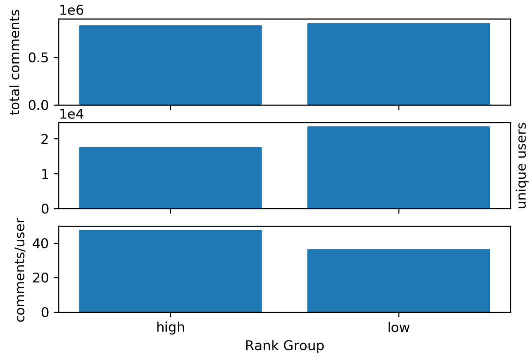

Figure 7: The total number of comments, unique users, and comments/user grouped by Rank Group on the `/r/nfl` subreddit during the 2016-2017 NFL regular season

The number of comments/user is plotted for each Rank Group across the games of the season in fig. 8. Users from the low rank group ($\mathcal{L}_{rank}$) generally post ≈ 1 comment less per user than those in the high rank group ($\mathcal{H}_{rank}$) after the 3rd game of the season. Both groups have a spike in discussion after the season completes. Possible explanations for these trends can be:

- Once high rank fans realize their team is winning, they become more excited to discuss about their teams
- Once the season ends, losing fans can discuss potential future directions (e.g. who was to blame, who to fire, who to trade for, etc.) while winning fans get excited about how far they can get in the playoffs

The validation of these reasons as causes of the observed trends is beyond the scope of this paper. It is simply important to note that comments from $\mathcal{L}_{rank}$ have a slightly lower sample size than comments from $\mathcal{H}_{rank}$ but that both groups of comments are fairly stable in sample size throughout the season. This is good news for statistical analysis on per t_{ij} intervals as it means the sample sizes should be fairly consistent for any t_{ij} (where Game Order represents i and Rank Group is an aggregation proxy for team j).

4 Methods (hypotheses to test)

As a reminder, $\mathcal{H}_{loyal}$ and $\mathcal{L}_{loyal}$ are the set of comments made by fans in the high loyalty group (YEB ≥ 2) and low loyalty group (YEB < 2) respectively. Similarily, $\mathcal{H}_{rank}$ and $\mathcal{L}_{rank}$ are the set

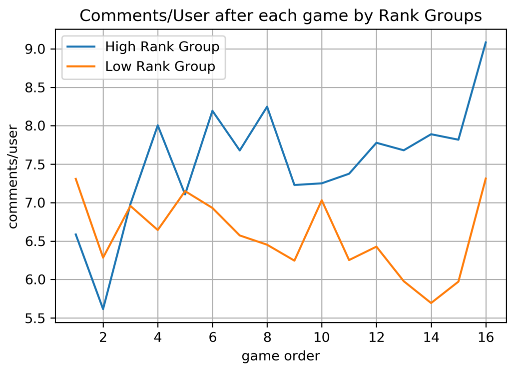

Figure 8: Comments per user across the 2016-2017 NFL season by game order; each line represents the aggregation of comments and users by Rank Group

of comments made by fans in the high rank group (Win Rank ≥ 13) and low rank group (Win Rank < 13) respectively.

An analysis of American Football fans in (Tapp, 2004) hypothesized that highly loyal fans may self identify with their club more than those of lower loyalty. This hypothesis is formally tested by looking at the "We" proportion of comments (w) as described in H1. In terms of linguistic sentiment, prior work was not found in terms of what hypothesis could be formed as linguistic sentiment differences between fans of different loyalty groups. Therefore, H2+ will first explore if there are any differences between $\mathcal{H}_{loyal}$ and $\mathcal{L}_{loyal}$ at all and then seeing in which direction the difference leans.

4.1 "We" proportion hypothesis

H1: Over the course of a season, loyal fans are more likely to associate the team with themselves than less loyal fans when controlled for the performance of their team $\Rightarrow$ For a given Rank Group $\mathcal{R} \in \{\mathcal{L}_{rank}, \mathcal{H}_{rank}\}$, the average of w for comments in $\mathcal{H}_{loyal} \cap \mathcal{R}$ will be higher than the average of w for comments belonging to $\mathcal{L}_{loyal} \cap \mathcal{R}$.

