# **ORIGINAL ARTICLE**

# Keeping it 100: Social Media and Self-Presentation in College Football Recruiting

Kristina Gavin Bigsby,<sup>\*</sup> Jeffrey W. Ohlmann, and Kang Zhao

# Abstract

Social media provides a platform for individuals to craft personal brands and influence their perception by others, including potential employers. Yet there remains a need for more research investigating the relationship between individuals' online identities and offline outcomes. This study focuses on the context of college football recruiting, specifically on the relationship between recruits' Twitter activities and coaches' scholarship offer decisions. Based on impression management theory, we analyze content posted by recruits and apply machine learning to identify instances of self-promotion and ingratiation in 5.5 million tweets. Using negative binomial regression, we discover that an athlete's level of engagement on Twitter is a positive and significant predictor of the number of offers. Also, both self-promotion and ingratiation are positively related to attracting new offers. Our results highlight the growing importance of social media as a recruiting tool and suggest that recruits' online self-presentation may have significant offline impacts. This research can benefit athletes and coaches by informing communication strategies during recruitment, and may also yield insight into the consequences of online impression management in other types of recruitment beyond sports.

Keywords: impression management; social media; sports analytics; recruitment; text mining

## Introduction

The proliferation of social media has given athletes the power to craft personas and influence how others perceive them in a virtual setting. However, questions remain as to whether these online identities have measurable offline impacts, as measured by sponsorships, salary, or draft success. In this work, we contribute to the sports analytics literature by investigating the relationship between college football recruits' behavior on Twitter and the number of scholarship offers they receive. Although sports media has reported on coaches' use of social media to strategically engage recruits<sup>1</sup> and to monitor players' behavior online,<sup>2</sup> there is no empirical research focusing on athletes' use of social media during recruitment.

College football is often a significant part of the public face of an institution. In 2017, more than 42 million fans attended Division I football games.<sup>3</sup> Top athletic departments generate multimillion-dollar profits, and success in athletics is linked to trends in general student body enrollment,<sup>4</sup> institutional reputation,<sup>5</sup> and donor behaviors.<sup>6</sup> On-field performance is highly dependent on recruiting. Bergman and Logan<sup>7</sup> have shown that individual recruit quality is a significant predictor of team success, with each additional five-star recruit worth an average of 0.306 wins, and the worth of a premium college football player has been estimated to be as high as \$2.3 million per season.<sup>8</sup> Thus, it is unsurprising that college football teams expend upward of \$1 million annually on recruiting activities.<sup>9</sup>

A significant portion of these resources is devoted to evaluating potential recruits and deciding whether to extend a scholarship offer. In 2014, Division I coaches recruited from a pool of more than 250,000 high school seniors, with only 2.5% advancing to play Division I football.<sup>10</sup> Recent evidence suggests that coaches increasingly turn to social media as a convenient way to screen the large number of prospects. In a survey<sup>11</sup> of 477 Division I, II, and III coaches across 19 sports, 85% reported searching for information about recruits online and 79% specifically used Twitter. In an *ESPN* 

Department of Management Sciences, University of Iowa, Iowa City, Iowa.

<sup>\*</sup>Address correspondence to: Kristina Gavin Bigsby, Department of Management Sciences, University of Iowa, Iowa City, IA 52242, E-mail: kristina-gavin@uiowa.edu

report<sup>1</sup> on social media monitoring, Oregon State football's director of player personnel, Darrick Yray, states, "Every single school does it. You have to, especially since you're investing almost \$500,000 in a player's development over a 4- or 5-year period." Social media has become an indispensible recruiting tool as it is not limited by physical location, can be accessed at any time, and presents an opportunity to gather detailed information while complying with National Collegiate Athletic Association (NCAA) rules<sup>12</sup> on the medium, timing, and frequency of communication between college coaches and high school athletes.

Mass media articles about the role of social media during the screening process have primarily focused on cases of negative behavior, including "racist, sexist, vulgar, or profane posts."2 While this concern is warranted-19% of coaches surveyed by Cornerstone Reputation<sup>11</sup> had rescinded an offer based on a recruit's online presence-anecdotal evidence suggests that social media misbehavior is decreasing over time. In a statement to Athletic Business,<sup>2</sup> Joe Dooley, then-head coach of men's basketball at Florida Gulf Coast University, suggests that athletes are increasingly aware of the scrutiny they are under both during recruitment and over the course of their college careers, "I do think, though, in the last couple of years student-athletes have gotten much more savvy and they understand that and they don't do that as much as they used to." Indeed, fewer than 4% of tweets in our data set were judged to be negative and/or offensive. Thus, we examine how recruits can leverage social media positively for networking and self-presentation. In the Cornerstone Reputation survey,<sup>11</sup> 90% of coaches reported finding something about a recruit online that gave a good impression, and 82% believed that athletes with strong, positive online presences had an advantage over other recruits. A survey<sup>13</sup> of Division I athletes noted an increase in players initiating contact during recruitment-37% of respondents had reached out to coaches via e-mail, mail, or sending a video highlight. Social media offers an easy way to engage in this type of self-promotion; athletes can instantly connect with coaches on Twitter and share information about themselves by replying, mentioning, and/or including links to news coverage or video highlights in their tweets.

Our study represents the first large-scale examination of athletes' online behavior during recruitment. We approach the recruiting activities and Twitter content posted by 2644 athletes in the college football recruiting class of 2016 with two main research objectives. First, we examine how athletes present themselves on Twitter during recruitment. We explore this issue from the perspective of impression management theory,<sup>14</sup> which describes individuals' conscious and subconscious attempts to influence the perceptions of others. Specifically, we identify instances of selfpromotion (posting positive content about the athlete's own athletic prowess, academic abilities, and recruiting achievements) and ingratiation (posting positive content about other recruits, college teams, or current college athletes) in 5.5 million athlete tweets. Second, we investigate the relationship between Twitter content and scholarship offers to determine whether certain online behaviors may lead to recruiting success. This work is the first to incorporate social media data into an explanatory model of the number of scholarship offers received by college football players. We model the number of new offers received during junior year as a function of the athlete's personal characteristics, recruiting activities, and Twitter content up to that point in time.

# **Related Work**

To position our contribution, we look to prior work on impression management in organizations, the role of social media in human resources (HR), and social media usage by college and professional athletes.

#### Impression management

First introduced in 1959, impression management theory<sup>14</sup> explains individuals' efforts to influence others' perceptions. We borrow from the organizational management and personnel psychology literatures, where impression management has been well-studied in the context of employment interviews. Although Goffman's original dramaturgical metaphor delineates "front stage" and "back stage" tactics, several taxonomies of impression management strategies have been proposed over time. In a laboratory study, Kacmar et al.<sup>15</sup> describe "self-focused" versus "other-focused" behaviors, and find that participants who engaged in self-focused impression management during interviews were rated higher and received fewer rejections than those using other-focused impression management. Tsai et al.<sup>16</sup> categorize verbal versus nonverbal strategies, finding that verbal self-promotion significantly influenced applicant ratings. Jones and Pittman<sup>17</sup> propose the most commonly used taxonomy, describing five primary tactics: self-promotion, ingratiation, exemplification, intimidation, and supplication. In a field study<sup>18</sup> of impression management during employment

interviews, interviewees used self-promotion tactics more often than ingratiation, and ingratiation did not significantly impact hiring managers' perceptions of suitability or the likelihood of receiving a job offer. We also utilize the Jones and Pittman<sup>17</sup> categories in our study, specifically identifying self-promotion and ingratiation in athletes' tweets.

