Computers in Human Behavior 48 (2015) 392–400

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets

Yang Yu, Xiao Wang *

Rochester Institute of Technology, USA

ARTICLE INFO

Article history:

Keywords:
Sports games
FIFA World Cup
Audience emotional reactions
Twitter
Big data
Sentiment analysis

ABSTRACT

The present project collected real-time tweets from U.S. soccer fans during five 2014 FIFA World Cup games (three games between the U.S. team and another opponent and two games between other teams) using Twitter search API. We used sentiment analysis to examine U.S. soccer fans' emotional responses in their tweets, particularly, the emotional changes after goals (either own or the opponent's). We found that during the matches that the U.S. team played, fear and anger were the most common negative emotions and in general, increased when the opponent team scored and decreased when the U.S. team scored. Anticipation and joy were also generally consistent with the goal results and the associated circumstances during the games. Furthermore, we found that during the matches between other teams, U.S. tweets showed more joy and anticipation than negative emotions (e.g., anger and fear) and that the patterns in response to goal or loss were unclear. This project revealed that sports fans use Twitter for emotional purposes and that the big data approach to analyze sports fans' sentiment showed results generally consistent with the predictions of the disposition theory when the fanship was clear and showed good predictive validity.

1. Introduction

On July 8, 2014, during the FIFA World Cup semi-final game at the Estadio Mineirao in Belo Horizonte, the host Brazil played the Germans. The German brilliance transformed into multiple goals and scored seven times and won the game by a wide margin of 7-1. The faces of the Brazilian fans in the stadium were captured by the camera and broadcast worldwide. Their faces turned from anticipation early in the game to surprise to sadness and to desperation after the many German goals. In contrast to the Germans' joy and celebration, the sobbing, mourning, and desperation of the many Brazilian fans, young and old, men and women, are probably still vivid among the many who watched the game.

Admittedly, sports fans' emotions change during games. Their faces and their body movements probably are natural ways to express their emotions. With the wide adoption of mobile devices and the easy access to the Internet, can sports fans' emotions manifest in their writing on social media and through their interaction with others through the Internet and mobile devices? Based on Raney's (2006) summary of sports fans' motivations to consume mediated games, recent research in social media and electronic communication indicates that sports fans are motivated to use social media for a variety of purposes, including emotional release (e.g., Wang, 2013, 2015). Such research, although informative, is often based on a survey method and provides only correlational data related to motivations and intentions. Although significant, the indirect relationship between using the social media for emotion release and intention was around .10 in Wang (2013), indicating that motivations may not translate into intentions or behavior. Such research has yet to be confirmed by an analysis of sports fans' behaviors, manifested by their tweets or messages on social media.

To that end, the present analysis examines sports fans' emotions and behaviors on the social media through a big data approach to examine viewers' responses to sports programming.

At the methodological level, we propose that instead of using a lab experiment, we can measure viewers' emotions and responses through social media or using social media messages as media users' responses to television shows or mediated sports programs. If results of our analysis are in line with our expectation of how fans would react based in disposition theory, it indicates that the social media are used for game-related emotions and that the big data analysis, as a fairly new text analysis method, is valid based on criterion validity. At the theoretical level, we examine whether the disposition theory of sports spectatorship can be used to examine cyber behaviors; that is, whether the emotional reactions to the games, manifested in tweets, change as a function of the game.

* Corresponding author at: School of Communication, Rochester Institute of Technology, 92 Lomb Memorial Drive, Rochester, NY, USA.
E-mail address: hsiaowang@gmail.com (X. Wang).

http://dx.doi.org/10.1016/j.chb.2015.01.075
0747-5632/

1.1. Sports enjoyment and emotional release

The present study focuses on the tweets with a location stamp from the United States, whereas the games took place in Brazil. Essentially, this means that the authors of the tweets viewed mediated games or gained information through various mediated sources instead of attending games in a stadium. Mediated sports are one form of entertainment and a source of enjoyment for sports fans. Because many people, including sports spectators, are probably chronically understimulated, they seek content to "psych themselves up" (Gantz, 1981). Raney (2006) called this "eustress" and stated that sports games provide a way for sports fans to experience emotional arousal or release emotions. Emotional responses can be both positive and negative.

In general, disposition theory states that sports enjoyment is a function of the content and one's disposition (Bryant, Comilsky, & Zillmann, 1981; Zillmann, Bryant, & Sapolsky, 1989). This theory defines the viewers' or participants' feelings and attachments toward the team as affective disposition. In sports, one's disposition toward a team is called fanship. Affective disposition determines one's reactions toward and enjoyment of the sports events. The affective disposition toward a team ranges from intense liking to intense disliking. More specifically, the theory predicts that participants' enjoyment of the event or games increases when their team or liked characters are linked to positive outcomes and their disliked teams experience negative outcomes, and their enjoyment of the events or games decreases when their team or liked characters experience negative events. If viewers are indifferent to a sports team, they are probably not excited or angered by that team's win or loss.

More specifically, Raney's (2006) stated that viewers view mediated sports as a means to be stimulated and to release emotions. First, sports games, as an entertainment source, may bring the spectators joy and happiness; the enjoyment or happiness was the highest when one's liked team won and when one's disliked team was defeated. Further, Zillmann et al. (1989) reviewed several studies and found that respondents' enjoyment of the games changed over the course of game and participants rated the game more enjoyment when their team scored and rated the game less enjoyable when the opponent team scored. The latter may lead to anger and sadness among the spectators. Second, as a source to release eustress, the exciting, suspenseful, and possibly violent nature can get fans aroused, excited, and be thrilled by the victory (e.g., Bernhardt, Dabbs, Fielden, & Lutter, 1998; see also Raney, 2006 for a review). That is, sports fans' emotional reactions during the game will vary as a function of the content and their own disposition.