Let $\hat{p}_{\mathcal{H}}$ be the sample proportion of comments where $w = 1$ in the comment set $\mathcal{H}_{loyal} \cap \mathcal{R}$ and $\hat{p}_{\mathcal{L}}$ be the sample proportion of comments where $w = 1$ in the comment set $\mathcal{L}_{loyal} \cap \mathcal{R}$. Also let $n_{\mathcal{H}} = |\mathcal{H}_{loyal} \cap \mathcal{R}|$ and $n_{\mathcal{L}} = |\mathcal{L}_{loyal} \cap \mathcal{R}|$. The null and alternative hypotheses are defined as (3) (where $p_{\mathcal{H}}$ and $p_{\mathcal{L}}$ are the true population proportions across all online discussion remarks across all platforms on the internet):

$$\begin{aligned} H_0 &: p_{\mathcal{H}} = p_{\mathcal{L}} \\ H_a &: p_{\mathcal{H}} \neq p_{\mathcal{L}} \end{aligned} \tag{3}$$

Assuming that Central Limit Theorem (CLT) holds given that our sample sizes are fairly large (in the realm of 1e5), then the distribution of $\hat{p}_{\mathcal{H}}$ and $\hat{p}_{\mathcal{L}}$ should be Normal and hence we can perform a Z test on their difference (which will also be Normal since the sum/difference of two Normal random variables is Normal). This allows the use of the Z-test where the test statistic Z is defined in (4).

$$Z = \frac{\hat{p}_{\mathcal{H}} - \hat{p}_{\mathcal{L}}}{\sqrt{\frac{\hat{p}_{\mathcal{H}}(1-\hat{p}_{\mathcal{H}})}{n_{\mathcal{H}}} + \frac{\hat{p}_{\mathcal{L}}(1-\hat{p}_{\mathcal{L}})}{n_{\mathcal{L}}}}} \tag{4}$$

4.2 Sentiment hypotheses

Although VADER measures the comments sentiment on a positive/negative axis, it does not summarize anything about the context of the comment. For example, a negative comment made about one's own team could mean frustration about their play while a negative comment made about another team could simply be putting that team down after one's own team crushed them in a game. Therefore, sentiment score differences will be considered for a controlled context. Context for the purposes of analysis in this papers refers to varying the binary fields of game result (Result), self team mention (f), and opposing team mention (o) described earlier in table 1 and table 2.

H2.00: When considering comments that do not mention ones own team and do not mention any other teams, the comment sentiment is significantly different between high and low loyalty fans regardless of the prior game's result or the overall seasonal performance of the poster's team $\Rightarrow$ For a given Rank Group $\mathcal{R} \in \{\mathcal{L}_{rank}, \mathcal{H}_{rank}\}$, Result $\mathcal{S} \in \{\text{Win}, \text{Loss}\}$, and for $f = 0, o = 0$, the average of s is significantly different between comments coming from $(\mathcal{H}_{loyal} \cap \mathcal{R} \cap \mathcal{S} \cap f \cap s)$ and $(\mathcal{L}_{loyal} \cap \mathcal{R} \cap \mathcal{S} \cap f \cap s)$; let these sample sentiment averages be $s_{\mathcal{H}}$ and $s_{\mathcal{L}}$ respectively.

H2.01: Same as H2.00 but for $f = 0, o = 1$

H2.10: Same as H2.00 but for $f = 1, o = 0$

H2.11: Same as H2.00 but for $f = 1, o = 1$

In each H2+, the null and alternative hypotheses are defined by (5).

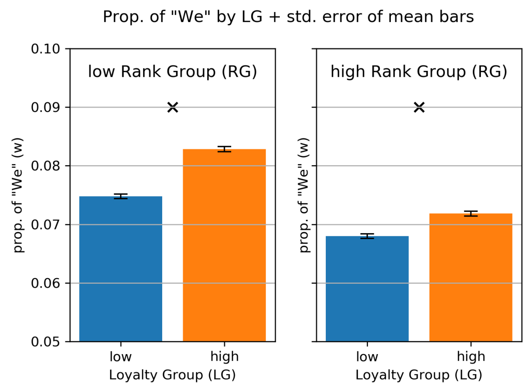

Figure 9: The proportion of "We" by Loyalty Group and Rank Group with standard error of mean bars. X's indicate statistical significance under 99% confidence

$$H_0 : s_{\mathcal{H}} = s_{\mathcal{L}}$$
$$H_a : s_{\mathcal{H}} \neq s_{\mathcal{L}} \tag{5}$$

Again, we can assume CLT since our sample sizes are large (lowest group has sample size $>$ 2800) and perform hypothesis testing using a Z test.