#### Social media in HR management

Social media has changed the landscape of HR recruiting, simultaneously offering individuals new avenues to engage in self-branding activities and presenting companies with additional information sources to evaluate potential employees. In a 2013 survey,<sup>19</sup> 64% of companies reported using social media to inform hiring decisions. Social media has been described as a "real-time search engine"<sup>20</sup> that is especially valuable for evaluating candidates' professionalism and writing skills.<sup>21</sup> Surveys<sup>22</sup> suggest that hiring managers are attentive to applicants' photographs on social media as well, using them to assess traits of extraversion and maturity. Broughton et al.<sup>23</sup> report that 35% of hiring managers have found material on social media that caused them not to hire a job candidate, especially "provocative" photographs. In recent years, companies have tripled their investment in professional social networking sites such as LinkedIn,<sup>24</sup> although there remains a need for more work examining the link between online behavior and offline HR recruiting outcomes. Roth et al.<sup>25</sup> state in a recent review of the literature on social media and employee selection, "Organizational practice has greatly outpaced the scientific study of social media assessments." In an experimental study,<sup>26</sup> participants rated candidates with family- or professionaloriented Facebook profiles higher than candidates with alcohol-oriented profiles. In addition, participants offered an average of \$2400 more to candidates who exhibited a professional online presence. Our study extends this prior research by utilizing observational data obtained by tracking the recruiting activities, social media content, and scholarship offers of college football recruits in the class of 2016.

## Social media and athletics

Previous studies suggest that college athletes may approach social media differently than their nonathlete peers. A survey<sup>27</sup> of 202 student-athletes and 419 non-athletes finds that, while athletes do not spend significantly more time on Facebook, they do tend to have more friends. Browning and Sanderson<sup>28</sup> perform

qualitative interviews with student-athletes, finding that keeping in contact, communicating with followers, and accessing information were the primary motivations for using Twitter. The authors touch briefly on the role of social media during recruitment, with one athlete recounting criticism by fans disappointed in his college choice. Our study is the first to quantify the prevalence of different impression management strategies in college recruits' social media profiles, and represents a significant addition to the literature on self-presentation in sports.

Studies of social media usage by professional athletes have investigated the types of impression management used across different sports<sup>29,30</sup> and differences by gender.<sup>31</sup> Professional athletes' social media content has previously been linked to fan attitudes.<sup>31,32</sup> Jin and Phua<sup>33</sup> conduct two experiments investigating the effects of a celebrity's number of Twitter followers and behavior on consumer influence. The authors find that prosocial behavior (e.g., philanthropy) is positively related to perceptions of credibility and attractiveness, and that this effect is intensified by increasing numbers of followers. Quantifying the return-on-investment of athlete endorsements, sports marketing firm Navigate Research<sup>34</sup> states that fans are 164% more likely to purchase a product if an athlete they follow endorses it on social media. Chung et al.<sup>35</sup> focus on the case of golfer Tiger Woods and Nike, finding that the company's decade-long relationship with Woods was associated with an increased share of the golf ball market, from 1.59% in 2000 to 10% in 2010. While the authors estimate that Nike's decision to publicly stand by the golfer after a 2009 adultery scandal led to losses of \$1.5 million in profit and 136,000 individual sales, they conclude that the sponsorship was overwhelmingly beneficial for the brand. Pegoraro and Jinnah<sup>36</sup> discuss several case studies of the offline impacts of athletes' social media usage, including the case of former professional football player Larry Johnson, who was fired by his team in 2009 after a barrage of critical and offensive tweets aimed at the team's coach. Our study extends this previous research by concentrating specifically on the relationship between an athlete's online behavior and offline recruitment prospects.

## Modeling and predicting recruiting outcomes

Previous works have utilized Twitter data to predict college football commitments<sup>37</sup> and decommitments,<sup>38</sup> but focus on social network data as opposed to tweets. Although Pitts and Rezek<sup>39</sup> do not use social media

data, they analyze football recruiting data to create a model of the offer process and conclude that the physical attributes of a player are the most significant predictors of the number of scholarship offers. In contrast to Pitts and Rezek,<sup>39</sup> who study the relationship between demographic characteristics and the total number of offers at the end of recruitment, our work models the number of new offers received during junior year of high school based on the athlete's personal characteristics, recruiting activities, and Twitter

#### Methods

content to-date.

We begin by identifying instances of self-promotion and ingratiation in the Twitter accounts of college football recruits. We use this information on athletes' impression management activities, in addition to data about each athlete's personal characteristics and recruiting activities before the start of junior year of high school, to build a statistical model explaining the number of offers received over the next year.

#### Data

This study utilizes data on the recruiting activities and Twitter content of college football recruits. The data sources selected offer both depth—tracking each athlete's online and offline behavior over the course of several months—and breadth—capturing information about all of the athletes in the recruiting market.

In August 2015, we obtained recruiting profiles for 2644 athletes in the class of 2016 from the sports media website 247Sports.com.<sup>40</sup> For each athlete, the site lists basic information (e.g., height, weight, hometown, position, and rating) and time lines of recruiting events (e.g., scholarship offers, visits, and commitments). We scraped these data using the Selenium package for Python,<sup>41</sup> and the profiles were visited again in March 2016 to retrieve final recruiting events and commitments.

In addition to providing a comprehensive source of college football recruiting data, 247Sports.com was selected because many of its pages contain links to athletes' Twitter profiles (1629 of the 2644 athlete pages had Twitter IDs embedded). We conducted manual search to locate missing Twitter IDs yielding 700 additional IDs. In total, 2329 athletes (88.1%) in the data set were linked to public Twitter accounts, 160 (6.1%) had protected accounts, and 155 (5.9%) had no Twitter presence. We also gathered 764 Twitter IDs for Division I coaches and 2397 IDs for current college football players from 247Sports.

Detailed Twitter information for the athletes with public accounts was collected using the Twitter REST API<sup>42</sup> and the Tweepy package for Python.<sup>43</sup> In September 2015, we gathered profile information, friend and follower lists, and tweets (up to 3200 historical tweets for each athlete). Updated friend and follower lists and new tweets were collected monthly between October 2015 and March 2016. This resulted in a data set of more than 5.5 million tweets, detailed in Table 1.

The full tweet data set includes both original tweets and retweets, reposting tweets by other users. Retweets represent 47.1% of all tweets in our data set, a slightly higher proportion than has been observed among general Twitter users.<sup>44</sup> Retweets from college coaches, current college football players, or other recruits in the class of 2016 comprise 7.4% of athletes' retweets. Replies, tweets directed at other users or in response to other users' tweets, constituted 24.0% of all tweets. This proportion is also higher than 15.3% for the general population.<sup>44</sup> Athletes directed 9.0% of their replies at coaches, college football athletes, or other recruits in the class of 2016. Mentions, references to other users, were extremely common, occurring in 67.6% of all tweets. This was much higher than the 22.3% mention rate observed among general Twitter users.<sup>44</sup> Mentions of coaches, college football players, or class of 2016 recruits represent 27.5% of the mentions in the data set. There were 627,789 self-mentions, defined as tweets or retweets that mention the athletes themselves. Self-mentions occur most often in a situation where athletes retweet someone else who has mentioned them in the original tweet. Overall, these results indicate that football recruits tend to engage in more interpersonal interactions on Twitter than the general user population. However, the low proportions of interactions with college coaches, current players, and other recruits in the class of 2016 suggest that, on average, athletes' use of Twitter is primarily personal in nature, rather than focused on recruitment.

Tweets may also contain additional entities, including hashtags, URLs, and media. Hashtags, a word or

Table 1.	Overview	of tweet	data	set
----------	----------	----------	------	-----

Category	Count
Retweets	2,615,108
Replies	1,332,622
Mentions	3,751,757
Hashtags	734,213
URLs	751,650
Media	983,759
Total	5,547,230

phrase used to denote a tweet topic, exist in 13.2% of tweets. Hashtags were cross-referenced with a handcurated dictionary to determine whether they were relevant to college football. Approximately 7.3% of hashtags referenced football topics. URLs exist in 13.6% of tweets in the data set, 11.8% of which linked to recruiting or sports media websites (e.g., Hudl, Rivals, or ESPN). Further analyzing the sports-related URLs, we were able to determine whether each URL referred to the posting user or another athlete or team. Self-referencing URLs constituted 46.5% of all sports URLs, while 53.5% referenced others. While we did not analyze the content of media (images, GIFs, videos, etc.), 17.7% of athletes' tweets contained media.