Recent research found that the use of social media can be used by emotional reasons as well. However, the research in this area is still limited. Our extensive literature review showed that using a survey method, Wang (2013) analyzed the motivations why sports fans used social media while watching mediated games and found that motivations related to using social media to release emotions had a significant, but weak or moderate relationship with sports fans' game enjoyment and their intentions to use social media during mediated games. As a cross-sectional study, Wang's study did not measure sports fan's actual behavior (i.e., authoring tweets) and as such did not provide evidence whether the sports fans indeed used the social media for emotional release purposes. To build on previous research, it is important to provide evidence for one's actual use of the social media.

1.2. Sentiment analysis of tweets

One way to examine sports fans' emotional behaviors on social media is to use the traditional content analysis method which relies on human coders to identify the emotions that manifest in the tweets. In the age of a vast number of social media messages and posts, the traditional method seems to be limited because it usually deals with a small number of sampled messages. In the present research, we used a "big data" approach to analyze the sentiments of the sports fans.

Sentiment analysis analyzes people's emotions, attitudes, or opinions toward various products or issues (Liu, 2012). Sentiment analysis and related research has increased considerably in the past decade. Existing sentiment analysis approaches are either based on linguistic resources or on machine learning. Sentiment analysis based on linguistic resources is centered on predetermined lists of positive and negative words and is more commonly used than machine learning (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011). By counting how many times a word appears, this approach recognizes words with positive polarity (expressing a favorable sentiment toward an object), negative polarity (expressing an unfavorable sentiment toward an object), and no polarity (neutral). It also detects words with specific emotions or moods such as joy, sadness, and anger. As such, sentiment analysis analyzes the words directly and avoids the traditional, costly, and time-consuming content analysis. Sentiment analysis has been applied to a variety of field such as business, education and politics (Ceron, Curini, Iacus, & Porro, 2014; Pang & Lee, 2008; Tumasjan, Sprenger, Sandner, & Welpe, 2010).

Twitter is a widely used social media platform and is a popular second screen. One distinctive feature of Twitter is that users can update contents instantaneously and frequently. Ji and Raney (2014) and Wang (2013) stated that users use tweets to share their real-time reactions and emotions. Thus, tweets can reflect users' thoughts in a real-time fashion and are a source of users' real-time activities and social media use. Tweets can be used to explain, detect, or predict various phenomena (Kalampokis, 2013). A growing number of scholars have used Twitter to examine social media sentiments (Liu, 2012; Pang & Lee, 2008; Yu, Duan, & Cao, 2013) or users' responses. In a recent study, Ji and Raney (2014) examined the role that morality played during the consumption of TV entertainment. Ji and Raney examined the tweets that viewers tweeted during and after viewing the Season 3 finale of the British drama Downton Abbey. Ji and Raney found that viewers tweeted more tweets related to the young protagonist Matthew who died during the finale (a major event) compared to other characters and that more morality-related tweets were found for those characters who were involved in moral events than those who were not. The results showed that tweets were consistent with the contents and events in the TV drama and can be used as a way to examine viewers' reactions.

1.3. The present investigation

The present analysis adopted a "natural experiment" approach to examine how U.S. sports fans reacted during the games when the U.S. National Soccer team competed in the FIFA World Cup 2014. We then used the sentiment analysis to examine the emotions manifested in the tweets (i.e., fans' reactions). The analysis was limited to the English language tweets with a location stamp originated from the United States instead of using all tweets in English language because of the following reasons: Previous literature review (e.g., Raney, 2006) indicates that fans' emotional reactions are determined by their affective disposition (fanship) and the content of the event. We assume that tweets in English from the United States would primarily root for the U.S. team. On the other hand, we are not certain about the teams that the authors of other English tweets were associated with. It should be acknowledged that not all English tweets with a U.S.-based location stamp were tweeted by a fan of the U.S. Soccer Team.