5 Results

5.1 "We" proportion hypothesis results

The raw w differences between high and low Loyalty Groups for a given Rank Group is shown in fig. 9 with standard error of mean bars. **In both control cases, the high Loyalty Group had a higher proportion of w indicating that indeed, high loyalty level fans tend to have higher team identification than low loyalty level fans when posting in online discussion boards.** The differences for both $\mathcal{R} = \mathcal{H}_{rank}$ and $\mathcal{R} = \mathcal{L}_{rank}$ were statistically significant at a 99% confidence level obtaining p-values of $7.84\text{e}{-12}$ and $1.21\text{e}{-43}$ respectively.

5.2 Sentiment hypotheses results

The s differences between high and low Loyalty Groups (LG) under different combinations of o, f, game results, and Rank Groups (contexts) are shown in fig. 10 with standard error of mean bars. Overall, low Loyalty Group (LG) posters are more positive than high LG posters across all contexts. In fact, if there is a difference between LGs the low LG posters are always the one with a higher sentiment (s). Loyalty level seems to be a statistically significant factor in determining comment sentiment across most contexts (11/16 contexts have a statistically significant s difference between LGs under a 95% confidence level as indicated by the x's). Some more context specific observations are:

	$o = 0$	$o = 1$
$f = 0$	0.0057	0.0088
$f = 1$	0.0120	0.0126

Table 3: Average differences in s between low LG and high LG (calculated as low-high) for each of the quadrants shown in fig. 10

- There is a larger sentiment difference between LGs right after a loss (Result="L") where the average difference in s between LGs is 0.0142 on average as opposed to 0.0054 after a win.
- There is a larger sentiment difference between LGs within worse performing teams (low Rank Group) where the average difference in s between LGs is 0.0162 on average as opposed to 0.0034 for high Rank Group posters
- The sentiment differences between LGs become more exaggerated as the context shifts from no team mentions ($f = 0, o = 0$), top left quadrant) to comments containing discussion of both ones own team and other teams ($f = 1, o = 1$, bottom right quadrant), the average difference in s across the 4 quadrants can be found in table 3

6 Validation

6.1 "We" as a proxy of team identification

To validate that w is actually a good linguistic marker for fan association with their team (team identification), 40 comments that were labelled as $w = 1$ were randomly sampled from the dataset and manually assessed. A comment was labelled as correct if the comment specifically used the word "we" to indicate association with the poster's team. The comment was labelled as incorrect otherwise. The results are shown in table 4 $\Rightarrow$ 77.5% of the sampled comments used the word "we" correctly to indicate team identification. It is therefore reasonable to conclude that w is a decent linguistic proxy of team association with roughly 77.5% precision.