#### Defining impression management

We base our definitions of impression management in athletes' tweets on the Jones and Pittman<sup>17</sup> taxonomy. Although the authors describe five categories of self-presentation, we focus on self-promotion and ingratiation, as they occur most frequently in recruits' social media and have been clearly linked to recruiting success in the HR domain.<sup>18</sup> In this study, we do not consider exemplification (a strategy intended to project "moral worthiness" through good acts), intimidation (aggressive behavior meant to evoke fear), and supplication (showing respect for another party by portraying oneself as weak or dependent).

Individuals who engage in self-promotion are attempting to enhance perceptions of competence.<sup>17</sup> For college football recruits, self-promotion on Twitter is likely to take the form of posting tweets publicizing their athletic ability and recruitment activities. Athletes commonly post links to their highlights on Hudl—an online athletic video service—or tweet at coaches to gain their attention. For example, 2-star recruit JaVaughn Craig posted a tweet directed at a Duke assistant coach, "@JeffFaris Hey Coach, can you follow back please. Thanks."

Ingratiation is intended to produce the "attribution of likability,"<sup>17</sup> and often takes the form of compliments or opinion conformity. Ingratiation may be intended for a broad audience or targeted to appeal to a specific person. In the context of college football recruiting, ingratiation is likely to take the form of mentioning coaches, current college players, and other recruits or positive tweets about other athletes and teams. Khris Pam, a 3-star recruit, tweeted his congratulations to another class of 2016 recruit from his home state of South Carolina, "@ShaedonMeadors Glad Y'all Beat Them Boys Bro."

Our study assumes that impression management strategies are nonexclusive, that is, a tweet may be simultaneously self-promoting and ingratiating. For example, Bryce Huff, a 2-star recruit, tweeted "I'm blessed to say that I've received an offer from Troy University!" This tweet could be interpreted as self-promotion, with Huff publicizing his recruiting success, or as ingratiating, intended to compliment Troy University by expressing his appreciation at being recruited by the program. Our approach intends to reflect the complexity of online self-presentation, in contrast to previous analyses of social media in employment screening<sup>26</sup> that placed social media profiles into narrowly defined, discrete categories.

As almost half of all tweets posted on athletes' time line were retweets, we consider both original content and retweets in our definitions of self-promotion and ingratiation. Michael Boykin, a 3-star recruit, reposted a tweet by another user promoting Boykin's own recruiting success and containing a mention of his Twitter ID, "#Cincinnati offers 2016 Stud DL Mike Boykin (Carrolton HS, GA) @mikebfeb\_13 @VarsityPreps @SouthRecruit1." While Boykin was not the original author of the tweet, posting it on his time line was a clear attempt at self-promotion. Mass media reports indicate that coaches pay as much attention to retweets as likes and follows,<sup>1</sup> a sentiment echoed by a coach responding to the Cornerstone Reputation survey,<sup>11</sup> "Certainly the company they keep or allow to follow and post on their walls/pages can have a negative or positive effect accordingly, as well. That is almost as important as the original posts the potential student-athlete makes."

# Classifying impression management from tweets

We construct two separate supervised classifiers to learn the features associated with self-promotion and ingratiation. We randomly selected 7000 tweets (both original and retweets) and manually labeled them as self-promotion, ingratiation, both, or neither. A second annotator independently labeled 100 tweets so that the robustness of the class definitions could be verified. For self-promotion, the annotators initially agreed on 83% of examples, yielding a Cohen's kappa coefficient<sup>45</sup> of 0.82. After discussion, four discrepancies were resolved in favor of the first annotator, resulting in 87% agreement. For ingratiation, the annotators initially agreed on 87% of examples ( $\kappa = 0.87$ ). After discussion, three discrepancies were resolved in favor of the first annotator, resulting in 90% agreement. Because of the high level of consensus, it was determined that the first annotator's labels could be used. Table 2 displays the

Table 2. Manually labeled tweet data

Class	Description	Coun	
Self-promotion	Tweet or retweet that promotes the athlete's recruiting success or athletic ability	1382	
Ingratiation	Tweet promoting or praising another athlete, coach, team, or school	852	
Both	Tweet or retweet that contains both self- promotion and ingratiation	153	
Neither	Tweet or retweet that contains neither self- promotion nor ingratiation	4613	
Total	. 2	7000	

number of manually labeled tweets belonging to each class. Self-promotion and ingratiation constituted 19.7% and 12.2% of the manually labeled tweets, respectively, with an additional 2.2% labeled as both self-promoting and ingratiating.

When constructing our tweet classifiers, we considered features representing both the structure (retweets, mentions, etc.) and the content of athletes' tweets (Table 3). The text of each tweet was preprocessed using the NLTK package for Python.<sup>46</sup> We removed non-ASCII characters (e.g., symbols, emojis), standard English stopwords (e.g., a, the, that), and the 0.1% most- and least-common words. Because hashtags were analyzed separately to determine their relevance

Table 5. Features of tweet classifier	Table	3.	Features	of	tweet	classifier
---------------------------------------	-------	----	----------	----	-------	------------

Туре	Feature	Description
Structural	Retweet	1 if tweet is retweet or quote, 0 otherwise
	Coach retweet	1 if tweet is retweet/quote of coach, 0 otherwise
	College retweet	1 if tweet is retweet/quote of college player, 0 otherwise
	Recruit retweet	1 if tweet is retweet/quote of recruit
	Reply	1 if tweet is reply, 0 otherwise
	Coach reply	1 if tweet is reply to coach, 0 otherwise
	College reply	1 if tweet is reply to college player, 0 otherwise
	Recruit reply	1 if tweet is reply to recruit, 0 otherwise
	Mention	1 if tweet contains mention, 0 otherwise
	Coach mentions	No. of coach mentions
	College mentions	No. of college player mentions
	Recruit mentions	No. of recruit mentions
	Self-mention	1 if tweet mentions athlete, 0 otherwise
Content	Sports hashtags	No. of hashtags referencing college football
	URL	1 if tweet contains URL, 0 otherwise
	Self-URLs	No. of football-related URLs about athlete
	Other URLs	No. of football-related URLs about others
	Media	1 if tweet contains media, 0 otherwise
Text	Words	Bag-of-words representation of tweet text

to college football, they were also removed. In addition, entities (e.g., mentions, URLs) were replaced with representative tokens (e.g., "CoachMention"). For example, the mention and URL in Austin McCall's tweet "I've picked Clemson to be my @drpepper #onefinalteam https://t.co/dluv6atkjg" were replaced to yield "I've picked Clemson to be my OtherMention Other-Hashtag." We utilized a simple bag-of-words model<sup>47</sup> to represent the tweet text, where each document is transformed into a k-vector containing the number of occurrences of each word, where k is the number of unique words in the corpus. Athletes' tweets were constructed out of  $\sim$ 750 unique words. We used stemming,<sup>47</sup> a rule-based process that removes English suffixes (e.g., -es, -ment, -ly) to decompose words to their base forms, to reduce the size and sparsity of the feature space. Applying a Porter stemmer reduced the number of unique words to  $\sim$  720.

We construct the classifiers using the scikit-learn package for Python.<sup>48</sup> As different classification methods are suited to different types of problems, we perform a preliminary test of several methods: logistic regression, decision tree, naive Bayes, support vector machines (SVM), artificial neural networks (ANN), and random forest. We use a grid search to select the appropriate parameters, listed below:

- Logistic regression: No regularization penalty was applied.
- Decision tree: Entropy was selected as the split criterion for the decision nodes. To prevent overfitting, the maximum tree depth was set to 50 levels.
- Naive Bayes: An additive smoothing factor of 0.25 was used to avoid the issue of zero estimates for unknown terms.
- SVM: We implemented a linear kernel function with a penalty weight of 1.
- ANN: The rectified linear unit function was selected as the transfer function between 20 layers, with the Adam algorithm for weight optimization.
- Random forest: The ensemble size was set to 100 trees, each using the Gini coefficient for split criterion.