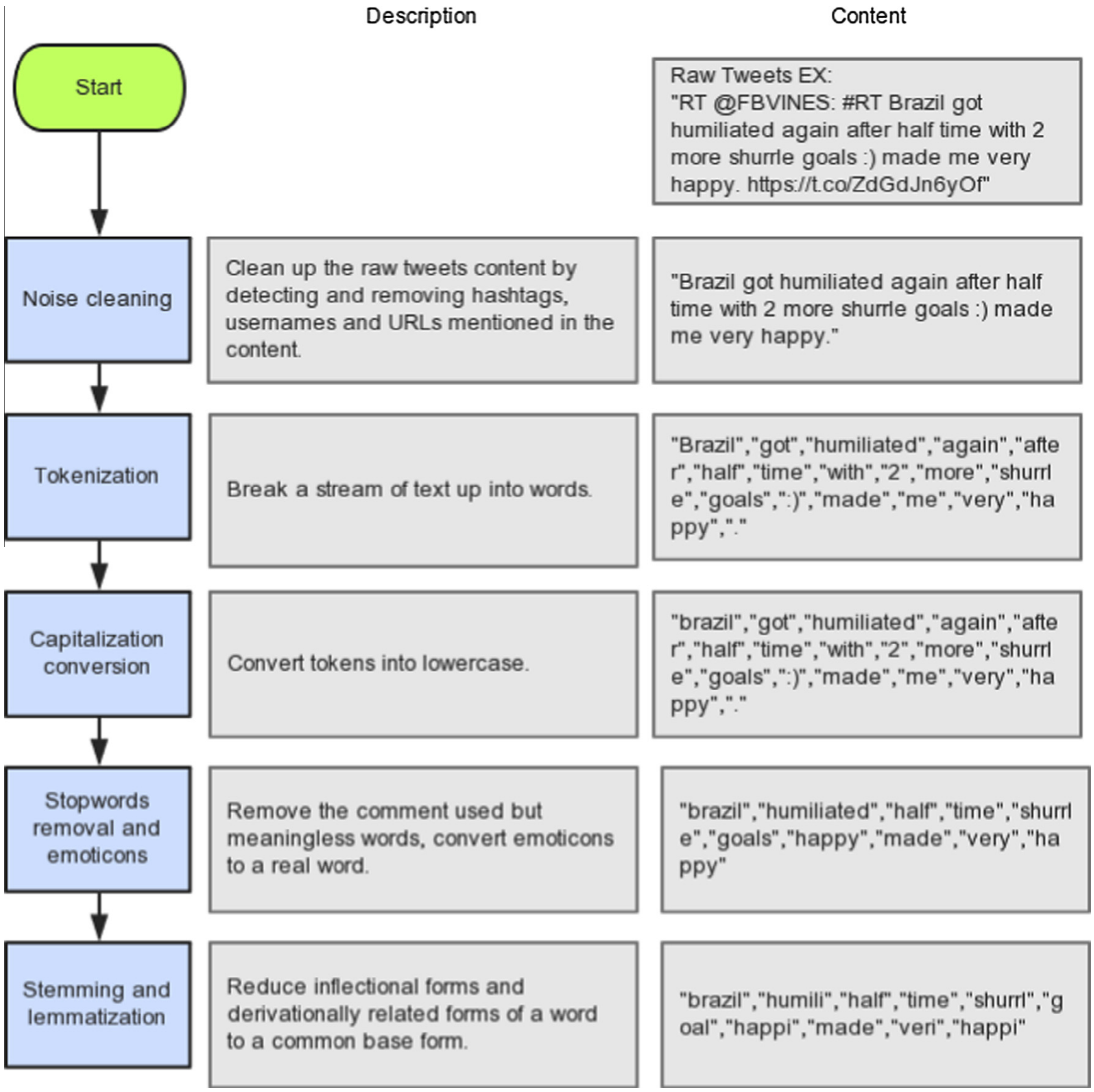

Fig. 1. Data cleaning procedure.

The tweets could be potentially tweeted by for example, Germans or British, living in the United States. However, the number of these foreign dwellers is small compared to the number of U.S. fans. That is, we acknowledge the potential bias that foreign dwellers may bring into our data, but we choose to use U.S. location stamps as a proxy for U.S. fanship.

Second, we previously reviewed that sports fans' emotional reactions and enjoyment change according to the win and loss of their own team (Bryant et al., 1981; Raney, 2006). As such, U.S. sports fans' affective disposition toward the U.S. soccer team will lead them to experience various emotions and enjoyment that are results of the win or loss of goals of the U.S. team. Fans would be excited and experience positive emotions (e.g., joy and excitement) when their own team scored and would experience negative emotions (e.g., anger, disappointment) when the opponent team scored. The tweets during and shortly after the games can be correlated with the win or loss of the U.S. team because we previously reviewed evidence that tweets can be indicative of viewers' reactions (Ji & Raney, 2006). On the other hand, the disposition theory states that if the sports spectators are indifferent, they are less likely to experience anger or joy. For example, the U.S. fans would less likely to be invested on the game between two unrelated teams, for example, a game between France and Nigeria on June 30, 2014. Thus, we hypothesized the following:

H1. Tweets with a U.S. location stamp (a) would show negative emotions (e.g., fear, anger) when the U.S. soccer team's conceded a goal and (b) would show positive emotions (e.g., joy, hope) when the U.S. soccer team scored a goal.

H2. Tweets with a U.S. location stamp will remain "indifferent" to the win or loss of other teams.

2. Method

2.1. Sample

We retrieved tweets from twitter.com via Twitter's Search API during the 2014 FIFA World Cup. Twitter Search API provides programmatic access to read and write Twitter data and provides approximately 1–2% of a random sample of all tweets. We designed a web crawler to collect and parse English tweets in real time using a list of predefined hashtags include #FIFA, #Football,

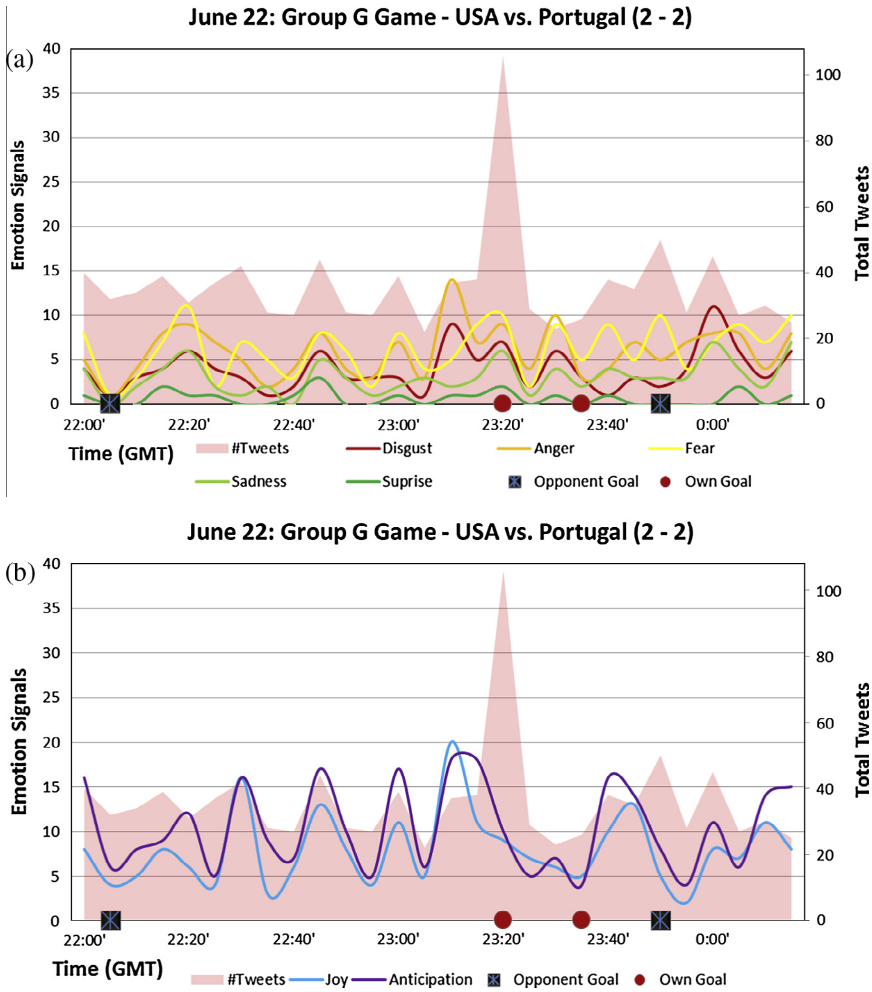

Fig. 2. (a and b) Emotional changes during the game between USA and Portugal. The final score is shown in the parentheses. Emotional signal = number of words that were coded as various emotions. Total numbers of tweets were on a different scale.