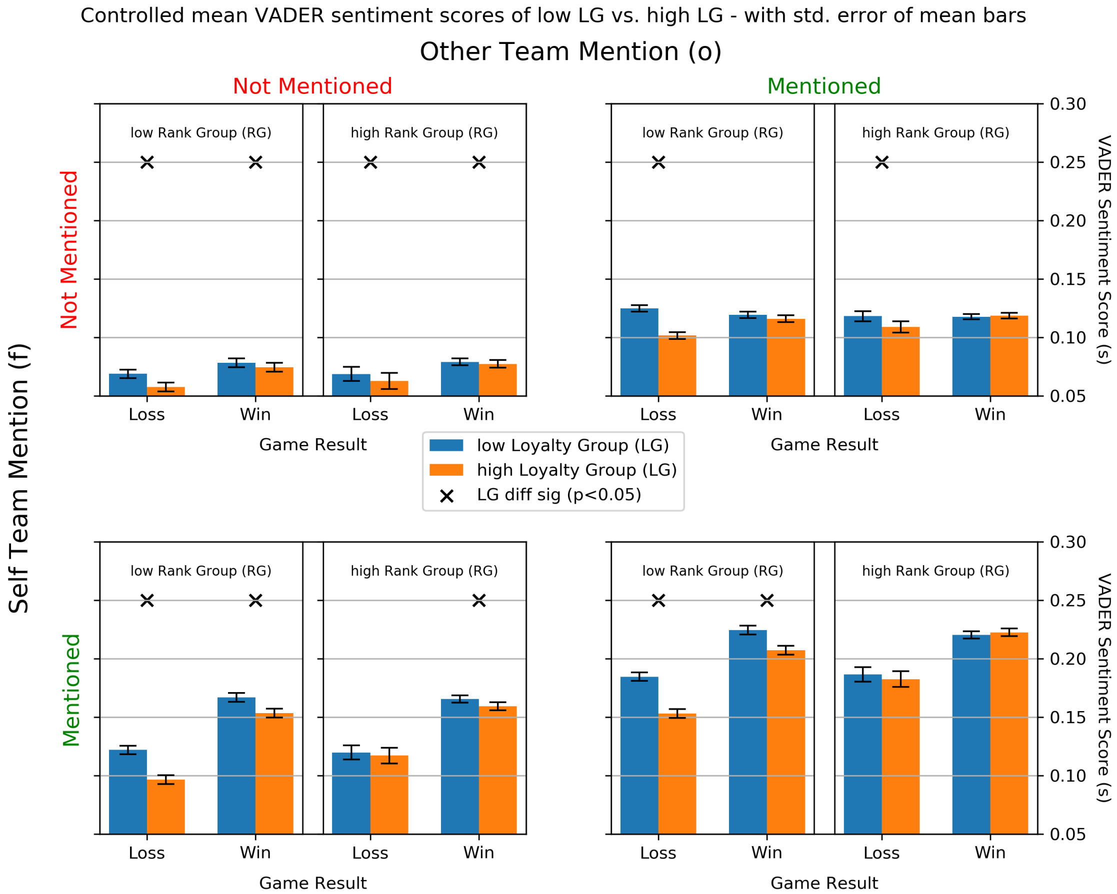

Figure 10: Mean VADER sentiment scores of low and high LG posters under various contexts. Each quadrant represents whether or not ones own team or another team was mentioned, and within each quadrant sentiment means are further separated by poster RG and the last game Result of the poster with respect to the time of the comment

	body	Proper We Usage
0	So do we	1
1	Did you even read the article?is what...	1
2	That's what we tell ourselves about the Schaub...	1
3	That is 100% true. That's why we didn't keep ...	1
4	> Dak isn't the reason this team is winning...	1
5	We had 2.1 points per drive, you put up 1.5.	1
6	Hey guys can we get like an inside joke in thi...	0
7	No we would get draft picks that will help is ...	1
8	I'd take Minnesota. I think we have better dep...	1
9	We got u fam	0
10	ahhh crap i didn't think this through.g...	1
11	And honestly it's looking great for you guys s...	1
12	If the Colts' injured DB's can return at ant p...	1
13	Oh fuck off lol. The Bears are way better than...	1
14	There we go skins	0
15	Yeah, teams bitch about their oline and you co...	1
16	God if we could pick a quarterback like Staffo...	1
17	If it were top 5 best players, I'd agree, but ...	1
18	Woah now, we aren't the raiders.	1
19	We are officially in the running for worst OL	1
20	And we made it about the patriots once again. ...	1
21	OK yes he's having a good season so far. How a...	0
22	Can we get Dak a necklace that says Dady, a la...	0
23	Dak's had a couple short relationships, becau...	0
24	I was skeptical at first but from what I've he...	1
25	He didn't really fall, we had to sign guys bec...	1
26	He's always been good. He had 788 yards and 4 ...	1
27	Oh man, he also said how we respond to 1-3 "st...	1
28	We no. We all seen it.	0
29	Whats funny was that if you watch the play, ro...	1
30	As the team that knocked the Raiders out of th...	1
31	That same year he missed 3 field goals @ home ...	1
32	I mean we do have Harrison. The man has payed ...	1
33	Goddamn this hurts. I really thought the worst...	1
34	The more I read stuff like this, the more I ac...	1
35	I'd be pretty upset as well If I made a play t...	1
36	> the **perceived** likelihood, ...	0
37	We appreciate the lucky break you gave us.	1
38	HOW THE FUCK ARE WE NOT ELIMINATED FROM THE PL...	1
39	We all make mistakes.	0

Mean: 31/40 = 0.775

Table 4: 40 random samples from $w = 1$ comments manually evaluated to see if the use of the word "we" was indeed used to indicate ones own team (Team association). Full comments can be found in the Jupyter notebook under the Verification section.

6.2 VADER as a measure of sentiment for the /r/nfl subreddit

Since VADER was created for Twitter data, it is a good idea to see if it's positive/negative sentiment measure works well on this paper's dataset from the `/r/nfl` subreddit. 40 random comments were sampled from the pool of all comments with a VADER score greater than the 75th percentile VADER score and manually assessed to see if they correctly represented positive comments. Similarly, 40 random comments were sampled from the pool of all comments with a VADER score less than the 25th percentile VADER score and manually assessed to see if they correctly represented negative comments. Using these percentile thresholds was an attempt to eliminate "neutral" comments to see if the more extreme comments were identified correctly (since these comments have the highest weight in the mean calculations of the results section).