We use the Monte Carlo cross-validation to assess classifier performance. Over 100 independent trials, the full set of 7000 labeled tweets was randomly split into equal-sized training and testing subsets. We oversample the training set by randomly selecting and copying positive instances until it contains 50% positive instances. Then, we train each classification method on the oversampled data, with consideration of the same set of features (Table 3), and evaluate on the unaltered testing set. We measure the performance of each classifier using standard metrics: accuracy (proportion of correctly predicted instances), precision (ratio of true positives to predicted positive instances), recall (ratio of true positives to actual positive instances), F1 score (harmonic mean of precision and recall), and area under the receiver operating characteristic curve (AUC), measuring the probability of ranking a randomly chosen positive instance higher than a randomly chosen negative instance.

# Modeling factors associated with offers

Scope of analysis. The task of creating an explanatory model of scholarship offers is complicated by the fact that offers are made over a long period of time. Although coaches cannot begin directly contacting athletes until the start of their junior year,<sup>12</sup> it is not unprecedented to make scholarship offers (sent to the athlete's high school coach or school) as early as freshman and sophomore year. Coaches' decision-making processes are almost certainly different when choosing whether to extend an offer to a freshman versus a se-

nior, especially considering the quantity and quality of information available in each case. Figure 1 displays the total number of scholarship offers received each month by athletes in the class of 2016. Seventy-eight percent of all scholarship offers were received during junior year of high school and following summer, and we elect to narrow our analysis to this period (September 1, 2014, to August 1, 2015).

The scope of our analysis is also limited by the amount of historical tweets available for each athlete. The Twitter API<sup>42</sup> allows retrieval of up to 3200 historical tweets for each user. Because account age and tweet frequency vary by individual, this figure covers different windows of time for each athlete. Measured as of the start of senior year (September 1, 2015), account ages for recruits with public Twitter accounts ranged from 6 days to 6.5 years (both mean and median account ages were ~ 2.5 years). Figure 2 displays the distribution of account ages for athletes in the class of 2016.

Athletes in the class of 2016 tweeted 130 times per month on average, although the distribution of Twitter activity is highly skewed. Measured as of September 1, 2015, athletes with public Twitter accounts posted between 0 and 1800 tweets per month, with a median of 66 (Fig. 3).





We were able to retrieve all historical tweets for 72.1% of athletes with public Twitter accounts and more than half of historical tweets for 85.9%. Figure 4 shows how the historical range of our data aligns with the college football recruiting time line. We possess one or more months of historical tweets posted before the beginning of junior year for 73.3% of athletes with public Twitter accounts. By restricting the scope of our analysis to junior year, we focus on the time period when most offers are occurring and ensure fairly consistent coverage of individual athletes' previous Twitter activities.

Statistical modeling. We create a set of instances, including all athletes with historical tweets available as of September 1, 2014, as well as those who do not possess a Twitter account or have protected accounts. This leaves 1792 athletes who received 7261 offers during junior year of high school (Fig. 5).

Because our outcome—the number of new offers received during junior year—is a count variable that exhibits overdispersion ( $\mu$ =4.05,  $\sigma$ =21.81), we utilize negative binomial regression implemented in R.<sup>49</sup> We create an instance for each athlete with independent variables tracking the personal characteristics, recruiting activities, and social media impression management up to the beginning of junior year of high school (Table 4).

We build a series of models to investigate the relationship between impression management and scholarship offers. The baseline (Model 0) models the number of new offers as a function of the athlete's demographic characteristics and recruiting activities up to the start of junior year. Model 1 considers the effect of Twitter engagement on the number of offers, regardless of the content of the athlete's tweets. Because Twitter activity varies widely by individual—12.5% of athletes posted 0 public tweets during their junior year, while 11.7% averaged 300 or more per month we create a categorical variable grouping the athletes with public Twitter accounts by average number of tweets per month (measured as of September 1,



2014). We assigned labels of low, medium, high, or very high activity approximately by quartiles. Both athletes with protected accounts and those who do not possess a Twitter account are assumed to have 0 tweets per month. To preserve the distinction between these two groups and athletes who rarely tweet, we also include levels representing account status (none, protected). Model 2 incorporates the proportion of tweets posted by the athlete up to the beginning of junior year that were labeled as self-promotion, and Model 3 considers the proportion of ingratiating tweets. Both athletes with protected accounts and those who do not possess a Twitter account are assumed to have 0 tweets per month and post no self-promoting or ingratiating tweets. We also create a combined model (Model 4), which includes all of the features of the previous models.

We also investigate whether these factors have different impacts on athletes of varying ability levels, as measured by star rating. Ratings of 2, 3, 4, or 5 stars are assigned by major recruiting websites as an estimate of a recruit's ability and potential. There are no 1-star recruits, although some athletes do not receive enough recruiting attention to earn a rating, and we treat these instances as a 0-star rating. Although it is an imperfect proxy for an athlete's latent talent, it is a measure that is readily observable and positions each individual on the same scale. An ANOVA shows that the number of offers varies significantly by star rating ( $p < 2 \times 10^{-16}$ ). A post hoc Tukey HSD test<sup>50</sup> confirms that there is no significant difference between 0- and 2-star recruits (p=0.2965), but that both earn significantly fewer offers than their higher rated peers. Similarly, athletes with 4 or 5 stars earn significantly more offers than lower rated athletes, but the difference between them was only marginally significant (p = 0.0817). We divide the full set of athletes into "low-rated" recruits with 0 or 2 stars (n=528), "midrated" recruits with 3 stars (n=1035), and "high-rated" recruits with 4 or 5 stars (n=229) and compare the results of applying the full model to each group.



#### Results

The goals of this study are to document the type of online impression management strategies used by athletes during recruitment and examine their relationship with offline recruiting outcomes. The following subsections report the results of tweet classification and the fitted negative binomial regressions modeling the number of offers received during junior year.

#### Tweet classification

As a preliminary test, we evaluated six different classification methods on the task of identifying selfpromoting and ingratiating tweets. Table 5 displays the comparative performance of each classifier, averaged over 100 trials using different training/testing partitions. Standard deviations are shown in italics next to each column.

These results indicate that logistic regression has the best overall performance for identifying self-promotion, achieving the highest AUC and F1 scores. *T*-tests demonstrate that these differences are significant. The logistic classifier has a significantly higher AUC ( $p < 2.2 \times 10^{-16}$  for decision tree, SVM, ANN, and random forest, p = 0.02771 for naive Bayes) and F1 score ( $p < 2.2 \times 10^{-16}$  for decision tree, naive Bayes, SVM, and ANN,  $p = 1.54 \times 10^{-9}$  for random forest) than all of the other methods. In addition, it has the second-highest scores for accuracy and recall, and the third-highest precision.

The logistic classifier also achieves the best overall performance for identifying ingratiation. It has significantly higher AUC ( $p < 2.2 \times 10^{-16}$  for decision tree, ANN, and random forest,  $p < 3.16 \times 10^{-13}$  for naive Bayes, p = 0.0003 for SVM) and F1 scores ( $p < 2.2 \times 10^{-16}$  for decision tree, naive Bayes, ANN, and random forest,  $p < 9.58 \times 10^{-9}$  for SVM) than all of the other methods. In addition, it earns the highest recall score, significantly outperforming the decision tree ( $p < 2.2 \times 10^{-16}$ ), SVM (p = 0.08), ANN ( $p < 2.2 \times 10^{-16}$ ), and random forest classifiers ( $p < 2.2 \times 10^{-16}$ ). Logistic regression has the second-highest accuracy and third-highest precision. We note that all five classification methods achieve lower AUC, precision, recall, and F1 scores when attempting



to identify ingratiating tweets, possibly due to the lower prevalence of ingratiation in the labeled training data.