#Worldcup, or #Soccer, ignoring case considerations. For each tweet or retweet, we parsed several key properties such as user's screen name, user ID, post time (Greenwich Mean Time, GMT), retweet frequency as well as the content. In addition, we collected location stamps for the tweets if available.[1]

Our analysis was limited to the tweets posted during five games. The U.S. team played three games during the group stage and an additional game during the round of 16. Because Tweets' location information was not collected for the first game, we limited our analysis to the last three games that the U.S. team played (June 22, June 26, and July 1). It should be acknowledged that on June 26, a game between Portugal and Ghana was played in the same time period when the U.S. team played Germany. The results of the Portugal and Ghana game might determine when the U.S. team would play in the round of 16. In addition, we collected the tweets during two non-U.S. games for comparison purposes: the game between France and Nigeria on June 30, 2014 and the quarter-final game between Brazil and Colombia on July 4, 2014. These two comparison games were randomly selected from games in the rounds of 16 and 8—we decided not to select additional games from the group stage because most of times, two games were played during the same time slot and not from the final elimination stage because these games usually feature football powerhouses and superstars (e.g., Brazil, Germany, Argentina and big names such as Messi or van Persie) In total, this project analyzed 1007, 1295, 2135 tweets for the three U.S. games (out of a total of 26,881, 26,014, and 49,576, respectively), and 461 and 468 tweets for the France–Nigeria and Brazil and Colombia (out of a total of 21,901 and 25,494, respectively). The total numbers of tweets listed in the parentheses included both tweets without location information and tweets with location stamp from other countries.

2.2. Analysis and coding procedure

The present project analyzed the emotions at the word level; that is, words with various discrete emotions were tabulated and counted. The more detailed procedure is as follows (Fig. 1): We chose Python and the Natural Language Toolkit (NLTK), a platform with resources and programming libraries suitable for linguistic processing (Bird, 2006). Given that Twitter messages contained useless or nonusable information, we designed a preprocessing work flow to clean and convert content of messages, include

[1] In the 31 days period of World Cup (June 12, 2014 to July 13, 2014), we collected a total of 13.58 million tweets, of which 7.05 million were original tweets and 6.53 million were retweets. In the 13,582,469 tweets we collected, 221,359 tweets (1.63%) had location stamps from 198 countries, of which 111,955 were posted from the United States, 34,339 were posted from the Great Britain and 23,946 were from Brazil.

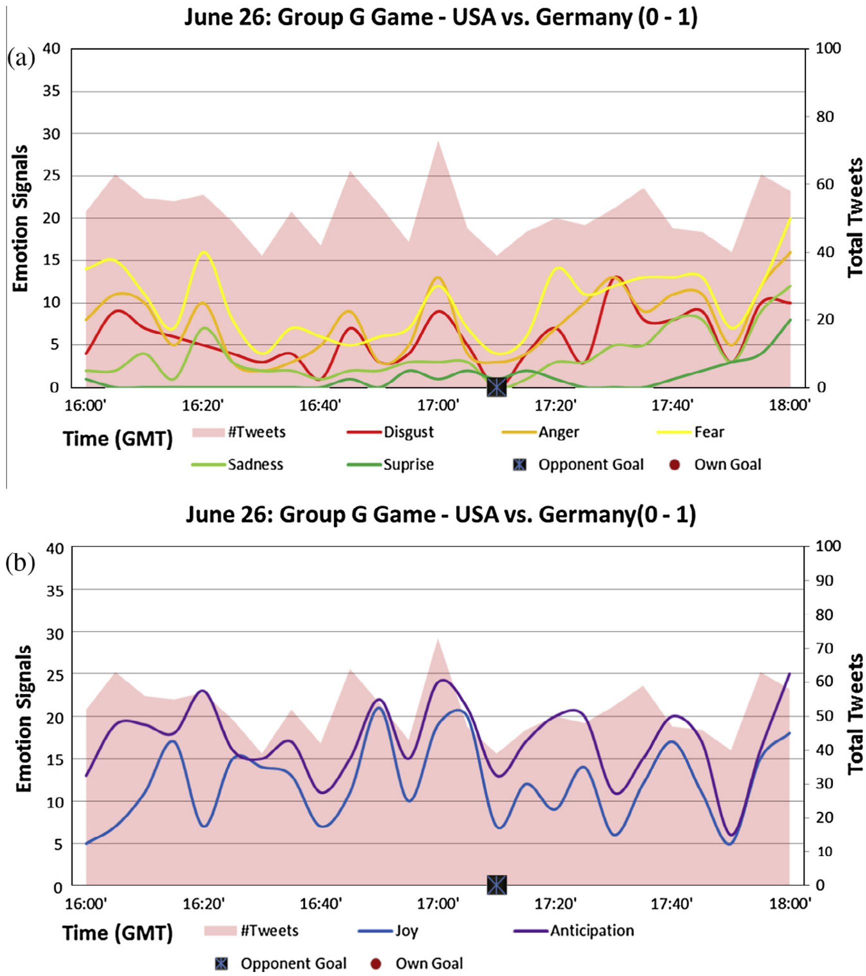

Fig. 3. (a and b) Emotional changes during the game between USA and Germany. The final score is shown in the parentheses. Emotional signal = number of words that were coded as various emotions. Total numbers of tweets were listed on a different scale.

modules such as user name/URLs/hashtags detection and removal, tokenization,[2] uppercase conversion, stemming & lemmatization,[3] stopword removal, and emoticons conversion (Liu, 2012).