The results for each verification task can be seen in table 5 and table 6. VADER had a low precision when identifying positive comments with an empirical sampled precision of 47.5%. Through manual inspection, the issue often revolves around VADER giving neutral commentary posts a very high VADER score. For example, comment 14 says:

"While what he said is false (not this year), the Pats went 3-1 while he was suspended. Willing, Montana won when his team went 3-0 with him missing, but we have to ask if Brady is most valuable if a player like Matt Ryan brings his team to the playoffs, which seems very likely. Does Brady deserve consideration over a guy like Ryan if they keep up the performances they're currently on?"

This comment was given a very high sentiment score by VADER of 0.8954 but the comment itself is more neutral commentary. Since neutral/critical commentary is an integral part of discussing sports, then clearly future work that involves sentiment analysis of the `/r/nfl` dataset should be trained/designed to label such commentary as neutral (close to 0) instead of overly positive. VADER as it stands clearly does not get a high enough precision for the positive comment case. In the negative comment case, VADER had a much better precision of 0.70 which is reasonable. The bad cases typically involved scoring constructive criticism as overly negative, but if you read the critics itself it is more factual than emotionally charged. An example is comment 11 which reads:

"Not sure I agree with the analysis in the article, I don't think there has been an oversaturation problem before. I think it's about the product being really bad with a lot of bad teams, coaching, and QB play. All of that is making issues like officiating and ridiculous commercial breaks feel worse than before. My theory is that before the NFL product was good enough to paper over those faults, and now this year with a lot of bad games in primetime, those cracks are coming through."

This comment was given an overly negative VADER score of -0.9672 despite just being a non-emotionally charged critic of an article and counter theory to what the article proposes.

7 Discussion and Future Work

This paper has shown that sports fans of varying loyalty levels do have differences in terms of the linguistic characteristics of their online discussion board comments (such as those on the `/r/nfl` subreddit). Fans who were more loyal used more possessive language than less loyal fans whereas less loyal fans tended to be a little more positive in their language sentiment overall. The differences in possessive language and linguistic sentiment between loyalty groups was larger within fans of lower ranked teams. Comments made right after a loss also exaggerated the difference in sentiment between different loyalty groups. Lastly, sentiment differences between loyalty groups were larger under contexts where ones own team or other teams were directly mentioned and smaller under contexts where no teams are directly mentioned.

Although these results don't confirm anything about the cause of such linguistic differences between high/low loyalty groups, they do show that ones loyalty "comes out" in the linguistic characteristics of one's online posts. Future work should involve trying to isolate the characteristics of posts at a more micro level (e.g. what are the lower loyalty fans more positive about?). Also, during validation it was found that VADER did not have great precision for positive comments, therefore future work should also focus on creating an alternative sentiment analysis tool that can handle online sports discussion text (e.g. be able to tag neutral sports commentary as neutral instead of positive).

	body	Proper Pos Usage
0	It’s funny because I’m like you, commercials d...	1
1	that used to be our kicker....:’)	1
2	And made the total under feel a lot safer too :D	1
3	Giants had close to an 11 minute drive(if not ...	1
4	Hey Dilfer won it all.	1
5	Seriously, this dude clearly knows where to fi...	1
6	Lately we’ve been experimenting with this whol...	1
7	Patriots somehow pull off a win.	1
8	For the love of god if the Browns want to stop...	0
9	I’m well aware of this. My point is that it’s ...	0
10	I knew what you meant, just giving you some fr...	1
11	Lol, sticking to your first impression I see.\...	0
12	This one always makes me laugh https://www.you...	0
13	Nah, outlier seasons happen, and people have s...	1
14	While what he said is false (not this year), t...	0
15	You are really proud of a meaningless win from...	0
16	I’d swap Lions and Giants until they play. I b...	0
17	If he ever writes one I hope to god that he is...	0
18	Please god Jimmy, go anywhere, but PLEASE GOD ...	0
19	Sometimes it feels like the Falcon’s name got ...	1
20	As lifelong UGA fan, fuck your bandwagoning in...	0
21	I LOVE TYREKE HILL	1
22	Overseen by Supreme Court Justice Roger Goodell.	0
23	Hard to think it can be a Super Bowl year thou...	1
24	Welcome to yesterday, where I realized that an...	0
25	You just honestly made hyped about that matchup	1
26	lmao	0
27	Surprised you didn’t. BB seems to love those u...	0
28	###[Kelvin Benjamin pulls in a reception for a...	0
29	lol, what was Sherman giving Ryan the stink-ey...	0
30	Lmao	0
31	But that was a legal move by the WR (even thou...	0
32	Yep. Justin Smith is the most underrated defen...	1
33	Luck getting his monies worth in Indy	1
34	>If the beard stay, he gets an extension. I...	1
35	Except they’ve looked that good pretty much al...	1
36	Actually, the average ypg from ’83 (when he en...	1
37	Well Duh the DNC wanted someone from their own...	0
38	so basically you always get a possession to at...	0
39	lol kk	0