Due to its consistently good performance on both tests, we utilize logistic regression to identify occur-

rences of self-promotion and ingratiation in the full set of 5.5 million tweets. After applying the trained logistic classifiers to all tweets, 1,065,735 (19.2%) are classified as self-promotion and 772,099 (13.9%) as ingratiation. These figures are consistent with the

Table 4. Offer regression independent variables

Туре	Variable	Description
Personal	Star	Recruit star rating (0, 2, 3, 4, 5)
	Height	Height (inches)
	Weight	Weight (pounds)
	BMI	Body mass index
	Position	Recruit position (ATH, DB, DL, LB, OB, QB, RB, ST)
	Region	U.S. census region of recruit (MW, NE, S, W)
Recruiting	Updates	No. of news updates on 247Sports.com to-date
5	Camps	No. of college camps attended to-date
	Unofficial	No. of unofficial visits to-date
	Coach	No. of coach visits to-date
	Offers	No. of offers received to-date
	FBS	Proportion of offers to-date from Division I, FBS schools
	Verbal	1 if athlete is currently committed, 0 otherwise
Impression management	Activity	Level of Twitter activity based on account status (none, protected) or average monthly tweets to-date ( $low = 1-20$ , $mid = 21-60$ , $high = 61-160$ , $highest = 160+$ )
5	Self-promotion	Proportion of self-promoting tweets posted to-date
	Ingratiation	Proportion of ingratiating tweets posted to-date

FBS, Football Bowl Subdivision.

Table 5. Tweet classifier testing results

	Method	Αςςι	ıracy	A	JC	Prec	ision	Re	call	F	1
		Mean	SD								
Self-promotion	LOG	0.866	0.005	0.792	0.007	0.711	0.017	0.660	0.014	0.684	0.010
	DT	0.859	0.006	0.762	0.011	0.719	0.023	0.588	0.024	0.647	0.017
	NB	0.844	0.005	0.790	0.007	0.630	0.015	0.695	0.015	0.661	0.010
	SVM	0.858	0.006	0.783	0.007	0.687	0.019	0.649	0.015	0.667	0.011
	ANN	0.847	0.008	0.761	0.012	0.665	0.028	0.610	0.028	0.635	0.016
	RF	0.873	0.006	0.774	0.010	0.773	0.030	0.598	0.025	0.673	0.014
Ingratiation	LOG	0.876	0.005	0.743	0.012	0.570	0.021	0.557	0.025	0.563	0.017
-	DT	0.872	0.006	0.690	0.013	0.572	0.031	0.435	0.031	0.493	0.021
	NB	0.858	0.006	0.731	0.010	0.502	0.021	0.554	0.021	0.527	0.016
	SVM	0.871	0.005	0.737	0.011	0.548	0.022	0.551	0.024	0.549	0.017
	ANN	0.870	0.006	0.706	0.012	0.554	0.030	0.476	0.028	0.511	0.018
	RF	0.884	0.005	0.693	0.013	0.646	0.034	0.425	0.030	0.511	0.021

ANN, artificial neural networks; AUC, area under the receiver operating characteristic curve; DT, decision tree; LOG, logistic regression; NB, naive Bayes; RF, random forest; SD, standard deviation; SVM, support vector machines.

proportions of self-promoting and ingratiating tweets observed in the labeled training data, 21.9% and 14.4%, respectively.

## Regression models of new offers

To estimate the relationship between the athlete's personal characteristics, recruiting activities, and Twitter content before the start of junior year and the number of new offers received during junior year, we constructed five negative binomial regression models. Table 6 lists the regression coefficients, significance, pseudo-R-squared,<sup>51</sup> and mean squared error (MSE) for each of the models. Note that the coefficients shown represent the additive effect of each feature on the log of the number of offers. We interpret the multiplicative effect of each feature on the number of offers by applying the exponential function ( $e^x$ ).

Examining Model 0, star rating has a positive and significant effect. In comparison with athletes with 0 stars, the number of offers increases 85%, 361%, 956%, and 1225% for 2-, 3-, 4-, and 5-star recruits, respectively. Region also has a significant effect. Compared with their peers from the Midwest, the number of offers increases 16.3% for athletes from the South and decreases 46.3% for athletes from the West. In terms of playing position, special teams (kickers and punters) receive 58.5% fewer offers than generalpurpose athletes. Attending training camps where the athlete can gain the attention of college coaches is associated with receiving additional offers, 3.6% for each camp, although the significance of this effect is not significant in subsequent models. The effects of other recruiting events (unofficial visits and coach visits) are not significant. The number and quality of offers received by the athlete before the start of junior year are also significant. For each current offer, the number of offers received during junior year increases 10.7%. Interestingly, attracting early attention from higher division programs is associated with decreased recruiting success during junior year. The number of new offers decreases 2.7% for each 1% increase in proportion of total offers from Football Bowl Subdivision (FBS) teams. Intuitively, having a current verbal commitment is associated with a 29.8% decrease in number of new offers compared with uncommitted athletes.

Model 1 focuses on the effect of Twitter account status and average activity level on the number of offers received during junior year. Compared with athletes without a Twitter account, a protected account is associated with a 118% increase in the number of new offers. Low activity—averaging 1–20 tweets per month—is associated with receiving 154% more offers. Relative to athletes without Twitter, the number of offers increased 179%, 158%, and 149% for athletes demonstrating mid (21–60 tweets on average), high (61–160 tweets), and very high activity (>160 tweets), respectively. All of the Twitter activity features are highly significant.

Model 2 incorporates data about the athlete's online self-promotion efforts. The proportion of self-promoting tweets—calculated as the number of self-promoting tweets to-date divided by total tweets to-date—is highly significant. Each 1% increase in self-promotion increases the number of offers by 1%.

Model 3 considers the impact of ingratiating behavior. Each 1% increase in proportion of ingratiating tweets posted to-date increases the log number of offers received by 1.6%.

#### Table 6. Offer regression results

Туре	Feature	Model 0	Model 1	Model 2	Model 3	Model 4
Personal characteristics	Intercept	1.0643	0.2202	-0.8879	0.0195	-0.6867
	2 Stars	0.6171***	0.5555***	0.5915***	0.6050***	0.5495***
	3 Stars	1.5286***	1.4360***	1.4958***	1.5070***	1.4265***
	4 Stars	2.3570***	2.2340***	2.3145***	2.3350***	2.2206***
	5 Stars	2.5842***	2.4310***	2.5803***	2.6293***	2.4429***
	Height	-0.0146	-0.0142	0.0114	-0.0011	-0.0011
	Weight	0.0103	0.0102	0.0055	0.0081	0.0080
	BMI	-0.0801	-0.0808	-0.0446	-0.0657	-0.0654
	NE	-0.1050	-0.0713	-0.0869	-0.0959	-0.0750
	S	0.1511*	0.1783**	0.1507*	0.1557**	0.1728**
	W	-0.6213***	-0.5680***	-0.6270***	-0.6249***	-0.5740***
	DB	0.1631	0.1861	0.1617	0.1706	0.1881
	DL	0.0649	0.1263	0.0707	0.0743	0.1259
	LB	0.0633	0.0853	0.0568	0.0460	0.0781
	OB	0.0717	0.0913	0.0540	0.0741	0.0886
	OL	-0.0507	0.0001	-0.0353	-0.0544	0.0027
	QB	-0.1372	-0.1543	-0.1771	-0.1735	-0.1722
	ST	-0.8787**	-0.8354**	-0.9070**	-0.8803**	-0.8500**
Recruiting events	Updates	0.0508	0.0460	0.0400	0.0375	0.0346
-	Camps	0.0351°	0.0279	0.0303	0.0279	0.0244
	Unofficial	-0.0178	-0.0202	-0.0238	-0.0271.	-0.0260
	Coach	-0.0111	-0.0098	-0.0146	-0.0132	-0.0130
	Offers	0.1018***	0.1005***	0.0955***	0.0975***	0.0968***
	FBS	-0.0276***	-0.0269***	-0.0274***	-0.0275***	-0.0269***
	Verbal	-0.3543**	-0.3670**	-0.3766**	-0.3671**	-0.3898**
Twitter activity	Low		0.9310***			0.7359***
	Mid		1.0250***			0.8875***
	High		0.9461***			0.8214***
	Highest		0.9134***			0.7920***
	Protected		0.7770***			0.7816***
Impression management	Self-promotion			0.0099***		0.0035
	Ingratiation				0.0162***	0.0085**
Pseudo-R-squared	5	0.3651	0.4025	0.3778	0.3811	0.4092
MSE		14.485	13.850	14.294	14.196	13.604

\*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05, ° p < 0.1.

BMI, body mass index; MSE, mean squared error.