The present project used two ways to analyze the emotions and sentiments in the words. First, NRC Word-Emotion Association lexicon (Mohammad & Turney, 2013) was used to measure the specific emotions including anger, fear, joy, sadness, disgust, surprise, trust, and anticipation, which are considered as eight basic and prototypical emotions (Plutchik, 1980). As a lexicon-based approach, we used the lexicon words' frequency in a tweet to measure of the strength of a specific dimension. For example, if two words in a tweet were coded as joy and one word was coded as anticipation, joy was counted twice and anticipation was counted once. An application written with R was created to extract features related to the NRC lexicon from a given tweet and then calculated score for each emotion by counting the words that matched the eight categories. Second, tweets were characterized by various emoticons that can express mood. For example, either ":)" or ":-)" expresses positive mood, and ":(" expresses negative mood. We also considered the emoticons in our extended list by mapping them to either "☺"or "☹". In this study, we extend the emotion list by considering ☹ to sadness and ☺ to joy. Examples of how emotions and sentiments were coded are presented in Table 1.

The data within each 5-min block were aggregated and were then plotted in Fig. 2.

3. Results

The first hypothesis of the present research stated that the U.S. sports fans' would experience more negative emotions after the U.S. team conceded a goal and that the fans would experience positive emotions after the U.S. team scored. We examined the patterns of the tweets to see whether they were consistent with our expectations. It should be acknowledged that during soccer games, there are multiple attacks and incidents of being attacked by the opponents, corners, free kicks, penalty kicks, and fouls that may also influence one's emotional responses. Fans may even start celebrating or experiencing joy when a foul is committed or two three minutes before a penalty kick is scored. The first two games in the group stage should also be understood relative the results of other games, whereas the game in the round of 16 was a game of elimination.

[2] Tokenization, or a tokenizer divides text into a sequence of tokens, which roughly correspond to "words." However, a sentence tokenizer uses an unsupervised algorithm to build a model for abbreviation words, collocations, and words that start sentences; and then uses that model to find sentence boundaries.

[3] The goal of both stemming and lemmatization is to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form. Stemming usually refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal correctly most of the time, and often includes the removal of derivational affixes. Lemmatization usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma.

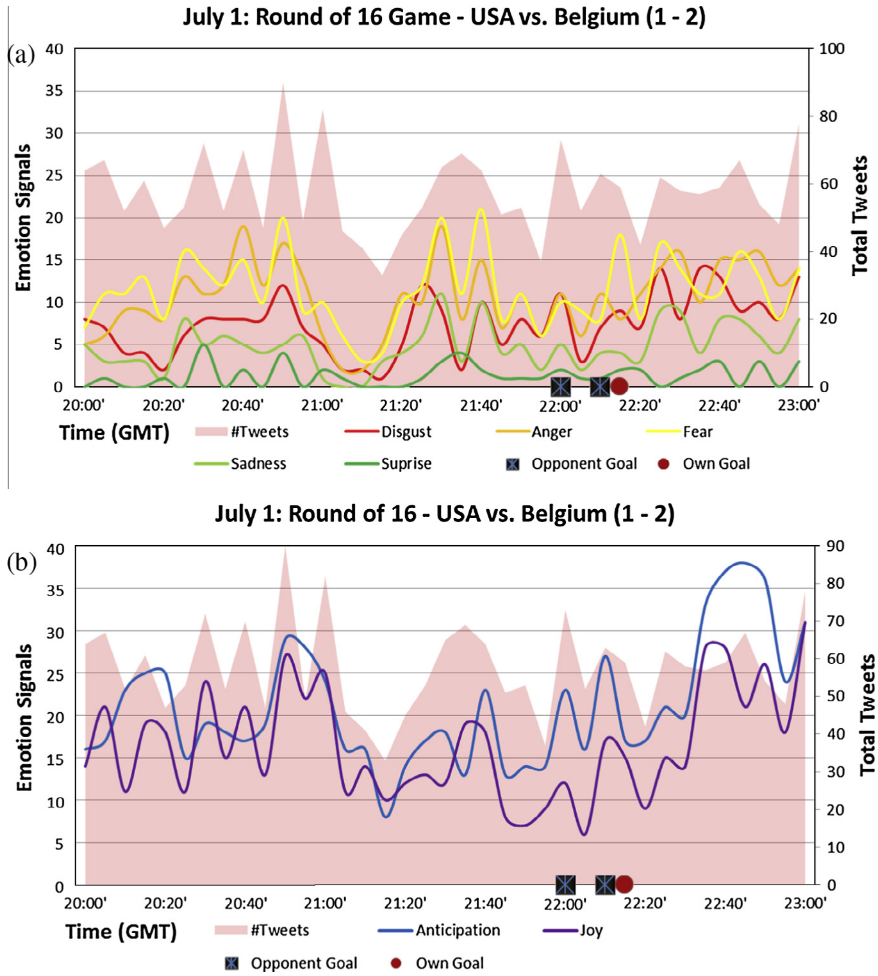

Fig. 4. (a and b) Emotional changes during the game between USA and Belgium. The final score is shown in the parentheses. Emotional signal = number of words that were coded as various emotions. Total numbers of tweets were listed on a different scale.

Table 1
Lexicon examples for emotional coding.

Emotions	Lexicon sample	Tweets sample
Anticipation	Believe, hope, expect	RT @BrosConfessions: I believe that we will win, go #USA
Fear	Afraid, worry, nervous	I am afraid US will never win the WC until they realize it's football not soccer
Anger	Hell, wft, shit, roar	Bored as shit…lemme fill my TL with soccer-unrelated tweets so I can read them and understand wtf I'm talking about
Surprise	Amazing, cool, what, surprise	It's amazing how many soccer fans come out during the World Cup
Sadness	Sadness, sad, sorrow, depress	Its a sadness, a player of such brilliance in creating chances for himself + others is such a toxic liability at times
Joy	Wonderful, happy, bliss	Allah o Akbar, Im happy for Algeria's progression to next round
Disgust	Slop, suck, sick, err	eeeeerrrrr your skills(or lack of skills) make me sick

Table 2
Total of number of words with emotions and total number of tweets for five games.