Mean: 19/40 = 0.475

Table 5: 40 random samples from s >75th percentile comments manually evaluated to see if the comments were indeed positive in sentiment. Full comments can be found in the Jupyter notebook under the Verification section.

	body	Proper Neg Usage
0	> With his head screwed onwas the ...	1
1	[I kinda thought that too until Mr. Plinkett s...	1
2	That has to be one of the worst signal-to-nois...	1
3	HEY GIANTS! GO SUCK AN EGG	1
4	Don't mean to hijack your comment, but serious...	1
5	Is it just me, or do the NBC streams not go fu...	1
6	Somebody can bring a car on the field and obli...	1
7	No thats just targeting gronk period this seas...	0
8	I bet the raiders lose a home game for the 3r...	1
9	do you mind changing your flair and retyping t...	1
10	That isn't true. The Super Bowl winner isn't ...	1
11	Not sure I agree with the analysis in the arti...	0
12	People on this sub forget about Gonzo all the ...	0
13	Oh for fucks sake	1
14	Just for you, u/Mikiflyr , I will continue to ...	0
15	Just repeat the last meeting, let Julio kick ...	1
16	not liking that reflection in the mirror stari...	1
17	Get your hat off the cameraman!! Wtf	1
18	Oldest trick in the book. Can't believe Belich...	1
19	For me, and apparently the rest of the thread,...	0
20	hmmm... Never seen that before :(	0
21	Ivan Poopybutthole, currently a sophomore in G...	0
22	Very rarely is a defense going to stop an offe...	1
23	Ebron is ballin'. He doesn't produce stats bec...	1
24	It's not roughing if you block the punt	1
25	> Problem is if he didn't try to ground it ...	1
26	Doubtful	0
27	Wait, I have to hate you guys the most? No one...	1
28	I was asking people for beetlejuice and they t...	0
29	two wrongs don't make a right fam. People hate...	1
30	I'm dreading when he actually retires.	0
31	That's what's stupid about mvp voting. Mvp doe...	1
32	Damn it sheard	0
33	That's just fucking horrendous, but how can th...	1
34	Some asshole hacked a bunch of accounts and is...	1
35	You're stuck with him.	1
36	Patriots eat ass	1
37	That's an old ad I haven't seen in a while, bu...	1
38	Fuck this I'm out	1
39	Oh I'm definitely being a bit facetious. I als...	0
		Mean: 28/40 = 0.70

Table 6: 40 random samples from $s <$25th percentile comments manually evaluated to see if the comments were indeed negative in sentiment. Full comments can be found in the Jupyter notebook under the Verification section.

References

Robert B Cialdini, Richard J Borden, Avril Thorne, Marcus Randall Walker, Stephen Freeman, and Lloyd Reynolds Sloan. 1976. Basking in reflected glory: Three (football) field studies. *Journal of personality and social psychology*, 34(3):366.

CJ Hutto Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*.

Stuck_In_the_Matrix. Every publicly available reddit comment for research. https://www.reddit.com/r/datasets/comments/3bxlg7/i_have_every_publicly_available_reddit_comment. Accessed: 2018-03-29.

Alan Tapp. 2004. The loyalty of football fans—we'll support you evermore? *Journal of Database Marketing & Customer Strategy Management*, 11(3):203–215.

Wikipedia. List of north american football nicknames. https://en.wikipedia.org/wiki/List_of_North_American_football_nicknames. Accessed: 2018-04-18.

Yang Yu and Xiao Wang. 2015. World cup 2014 in the twitter world: A big data analysis of sentiments in us sports fans' tweets. *Computers in Human Behavior*, 48:392–400.