# Online and Off the Field: Predicting School Choice in College Football Recruiting from

# Social Media Data

Kristina Gavin Bigsby<sup>1\*</sup>, Jeffrey W. Ohlmann<sup>2</sup>, Kang Zhao<sup>3</sup>

<sup>1</sup>Interdisciplinary Graduate Program in Informatics, University of Iowa, Iowa City, IA, 52242, kristinagavin@uiowa.edu. \*Corresponding author.

<sup>2</sup>Department of Management Sciences, University of Iowa, Iowa City, IA, 52242, <u>jeffrey-ohlmann@uiowa.edu</u>.

<sup>3</sup>Department of Management Sciences, University of Iowa, Iowa City, IA, 52242, <u>kang-zhao@uiowa.edu</u>.

# Abstract

The proliferation of social media provides new opportunities to study how people make decisions. This research explores predictors of school choice decisions in American college football recruitment. We combine data about individual athletes' recruiting activities with social media data to predict which school the athlete will choose from among those that have offered him a scholarship. While previous works have modeled school choice as a rational decision process, we find that considering decision-making heuristics can improve school choice predictions, a result that may be useful for predicting decisions in other recruiting contexts. We also investigate social factors in school choice, and find that models incorporating social media features consistently outperform the baseline model with only recruiting features. In addition to better understanding the school choice decision, this work is intended to help coaches effectively allocate recruiting resources and inform social media strategies during recruitment. Therefore, we also explore how the actions taken by athletes on during recruitment can be interpreted as early signals of athletes' preferences. Our final model, combining these different groups of offline and online features, achieves an AUC of 0.720.

## **Keywords**

Social media; School choice; College football; Applications; Multi-objective decision-making; Bounded rationality; Heuristics; Signaling

# Citation:

Bigsby, K. G., Ohlmann, J. W., & Zhao, K. (2017). Online and Off the Field: Predicting School Choice in College Football Recruiting from Social Media Data. *Decision Analysis*, 14(4), 261–273. https://doi.org/10.1287/deca.2017.0353

### **1. Introduction**

Social media provides detailed data about individuals' behaviors, preferences, and online social networks, presenting new opportunities to study how people make decisions. In this study, we analyze the school choice decisions of American college football recruits in the class of 2016. Athletic recruitment presents an interesting context for study due to its high stakes—college football recruiting budgets range from hundreds of thousands to millions of dollars per year (Sherman 2012), a high level of public interest on this topic, and availability of relevant data.

Previous work predicting school choice has mainly used rational decision-making models, presuming that an athlete will select a school that maximizes the expected utility of attendance (e.g. Dumond et al. 2008). However, we contend that school choice decisions occur under significant constraints in terms of time, information, and cognitive resources, suggesting that a bounded rationality model incorporating simple heuristics may be more appropriate.

Furthermore, the decisions of individuals are often related to their social networks. Both anecdotal and empirical evidence indicates that parents (e.g. Croft 2008), high school coaches (e.g. Prunty 2014), and other players (e.g. Myerberg 2015) have an important role in school choice decisions. Social media offers an unprecedented opportunity to gather data on the social networks of individuals, and we investigate the predictive power of Twitter data, specifically the connections and content posted by college football recruits in the class of 2016.

To account for these multiple, conflicting factors, we combine data about the athlete's choice set (the schools that have offered a scholarship), the athlete's recruiting activities, and the athlete's social media profile. We seek to predict which school an athlete will commit to from among those that have offered him a scholarship. We construct several models using different

sets of features and compare their performance in order to measure the value added by social media data.

While we focus on college football recruitment in this study, we contend that our findings may be generalized to other recruiting decisions in other domains, such as human resources (HR), military, or academic settings. We review background information on school choice in college football recruiting and related research in the remainder of Section 1. Section 2 outlines the data and methods of this study, and Section 3 presents results. Section 4 contains a discussion of findings and a practical application of our model. We make final conclusions and suggestions for further research in Section 5.

# **1.1. Modeling the School Choice Decision**

College football recruiting can be categorized into two sequential decision-making stages. In the first stage, schools identify and evaluate potential recruits and decide whether to extend a scholarship offer. In the second, athletes select a school from among the scholarship offers and announce a commitment. We focus on this latter stage, identifying factors predicting school choice decisions. But how do we model the school choice decision?

We may begin by considering school choice as a multi-objective decision-making process. That is, school choice decisions are likely to be based on several, potentially conflicting objectives. Indeed, surveys of college athletes have identified economic benefits (e.g. Doyle and Gaeth 1998), geographic proximity (e.g. Barden et al. 2013, Lujan 2010), probability of achieving a professional career (e.g. Croft 2008, Treadway et al.), and educational quality (e.g. Popp et al. 2012) as significant factors in school choice decisions. However, some of these objectives may be abstract or difficult-to-quantify. For instance, an athlete may base his commitment decision in part on his intention to pursue a professional football career, but the relationship between school choice and obtaining this goal is uncertain. The athlete may elect to use a means objective, an intervening factor that is related to this fundamental objective, but is easier to measure (Huynh and Simon 2016). In the case of maximizing the likelihood of an NFL career, the athlete may estimate the benefits of attending a given school by looking at the school's previous draft success or post-season appearances.