Model 4 includes all of the features tested in the previous models. Overall Twitter activity remains highly significant, and increasing levels of engagement are associated with receiving 109%–143% more offers during junior year. In comparison with recruits with no Twitter account, a protected account is associated with an increase of 118%. Interestingly, self-promotion is not a significant predictor in the final model, although its coefficient shows the expected sign. We find that each 1% increase in proportion of ingratiating tweets to-date is associated with a 0.9% increase in number of new offers.

We also compare the models based on measures of fit. First, we consider pseudo-R-squared.<sup>51</sup> All of the models incorporating social media data achieve higher values than the baseline containing only recruiting data. Merely considering the effect of Twitter account status and activity level (Model 1) achieves a 10% increase in pseudo-R-squared over the baseline. The self-promotion model (Model 2) has a slightly lower

pseudo-R-squared value than the ingratiation model (Model 3), indicating that ingratiation explains more of the variation in number of offers than selfpromotion. Ultimately, the combined model, including all of the recruiting and social media features, achieves the highest pseudo-R-squared and lowest MSE. The distinguishability and goodness of fit of the three models are also compared using a Vuong test.<sup>52</sup> This test calculates whether models equally approximate the true data generating process, against the alternative that one model is closer. All of the models incorporating Twitter data significantly improve the fit of the model  $(p=8.68\times10^{-8} \text{ for Model 1}, p=0.001 \text{ for})$ Model 2, p = 0.0005 for Model 3, and  $p = 8.95 \times 10^{-9}$ for Model 4). Comparing Model 0 with Models 1 and 2, the Vuong test determines that Model 1 is closer (p=0.001 for Model 2 and p=0.0017 for Model 3),indicating that adding data about the athlete's overall Twitter activity level contributes more to the fit of the model than adding either of the impression

management features alone. Model 4 significantly outperforms all of the other models (p=0.015 for Model 1,  $p=6.61 \times 10^{-7}$  for Model 2, and  $p=3.79 \times 10^{-6}$  for Model 3), suggesting that a combination of Twitter features is a better fit for the data.

Effects of self-presentation by star rating Table 7 displays the results of applying the selfpromotion and ingratiation models to the star-specific samples, including regression coefficients, significance, pseudo-R-squared,<sup>51</sup> and MSE.

In contrast to the full set of recruits, attendance at training camps appears to be most beneficial for lowrated (0- and 2-star) recruits. Each camp is associated with a 16.3% increase in the number of new offers received during junior year. The number of current offers is significant for the midrated (3-star) and high-rated (4- and 5-star) groups, and the proportion of previous offers from FBS schools is negative and significant across all of the star-specific samples. Interestingly, the coefficient representing a current verbal commitment is only significant for 4- and 5-star recruits, associated with a 36.2% decrease in the number of new offers. This is likely explained by the fact that the distribution of early commitments varies by star rating. Only 0.004% of low-rated and 2.9% of midrated athletes were committed at the start of junior year. In contrast, 14.8% of athletes with 4 or 5 stars were verbally committed.

The effect of Twitter activity is amplified for lowrated recruits. In comparison with low-rated recruits without a Twitter account, a protected account is associated with a 259% increase in number of new offers, and increasing levels of activity are associated with receiving 116%–187% more offers during junior year. The effect of self-promotion is also higher for lower rated recruits, with each 1% increase in proportion of self-promoting tweets associated with a 1.2% increase in number of new offers. While the effect of selfpromotion is not significant for the mid- and highrated samples, the coefficients show the expected sign. Interestingly, the effect of ingratiation is not significant for any of the star-specific samples, although its coefficient is consistently positive.

#### Discussion

This study examines the use of online social media by college football recruits as a platform to engage in selfpromotion and ingratiation and analyzes the impact of these strategies on the number of scholarship offers re-

Table 7. Star-specific regression

Туре	Feature	Low star	Mid star	High star
Personal	Intercept	0.6769	-6.8652	14.0142
characteristics	Height	-0.0140	0.1004	-0.1653
	Weight	0.0036	-0.0060	0.0338
	BMI	0.0353	0.0460	-0.2720
	NE	-0.1591	-0.0763	0.0538
	S	0.0926	0.2295**	-0.0132
	W	-0.7068**	-0.6654***	-0.2941*
	DB	0.0128	0.2616°	0.4053°
	DL	0.1182	0.0973	0.3876°
	LB	-0.0527	0.1172	0.3900°
	OB	0.0622	0.0994	0.2194
	OL	0.0596	0.0091	0.2085
	QB	-0.2408	-0.1407	0.1496
	ST	-0.6150	-1.0137**	0.0000
Recruiting events	Updates	-0.1873	0.0180	-0.0168
	Camps	0.1512**	-0.0206	0.0084
	Unofficial	-0.0187	-0.0086	-0.0123
	Coach	0.3925	-0.0251	-0.0031
	Offers	0.5088	0.3436***	0.1042***
	FBS	-0.0404*	-0.0364***	-0.0332***
	Verbal	1.1919	-0.1672	-0.4500***
Twitter activity	Low	0.7716***	0.7895***	0.5733*
	Mid	1.0541***	0.8673***	0.7475**
	High	0.9513***	0.8197***	0.6825**
	Highest	0.8660***	0.8269***	0.7779**
	Protected	1.2777***	0.6422***	0.7102**
Impression	Self-promotion	0.0116*	0.0025	0.0025
management	Ingratiation	0.0077	0.0056	0.0058
Pseudo-R-squared		0.1792	0.2796	0.6660
MSE		4.3265	15.529	18.374

\*\*\*p < 0.001, \*\*p < 0.01, \*p < 0.05, °p < 0.1.

ceived over time. While the role of in-person impression management in job interviews has been widely studied, this research is the first to examine impression management in the context of college football recruiting. Our work has significant implications for college football recruits and can inform their online communications during recruitment.

Building supervised classifiers to identify impression management in 5.5 million tweets posted by athletes in the class of 2016, we find that athletes tend to post more self-promoting (19.2%) than ingratiating content (13.9%). This result is consistent with previous studies examining impression management in employment interviews,<sup>18</sup> supporting our application of theory from the HR domain to the athletic recruiting context.

We use this labeled tweet data, in addition to recruiting data from 247Sports.com, to perform a statistical analysis of the relationship between the number of offers received during junior year of high school and the athlete's personal characteristics, recruiting activities, and Twitter content before the start of junior year. Focusing on static, demographic features, we find that star rating is a highly significant predictor of the number of new offers. In contrast to previous work on the offer process in college football recruiting,<sup>39</sup> the athlete's physical features do not have a significant effect on the number of offers received during junior year. We find evidence of a regional effect, particularly among midrated recruits, with athletes from the West earning fewer offers and athletes from the South garnering significantly more offers than their peers. This result is partially consistent with Pitts and Rezek's<sup>39</sup> analysis, which found that recruits from Florida and Texas (both located in the South) receive more scholarship offers than those from other states.

While the features representing news updates and coach visits to-date are not significant, attendance at training camps is a positive predictor of new scholarship offers in the baseline model. Indeed, such events function as opportunities for athletes and coaches to evaluate each other and engage in offline impression management, similar to a job interview. Recruits have the chance to show off their athletic skills and coaches can gauge "intangible" qualities such as attitude and maturity through interpersonal interactions. Our analysis suggests that these opportunities are particularly important for lower rated (0- and 2-star) recruits. Notably, NCAA regulations<sup>12</sup> severely limit coaches' chances to communicate with freshman and sophomore athletes outside of the training camp setting.

We also find that the number of offers received before the start of junior year has a positive and significant effect on new offers received over the next year, especially for 3-, 4-, and 5-star recruits. Intuitively, the most valuable recruits are likely to already possess one or more scholarship offers. However, the proportion of offers from FBS schools is a negative factor across all recruits in the class of 2016, suggesting that college teams are reluctant to expend resources recruiting an athlete who already has multiple (possibly superior) alternatives. Highly rated (4- or 5-star) recruits who were already verbally committed to a school were also less likely to receive more new offers than their similarly rated peers.