	Disgust	Anger	Fear	Sadness	Surprise	Joy	Anticipation	#Tweets
USA vs. Portugal	112	168	176	86	19	223	293	1007
	0.11	0.17	0.17	0.09	0.02	0.22	0.29	
USA vs. Germany	146	192	250	94	29	303	428	1295
	0.09	0.13	0.14	0.07	0.01	0.17	0.23	
USA vs. Belgium	275	395	425	177	49	614	781	2135
	0.13	0.19	0.20	0.08	0.02	0.29	0.37	
France vs. Nigeria	43	65	68	29	4	189	222	461
	0.09	0.14	0.15	0.06	0.01	0.41	0.48	
Brazil vs. Colombia	90	93	78	47	10	98	113	468
	0.19	0.20	0.17	0.10	0.02	0.21	0.24	

Note. In each cell, the top number is the total number of words with an emotional signal. The bottom number is the total number of words with a given emotional signal divided by the total number of tweets with a U.S. location stamp.

During the game on June 22, 2014, the U.S. soccer team played Portugal. The Portuguese team scored at 5′ and 90 + 5′, whereas the U.S. team scored at 64′ and 81′. According to the disposition theory, and if tweets reflect fans' emotions, we should observe a surge of fear and anger and other negative emotions at 5′ and 90 + 5′. Fig. 2a showed that fear and anger were the more frequently detected negative emotions compared to sadness and surprise. Furthermore, after five minutes into the game when the Portugal scored its first goal, the number of tweets with anger and fear went up and reached their peak in the following twenty minutes. These emotions fluctuated throughout the first half and the half time. Then anger and fear reached their second peaks in the beginning the second half, respectively. And fear and anger declined after the first U.S. goal and again declined after the second U.S. goal. These negative emotions went up after the final Portuguese goal in the extra time.

Fig. 2b showed the emotions of anticipation and joy during the game. Anticipation was constantly up after the U.S. conceded the first goal (i.e., expecting the U.S. team to level the game) and waned off around the time the U.S. team scored its first goal and again went up around the time that the U.S. team scored its second goal (i.e., probably expecting the U.S. team to maintain the lead). Overall, the check points based on the goal time seemed to show a pattern that was roughly consistent with the expectations and sport psychology.

Fig. 3a and b showed the negative and positive emotions in the tweets, respectively. Fear increased at the 55′ when Germany scored and again increased to the highest level at the end of the game and shortly after the game. Fig. 2b showed that anticipation and happiness dipped when Germany scored the only game-winning goal at 55′ and was the lowest shortly after the game. However, anticipation and happiness surged about 20 or 30 min later (2 h 10 min since the start of the game)—it was probably the time that most U.S. fans realized that the United Stated advanced to the second round despite a close loss to Germany.

Fig. 4a and b showed the emotional patterns during the third game. This game was during the round of 16 (i.e., elimination stage). Tweets showed a mix of anger and fear throughout the game. Neither team scored during the regular time. At the beginning of the extra time (after approximately 95 min and half time), anticipation reached its peak. However, anticipation and joy went down, fear and anger surged up shortly afterward when the Belgium scored its first goal at 92′ and again went up when the Belgium scored the second time at 105′. At 105′, the U.S. fans showed low anticipation and joy and high anger and fear. At the end of the game, joy, however, went up again, indicating that a close game, despite a loss, can still be a source of enjoyment.

We selected a comparable round-of-16 game (France vs. Nigeria) and also a quarter-final game (Brazil vs. Colombia) to examine U.S. Twitter users' emotional reactions. For these games, the overall tweet counts remained lower. The indicators based on the number of total emotion words and total tweets showed that negative emotions shown during these two games and during the three U.S. games remained similar (see Table 2). The patterns of emotional reactions were difficult to predict potentially because these sports fans could be motivated by other factors (e.g., love for

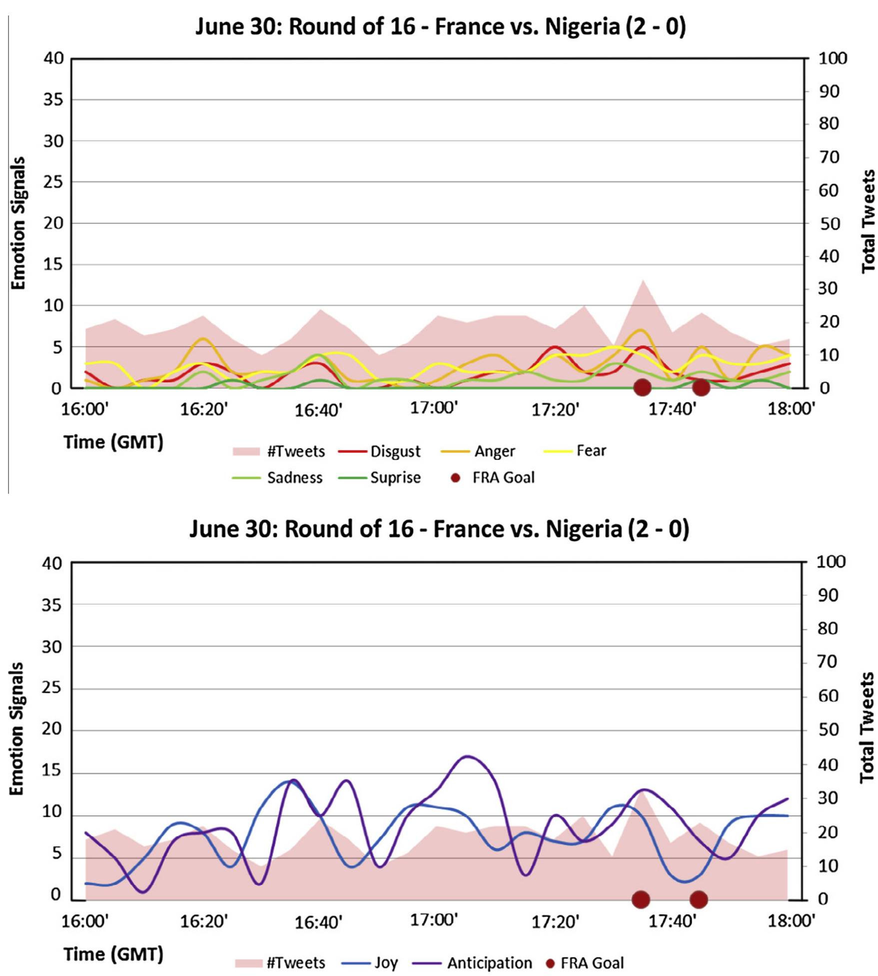

Fig. 5. (a and b) Emotional changes during the game between France and Nigeria. The final score is shown in the parentheses. Emotional signal = number of words that were coded as various emotions. Total numbers of tweets were listed on a different scale.