Yet we contend that such a rational decision-making model does not capture the complexity of the school choice decision. The underlying assumptions of the rational decision-making model are: (1) that athletes possess sufficient time to make rational choices, (2) that athletes possess sufficient information to make rational choices, and (3) that athletes possess the cognitive ability and desire to make rational choices. Our work questions these assumptions and investigates the role of heuristics and social factors in the commitment decision-making process.

Contradicting the first assumption, commitment decisions are often made under time constraints. While athletes can announce a verbal commitment at any time, they begin signing Letters of Intent on National Signing Day—the first Wednesday in February. The Letter of Intent is essentially a contract between the athlete and school where the athlete agrees to attend for one year and the school agrees to provide financial support for one year, and National Signing Day acts as a *de facto* deadline for commitments. Because schools can only award a maximum of 25 scholarships to incoming freshmen (NCAA 2015b), athletes may feel pressure to commit quickly in order to secure a scholarship. Indeed, 15% of college athletes report being given less than one week to accept a scholarship offer (Sander 2008). The high costs of recruitment can also encourage quick commitments. For instance, while the NCAA allows unlimited unofficial visits during recruitment—46% of respondents to Sander's (2008) survey went on at least 3 unofficial visits—athletes' families bear the costs of such visits. The father of a class of 2016 quarterback

commit estimated spending \$40,000 on travel expenses for camps, combines, and unofficial visits (Elliott 2015).

Challenging the second assumption, athletes often make school choice decisions with limited information. Recruits evaluate prospective colleges by taking visits, communicating with coaches and current college athletes, and researching the team and institution. As noted above, unofficial visits present an economic hardship to athletes' families and it may be impossible to for an athlete to visit each school that is recruiting him. While athletes can also take official visits paid for by the recruiting school, they are only allowed a maximum of 5 (NCAA 2015b). Additionally, athletes in the class of 2016 could not take official visits until fall of their senior year, although this rule was recently relaxed to allow spring visits during junior year. Gathering information about college options via communication with coaches is also fraught with difficulties. NCAA recruiting regulations limit contact between college coaches and recruits, presenting a significant impediment to communication and information exchanges. The football recruiting calendars is divided into four alternating periods: quiet, contact, evaluation, and dead (NCAA 2015a). During "quiet" periods, athletes may make visits to colleges, but coaches may not visit athletes off-campus. Coaches are allowed to visit with athletes and their families during "contact" periods, while "evaluation" periods allow only visits to an athlete's high school. "Dead" periods are the most constrained, with only telephone and written communication allowed. In a study of 98 high school athletes, only 18.4% of those reporting an intention to play in college had actually spoken with a college coach (Lujan 2010).

Although the proliferation of online recruiting databases such as Rivals.com or 247Sports has increased the amount of information available for decision-making, previous work suggests that athletes are ill-informed of college options. In Lujan's (2010) survey, 65% of high school

4

athletes reporting an intention to play in college had spent little or no time researching colleges. Yet accurate information on the athletic job market is vital, as only a fraction of high school athletes will actually earn a spot on a college roster. For example, the NCAA estimates that 6.5% of high school football players will play in college at any level, with only 2.5% in Division I (NCAA 2015c). Of those who do advance to the college level, only 1.6% will be drafted into the NFL (NCAA 2015c). These figures are highly disproportionate compared to the expectations of high school athletes—51% reported definitely planning to participate at the college level, with an additional 35.7% replying "maybe" (Lujan 2010).

This process of evaluating and selecting a college is also complicated by the difficulty of measuring subjective characteristics of a school, e.g. athlete-school fit. Athletes can consult information about a school's observable traits such as academic rating, majors offered, and team playing record, which are referred to as *indices* in Spence's job market signaling model (1973). However, indices will not capture all of the information relevant to the school choice decision. In this case, Spence's (1973) model predicts that the schools will take an action, or *signal*, in order to communicate information to the athlete. Applying this theory to college football recruiting, we hypothesize that athletes and coaches will engage in signaling in order to convey their relative levels of "interest" in the other party. For colleges, interest relates to the priority placed on a specific athlete relative to others, which will impact the athlete's expected utility of attendance at that school via intervening outcomes like availability of scholarship funds and future playing time. Conversely, the school risks turning off the athlete by signaling that he is not a priority. A top recruit described the worst recruiting pitch he received from a college team as "when it told him it offered three other QBs on the same day" (Davenport 2015). We contend that athletes will consider signals of interest from the recruiting schools communicated via social media when

making commitment decisions. We expect that instances where coaches and current players at the school demonstrate interest in the athlete will be more likely to result in commitment.