Recruits' usage of social media varied widely; from those with no Twitter accounts, to posting rarely (24% of athletes with public accounts posted fewer than 10 times per month), to posting almost incessantly (6% of athletes with public accounts posted more than 10 times per day). We find that Twitter account status and activity level (based on average number of tweets per month) are significant predictors of the number of new offers earned during junior year. In general, more Twitter activity is associated with higher odds of receiving a scholarship offer, but even a protected (nonpublic) Twitter account represents a significant advantage. This finding is notable, as it contradicts previous statements by coaches expressing suspicion in regard to protected social media accounts. In the Cornerstone Reputation survey,<sup>11</sup> a college coach states that "For me, when a recruit has a Twitter or Instagram account that is private it sends up a red flag. Anything they post should be visible to coaches or they shouldn't be posting it!"

While we find that posting self-promoting and ingratiating content online consistently has a positive effect on the number of new offers, regression results suggest that neither strategy is as important as overall Twitter activity. The features tracking account status and average activity level explained  $\sim 4\%$  of the variation in number of offers, more than self-promotion and ingratiation combined. Comparing the models focusing on self-promotion and ingratiation, the coefficients suggest that the latter may have a slightly larger impact on recruiting success during junior year. In addition, self-promotion is not significant in the full model when applied to all recruits. This result is not consistent with previous studies of impression management in HR recruiting, where self-focused tactics resulted in higher interviewer ratings<sup>18</sup> and fewer rejections<sup>15</sup> than other-focused tactics. The larger effect of ingratiation in our analysis may be related to the ready availability of physical measurements and performance statistics on platforms such as 247Sports and Rivals. Thus, self-promotion efforts would be unlikely to impress coaches and offer little new information for decision-making. It should be noted that we focus on offline and online events that occur early in the recruiting process-before the start of junior year-a period when coaches have little direct access to recruits. Ingratiating content can signal intangible factors such as academic ability or good character that would be difficult to judge without face-to-face communication. Ninetynine percent of coaches surveyed by Cornerstone Reputation<sup>11</sup> rated character as important or very important, and 85% believed that online presence gave a "better sense of a recruit's character and personality."

Our results also suggest that the positive effect of social media activity varies by recruits' abilities. In comparison with the full set of recruits, the effects of Twitter account status and activity level are higher for the lower rated sample (0- and 2-star recruits), suggesting that social media may play a bigger role in their recruiting success in comparison with their higher rated peers. Any recruit who has earned a 4- or 5star rating is almost certain to attract new offers, regardless of the engagement on Twitter. Social media is a convenient and wide-reaching medium that offers unheralded athletes with a way to stand out from the crowd. In contrast to the full set of recruits, selfpromotion appears to yield the most value for lower rated athletes. Furthermore, the effect of selfpromotion is relatively universal. Posting a Hudl highlight video on Twitter allows the material to be disseminated to many coaches at once and increases the chances of garnering multiple offers, while sending the same video in the mail to a specific college would only increase the chances of getting an offer from that college.

We note several limitations of this study. First, the accuracy of the recruiting data analyzed is dependent on the source. 247Sports incorporates both usersupplied and expert data, which we believe makes it both a comprehensive and up-to-date source of recruiting information. Second, we use data on a single recruiting class, and results may differ for other recruiting years, especially due to the constantly evolving nature of NCAA rules. Since the period of data collection for the class of 2016, the NCAA has lifted its ban on text messaging between college football coaches and recruits.<sup>53</sup> While there is no indication that this development has resulted in social media taking a smaller role in the recruiting process, it does represent a change to the environment that should be investigated in future work. Furthermroe, starting in 2017, the NCAA established an early signing period consisting of 3 days in December in which athletes can sign a Letter of Intent rather than waiting until the following February.<sup>54</sup> Finally, as this is not an experimental study, no causality can be inferred from the results. Our findings indicate that, in addition to data available from recruiting websites such as 247Sports, athletes' social media content adds value to explanatory models of the number of scholarship offers received during junior year. While our findings suggest that overall Twitter engagement and impression management are associated with increasing number of offers, we cannot state that these variables cause offers.

# Conclusion

Overall, we find that considering social media data, specifically account status, activity level, and impression management strategies used by athletes on Twitter, adds value to an explanatory model of scholarship offers and suggests that athletes' online actions may have offline consequences. Our regression results indicate that having any type of online presence represents a significant benefit to athletes during recruitment.

This work represents a promising step in modeling and predicting scholarship offers in college football and provides a framework for using candidates' online behaviors to forecast offline recruiting outcomes. While this study focuses on football, our approach could easily be applied to other sports and personnel recruiting. Further investigation into different types of online self-presentation may prove fruitful, including using topic discovery to identify new categories of expression. It should be noted that our study focuses on positive self-presentation on social media, and it is possible that exhibiting bad character traits through racist, sexist, vulgar, or aggressive content could have a much larger effect on scholarship offers. Currently, our analysis is school-neutral, aimed at modeling and predicting new offers from any college. In application, athletes may be more interested in the question of whether they will receive an offer from a specific college. Focusing on the relationship between targeted self-presentation strategies and offers (e.g., positive tweets about Iowa and the likelihood of receiving an offer from Iowa) is an intriguing extension of this work. Impression management relates naturally to questions of person-organization and person-job fit. Deeper analysis of tweet content, such as estimating the topical similarity between coach and recruit tweets, may provide insight into athlete-school fit. Gathering additional data on athletes' postrecruitment outcomes represents another promising angle for future work. Personnel management and player evaluation are major drivers of the sports analytics revolution, popularized by works such as Moneyball.55 Linking data available during the recruitment process about athletes' personal characteristics, recruiting activities, and social media to later outcomes (e.g., on-field success, draft prospects, academic ineligibility, legal issues) would be a novel extension of this research with the potential to provide significant benefits to college teams.

# Acknowledgments

Some of the material in the article is taken from the PhD dissertation of K.G.B. (University of Iowa, Information Science, 2018). From Hashtags to Heismans: Social Media and Networks in College Football Recruiting

(10837758). It is available from ProQuest (https://search. proquest.com/docview/2118481920) as well as the University of Iowa institutional repository (https://ir.uiowa. edu/etd/6371).

#### **Author Disclosure Statement**

No competing financial interests exist.

#### References

- Crabtree J. 2016. The 'social' science of recruiting. ESPN. Available online at www.espn.com/college-football/recruiting/story/\_/id/14646545/ social-media-becomes-powerful-aide-dangerous-connectionrecruiting (accessed January 1, 2018).
- Caldwell D. 2015. Missteps on social media can ruin recruitment. Athletic Business. Available online at www.athleticbusiness.com/web-social/ missteps-on-social-media-can-ruin-recruitment.html (accessed January 1, 2016).
- National Collegiate Athletic Association. 2018. 2017 football attendance. National Collegiate Athletic Association. Available online at www.ncaa.org/championships/statistics/ncaa-football-attendance (accessed June 1, 2018).
- Pope DG, Pope JC. The impact of college sports success on the quantity and quality of student applications. South Econ J. 2009;75: 750–780.
- Goidel RK, Hamilton JM. Strengthening higher education through gridiron success? Public perceptions of the impact of National Football Championships on academic quality. Soc Sci Q. 2006;87: 851–862.
- Baade RA, Sundberg JO. Fourth down and gold to go? Assessing the link between athletics and alumni giving. Soc Sci Q. 1996;77:789–803.
- Bergman SA, Logan TD. The effect of recruit quality on college football team performance. J Sport Econom. 2014. DOI:10.1177/ 1527002514538266
- Hunsberger PK, Gitter SR. What is a blue chip recruit worth? Estimating the marginal revenue produce of college football quarterbacks. J Sport Econom. 2015;16:664–690.
- Sherman M. Balancing the recruiting budget. ESPN 2012. Available online at http://espn.go.com/college-sports/recruiting/football/story/\_/id/ 8041461/the-cost-recruiting (accessed January 1, 2018).
- National Collegiate Athletic Association. Estimated probability of competing in college football. National Collegiate Athletic Association. 2015. Available online at www.ncaa.org/about/resources/research/ football (accessed June 1, 2016).
- Safian N. Cornerstone Reputation college athletics survey report. Somerville, MA: Cornerstone Reputation. 2016.
- 12. National Collegiate Athletic Association. 2015–2016 NCAA Division I Manual. Indianapolis: National Collegiate Athletic Association. 2016.
- Sander L. For college athletes, recruiting is a fair (but flawed) game. Chron High Educ 2008;55:A1.
- 14. Goffman E. The presentation of self in everyday life. New York: Anchor Books 1959.
- Kacmar KM, Delery JE, Ferris GR. Differential effectiveness of applicant impression management tactics on employment interview decisions. J Appl Soc Psychol. 1992;22:1250–1272.
- Tsai W, Chen C, Chiu S. Exploring boundaries of the effects of applicant impression management tactics in job interviews. J Manage 2005;31: 108–125.
- Jones EE, Pittman TS. Toward a general theory of self-presentation. Suls J. (Ed.): Psychological Perspectives on the Self, Vol. 1. Hillsdale, NJ: Lawrence Erlbaum Associates 1982. pp. 231–262.
- Stevens CK, Kristof AL. Making the right impression: A field study of applicant impression management during job interviews. J Appl Psychol. 1995;80:587–606.
- HRreview. 60% of job seekers do not like recruiters searching social media profiles. HRreview 2013. Available online at www.hrreview.co.uk/hrnews/recruitment/60-of-job-seekers-do-not-like-recruiters-searchingsocial-media-profiles/49324 (accessed January 1, 2018).