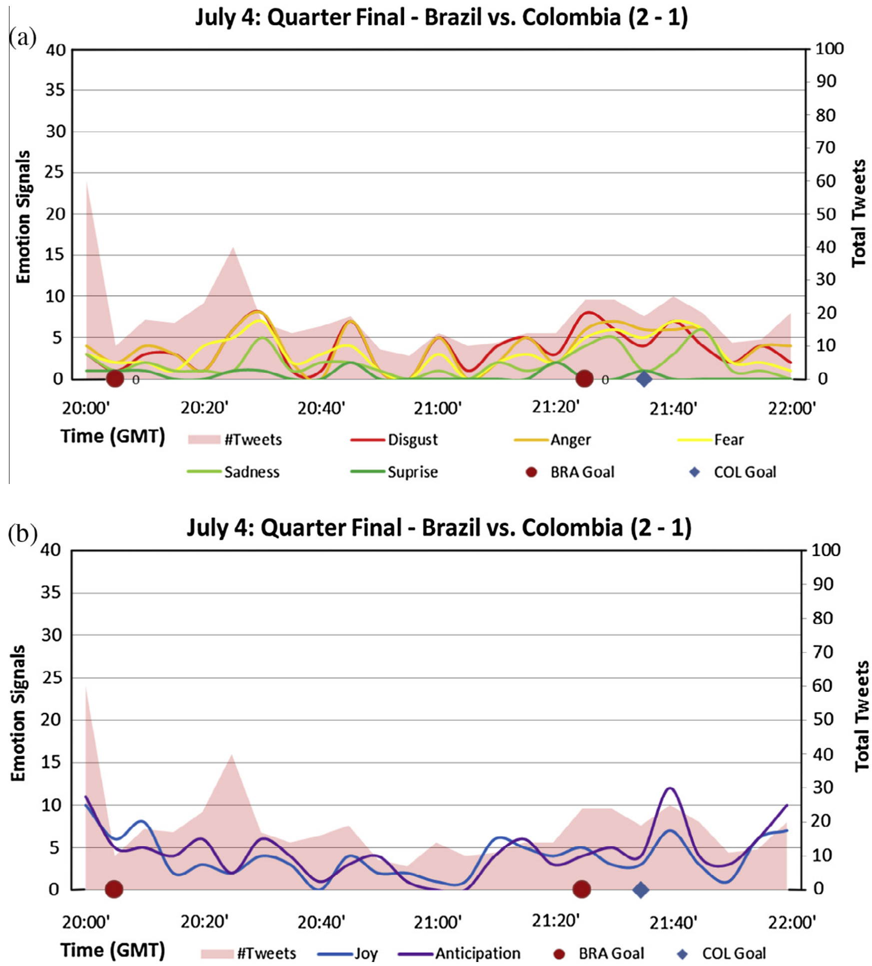

Fig. 6. (a and b) Emotional changes during the game between Brazil and Colombia. The final score is shown in the parentheses. Emotional signal = number of words that were coded as various emotions. Total numbers of tweets were listed on a different scale.

soccer) or could be a fan of any of the four participating teams in these two games. For example, during the round-of-16 game between France and Nigeria, viewers' anticipation and joy were the dominant emotions and were higher than the rest four games.

4. Discussion

The present paper reports on U.S. soccer fans' emotions in tweets by using a "natural experiment" and a big data approach to analyze the real-time sentiment in the sports fans' tweets. Below, we highlight and discuss three major findings and issues: the patterns of tweets and the games, the use of real events or "natural experiments," and the use of the big data approach to analyze the vast number of tweets.

First, we found the emotional patterns of the tweets largely consistent with our expectations and the predictions of the disposition theory (Zillmann et al., 1989). Regarding U.S. fans' emotional reactions to their own teams' loss or draw, the number of the U.S. sports fans' negative emotions, mainly anger and fear, increased when the opponents scored and decreased when the U.S. team scored and was consistent with the disposition theory of sports spectatorship. When putting in the context, anticipation was high when the U.S. needed some "push" or when the U.S. team showed some positive signs (e.g., game on June 22, 2014).

We also found that for a non-U.S. related games, the amount of negative emotions was much lower and remained constant throughout the games. On the other hand, tweets showed more positive emotions. Taken together, fanship heightens one's involvement with the results of the games and introduces fear, anger, and other negative motions, as well as providing enjoyment for the fans. On the other hand, a non-related game (e.g., France–Nigeria) did not seem to affect sports spectators' sense of worry (i.e., indifferent to gain or loss) and is mainly a source of anticipation and enjoyment. That is, our results seem to indicate that although one may not be a fan of the competing teams, one may still enjoy the games and results as a fan of a given sport. To state at a deeper level, fanship of a team enhances one's worries and involvement (e.g., anger and fear) and enjoyment (e.g., happiness) related to win or loss of the team, whereas being a fan of a given sport is mainly tied to enjoyment (e.g., happiness) and emotional release (anticipation).

It should also be noted that our research focused on the English tweets from the United States. If social media posts for the semi-final game between Germany and Brazil were analyzed, we might be able to find that social media posts from Germany may exhibit little fear or anger because the German team established a lead early in the match and won by a large margin. The competitiveness of the games may determine sports' fans' emotions.