- Penttila C. Build a social media hiring strategy. Entrepreneur 2009. Available online at www.entrepreneur.com/article/202466 (accessed January 1, 2018).
- Schawbel D. How recruiters use social networks to make hiring decisions now. TIME 2012. Available online at http://business.time.com/2012/07/ 09/how-recruiters-use-social-networks-to-make-hiring-decisions-now (accessed January 1, 2018).
- 22. Caers R, Castelyns V. LinkedIn and Facebook in Belgium: The influences and biases of social network sites in recruitment and selection procedures. Soc Sci Comput Rev. 2010;29:437–448.
- Broughton A, Foley B, Ledermaier S, et al. The use of social media in the recruitment process. Brighton, United Kingdom: Institute for Employment Studies 2013.
- Krider J, O'Leonard K, Erickson R. What Works brief: Talent acquisition factbook. Oakland, CA: Bersin by Deloitte 2015. Available online at https://legacy.bersin.com/uploadedfiles/042315-ta-factbook-wwbfinal.pdf (accessed January 1, 2016).
- Roth PL, Bobko P, Van Iddekinge CH, et al. Social media in employeeselection-related decisions: A research agenda for uncharted territory. J Manage. 2016;42:269–298.
- Bohnert D, Ross WH. The influence of social networking Web sites on the evaluation of job candidates. Cyberpsychol Behav Soc Netw. 2010;13: 341–347.
- Lampe C, Ellison NB. Student athletes on Facebook. 2010 ACM Conference on computer supported cooperative work. New York: Association for Computing Machinery 2010. pp. 193–196.
- Browning B, Sanderson J. The positives and negatives of Twitter: Exploring how student-athletes use Twitter and respond to critical tweets. Int J Sport Comm. 2012;5:503–521.
- Pegoraro A. Look who's talking—Athletes on Twitter: A case study. Int J Sport Comm. 2010;3:501–514.
- Sanderson J. Stepping into the (social media) scene: Building athlete identity via Twitter. In: Luppicini L (Ed.) Handbook on Research on Technoself: Identity in a Technological Society, Vol. 2. Hershey, PA: IGI Global, 2013. pp. 419–438.
- Lebel K. Professional athlete self-presentation on Twitter (Dissertation). University of Western Ontario, London, Ontario, Canada. 2013.
- Narchum JA, Havard CT, Mason KH. The impact of acts of Twitter transgression on an athlete's brand. J Bus Admin. 2016;14. Available online at https://www.atu.edu/jbao/spring2016/JBAOMason.pdf (accessed January 1, 2018).
- Jin SA, Phua J. Following celebrities' tweets about brands: Impact of Twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. J Advert. 2014;43:181–195.
- Navigate Research. What we know about measurement. Navigate Res. 2014. Available online at www.navigateresearch.com/ana (accessed January 1, 2018).
- Chung KYC, Derdenger TP, Srinivasan K. Economic value of celebrity endorsements: Tiger Woods' impact on sales of Nike golf balls. Mark Sci. 2013;32:271–293.
- Pegoraro A, Jinnah N. Tweet 'em and reap 'em: The impact of professional athletes' use of Twitter on current and potential sponsorship opportunities. J Brand Strat. 2012;1:85–97.
- Bigsby KG, Ohlmann JO, Zhao K. Online and off the field: Predicting school choice in college football recruiting from social media data. Decis Anal. 2017;14:261–273.
- Bigsby KG, Ohlmann JO, Zhao K. The turf is always greener: Predicting decommitments in college football recruiting using Twitter data. Decis Support Syst. 2018;116:1–12.
- Pitts JD, Rezek JP. Athletic scholarships in intercollegiate football. J Sport Econom. 2012;13:515–535.
- CBS Sports. FB Recruiting. 247Sports n.d. Available online at http:// 247sports.com (accessed January 1, 2018).
- Muthukadan B. Selenium with Python 2011. Available online at http:// selenium-python.readthedocs.io. (accessed January 1, 2016).
- Twitter. REST APIs. Twitter n.d. Available online at https://dev.twitter.com/ rest/public (accessed June 1, 2016).
- 43. Roesslein J. Tweepy 2009. Available online at www.tweepy.org (accessed June 1, 2016).
- 44. Wood-Doughty Z, Smith M, Broniatowski DA, Dredze M. How does Twitter user behavior vary across demographic groups? Proceedings of the

Second Workshop on Natural Language Processing and Computational Science. Stroudsburg, PA: Association for Computational Linguistics 2017. pp. 83–89. (accessed June 1, 2016).

- 45. Cohen JA. coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
- Loper E, Bird S. NLTK: The natural language toolkit. Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics, Vol. 1. Stroudsburg, PA: Association for Computational Linguistics 2002. pp. 63–70.
- 47. Manning CD, Raghavan P, Schutze H. Introduction to information retrieval. Cambridge, United Kingdom: Cambridge University Press 2008.
- Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2008. Available online at www.R-project.org (accessed January 1, 2016).
- Tukey JW. Comparing individual means in the analysis of variance. Biometrics. 1949;5:99–114.
- McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P (Ed.): Frontiers in Econometrics. Berkeley, CA: Academic Press, 1974. pp. 105–142.
- 52. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica. 1989;57:307–333.
- USA Today High School Sports. NCAA opens door for unlimited texting between recruits, football coaches. USA Today 2016. Available online at

https://usatodayhss.com/2016/ncaa-opens-door-for-unlimited-textingbetween-recruits-football-coaches (accessed January 1, 2018).

- Hostick MB. DI council adopts new football recruiting model. National Collegiate Athletic Association 2017. Available online at www.ncaa.org/ about/resources/media-center/news/di-council-adopts-new-footballrecruiting-model (accessed January 1, 2018).
- 55. Lewis M. Moneyball: The art of winning an unfair game. New York: W.W. Norton & Company 2003.

**Cite this article as:** Bigsby KG, Ohlmann JW, Zhao K (2019) Keeping it 100: Social media and self-presentation in college football recruiting. *Big Data* 7:1, 3–20, DOI: 10.1089/big.2018.0094.

#### Abbreviations Used

- ANN = artificial neural networks
- $\mathsf{AUC} = \mathsf{area}$  under the receiver operating characteristic curve
- BMI = body mass index
- $\label{eq:FBS} \mathsf{FBS} = \mathsf{Football} \; \mathsf{Bowl} \; \mathsf{Subdivision}$
- HR = human resources
- MSE = mean squared error
- $\mathsf{NCAA} = \mathsf{National} \ \mathsf{Collegiate} \ \mathsf{Athletic} \ \mathsf{Association}$
- $\mathsf{SVM} = \mathsf{support} \ \mathsf{vector} \ \mathsf{machines}$