Second, the present research used a "natural experiment" approach and used fans' real-time tweets as the dependent variable. One benefit of the natural experiment is to provide good ecological validity of the results: how the goals and losses are related to sports fans' emotional reactions in tweets. The emotional reactions in the tweets could be based on the goals and losses and could also be based on the interaction with other fans in their homes, in a bar, or through social media. That is, our approach provides real-life interactions and reactions from the sports fans, and is an alternative to the controlled experiments. The controlled experiments often lack good ecological validity because researchers artificially manipulate various variables in the experiments and the participants may perform because of the Hawthorne effect or demand characteristics. However, unlike controlled experiments, we cannot control or manipulate the win or loss of a game or how the games are played. In soccer games, the number of goals (or points) is low. Many plays such as attacks, counter-attacks, corner kicks, shots on the goal, and fouls even half-time commentary, may result in fans' emotional reactions. It is also possible that during the games, broadcasters encourage fans to tweet, which may result in a higher number of tweets, with both positive and negative emotions. Such emotional reactions build up and go away. Not surprisingly, the graphs show that many peaks and valleys of the several emotions during the games were not in response to goals. They were probably the responses to other acts and plays during the games. But we did find that emotions manifested in tweets were largely consistent with the goal times. Overall, the analysis of real-time tweets or other social media messages through sentiment analysis can be an alternative to controlled lab experiments in media psychology.

Third, our study showed that the use of a big data analysis of the sentiments in tweets is consistent with our expectation and can be reasonably explained. The big data analysis allowed us to analyze a relatively large number of tweets in this case and allowed us to examine the sports fans' natural emotional reactions in the real life (i.e., unlike the emotional reactions due to response to experimental manipulation, which may lead to demand characteristic and Hawthorne effects). However, our analysis of the sentiments has its own limitations as well. First, we analyzed the emotions through analyzing each word. We did not consider the semantic meaning of the whole tweets. We note that sentiment lexicon cannot cover the complete domain knowledge and cannot extract the exact meaning of each word in a specific scenario. For example, in "I am searching a good game to watch," "good" here does not express either a joy or trust emotion on any particular game. Second, we examined the tweets as a way to manifest emotions. Many sports fans may not rely on tweets to manifest emotions. At the moment, we do not know the percentage of sports fans that engage in such behaviors and the characteristics of these sports fans. That is, the big data analysis may only represent some sports fans, but not all. Third, we used the location stamp as the proxy of fanship, which would bring bias/noise in the results. In addition, this approach causes information loss because only 1.6% tweets provided location information. Fourth, regarding U.S. Twitter users' emotional signals in tweets, the pattern was hard to interpret. We found that in general, the numeric indicators based the number of words with emotional signals and total tweets for a given game were comparable. However, the patterns based on Figs. 5 and 6 were less clear. The reason could be that there were fewer tweets compared to the other three games, and possibly only those who were strong fans of soccer or a particular team watched the game. Fifth, our research focused on the World Cup 2014 and soccer-related tweets from soccer fans and did not analyze other types of tweets or social media messages (e.g., basketball or sentiments toward a political candidate). Future research may want to extend our projects to analyze other social messages for other events, that is, to explore the underlying emotion expression pattern in various media sources and among different segments of publics. Lastly, we used results of sentiment analysis as the dependent variable and examined how sentiments in tweets varied based on media events. Sentiments revealed in tweets can also be used to predict firm equity (Yu et al., 2013) or to predict whether a policial candidate is elected.

In conclusion, our study took a novel approach to examine the emotions manifested in the tweets during several games. We compared and contrasted U.S. sports fans' emotional reactions toward both games in which the U.S. team played and the games between two non-U.S. teams. Our analysis, based on sentiment analysis, provided support for the disposition theory of sports spectatorship such that negative emotions increased after one's team loss of a goal and decreased after one's team scored a goal. The results also showed that sports fans, although less concerned about the loss or goals of non-affiliated teams, enjoyed other games by showing anticipation and joy over the tweets. Finally, the sentiment analysis using the Mohammad and Turney (2013) lexicon showed some results as predicted by the disposition theory, which provided some predictive validity for their lexicon.

References

Bernhardt, P. C., Dabbs, J. M., Fielden, J. A., & Lutter, C. D. (1998). Testosterone changes during vicarious experiences of winning and losing among fans at sporting events. *Physiology and Behavior, 19*, 263–268.

Bird, S. (2006). NLTK: The natural language toolkit. In *Paper presented at the proceedings of the COLING/ACL on Interactive presentation sessions*.

Bryant, J., Comilsky, P. W., & Zillmann, D. (1981). The appeal of rough-and-tumble play in televised professional football. *Communication Quarterly, 29*, 256–262.

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media and Society, 16*, 340–358.

Gantz, W. (1981). An exploration of viewing motives and behaviors associated with television sports. *Journal of Broadcasting, 25*, 263–275.

Ji, Q., & Raney, A. A. (2014). Morally judging entertainment: A case study of live Tweeting during Downton Abbey. *Media Psychology*, 1–22. http://dx.doi.org/10.1080/15213269.2014.956939.

Kalampokis, E. (2013). Understanding the predictive power of social media. *Internet Research, 23*, 544–559.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*, 1–167.

Mohammad, S. M., & Turney, P. D. (2013). NRC Emotion Lexicon.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Harper & Row.

Raney, A. A. (2006). Why we watch and enjoy mediated sports. In A. A. Raney & J. Bryant (Eds.), *Handbook of sports and media* (pp. 313–329). New York: Lawrence Erlbaum Associates.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*, 267–307.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM, 10*, 178–185.

Wang, X. (2013). Applying the integrative model of behavioral prediction and attitude functions in the context of social media use while viewing mediated sports. *Computers in Human Behavior, 29*, 1538–1545.

Wang, X. (2015). Using attitude functions, self-efficacy, and norms to predict attitudes and intentions to use mobile devices to access social media during sporting event attendance. *Mobile Media and Communication, 3*, 75–90.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems, 55*, 919–926.

Zillmann, D., Bryant, J., & Sapolsky, B. (1989). Enjoyment from sports spectatorship. In J. H. Goldstein (Ed.), *Sports, games, and play: Social and psychological viewpoints* (2nd ed., pp. 241–278). Hillsdale, NJ: Lawrence Erlbaum.