Applying completely rational decision-making models to school choice also assumes that athletes possess the cognitive ability and desire to make rational choices. In addition to the time and information constraints caused by NCAA policies, the recruiting calendar requires athletes to divide their attention between high school academics, sports, social lives, and the college recruiting process. While coaches may begin sending materials to an athlete during his junior year of high school, the first "contact" period does not occur until November of the athlete's senior year (NCAA 2015a). Thus, the most intense periods of recruitment directly conflicts with the school year and high school football season. Furthermore, a large body of literature explores age-specific differences in psychology and decision-making (e.g. Steinberg and Cauffman 1996), suggesting that adolescents are likely to deviate from rational decision-making, relying instead on emotional and social factors. While, anecdotal and empirical evidence indicates that parents (e.g. Croft 2008), high school coaches (e.g. Prunty 2014), and other players (e.g. Myerberg 2015) have an important role in school choice decisions, only one previous predictive work has considered athlete's social networks. Mirabile & Witte's (2015) study identified family connections between athletes and schools, finding that having a family member who played or coached at a school increased the likelihood of commitment between 96% and 253%. Similarly, we expect that athletes with social ties to a given school will be more likely to select that school. However, our work uses automated processes and social media data to track connections between athletes and colleges.

Given these constraints on rational decision-making in school choice, our project also considers the role played by heuristics and biases. Heuristics are mental shortcuts utilized when

6

making judgments and decisions under conditions of bounded rationality, and biases are systematic tendencies to deviate from rational decision-making. Specifically, we explore how the availability heuristic might impact school choice. The availability heuristic holds that decisionmakers will select the most memorable option (Tversky and Kahnemann 1973). Availability may be influenced by several factors, including sequence, frequency, and vividness. Because our data includes timelines of recruiting events, we focus on the relationship between school choice and sequence, tracking the first and last events in each series (e.g. offers, visits).

Ultimately, this project has two main objectives: (1) to better understand the school choice process of college football recruits, and (2) to build predictive models that can assist coaches in identifying athletes who are most likely to commit to their school, which can inform their recruiting strategies. With this second goal in mind, we also consider the how athletes' actions during recruitment can be interpreted as early signals of interest in a specific school. For example, schools may make offers to other athletes at the same position, creating direct competition for scholarships and future playing time. Additionally, offers may be rescinded—a rare but not unheard-of event. An athlete may consciously elect to communicate his interest to a college team (e.g. attending a college camp or taking an unofficial visit) in an effort to avoid these negative outcomes. In addition to such offline channels for engaging in signaling, we contend that athletes will also use online social media as a medium to communicate their interest in a school, whether by connecting with coaches and athletes at the school or posting content about the school. Spence (1973) describes signals as conscious actions meant to communicate information, but we also consider the possibility that social media signals are unconscious. The HR literature has investigated how organizations use candidates' social media profiles to screen for traits like "character" (Broughton et al. 2013) and "professionalism" (Schwabel 2012). We

focus on the social media behaviors most likely to be interpreted as signals of interest, including following, replying, retweeting, and tweet content. Social media offers a fertile environment to study social factors in decision-making because it enables unobtrusive observation of public connections and communication, information that would be difficult to obtain offline.

#### **1.2. Related Research**

Our study is the first of its kind to examine social media predictors of school choice in college football. In addition to contributing to the growing body of sports analytics literature, our work has more general implications about the role of social factors and heuristics in recruiting decisions, and we consider previous work from the athletic recruiting, HR recruiting, and decision analysis domains.

This project uses college football recruiting as a case study to explore the use of social media data for analyzing and predicting recruiting decisions in other contexts. We are informed by previous studies of college athletic recruiting identifying significant factors in school choice decisions (e.g. Croft 2008, Doyle and Gaeth 1998, Popp et al. 2012). These works present multiple, conflicting objectives motivating school choice, and we also consult previous research on multi-objective decision analysis in personnel management (e.g. Dees et al. 2013). However, we focus on inferring the hierarchy of the objectives via logistic regression, rather than eliciting preferences from decision-makers. Since our models use available data about an athlete's college options to predict which school he will choose, we do not have access to information about the decision-maker's fundamental objectives, so we also draw upon prior work on means objectives in multi-objective decision analysis (e.g. Huynh and Simon 2016).

Given constraints on time, information, and cognitive resources during recruitment, we explore school choice from the perspective of bounded rationality. The extant literature on school choice and job choice has primarily focused on applying rational decision models, overlooking heuristics and biases (Highhouse et al. 2014). Hogarth and Karelaia (2006) have previously studied the utility of heuristics for predicting decisions in situations of bounded rationality, though we focus on the impact of availability (Tversky and Kahnemann 1973) on school choice decisions rather than the elimination-by-aspects and take-the-best heuristics applied in their work. We also investigate the role of social factors in commitment decisionmaking. With the exception of Mirabile and Witte (2015), previous work on athletic commitments has overlooked social network predictors (e.g. Dumond et al. 2008), and our project is the first to include social media data in a predictive model of school choice. The HR research literature has paid a significant amount of attention to social networks in job-seeking and job choice (e.g. Granovetter 1973). Indeed, social ties between a candidate and organization can increase applicant attraction and job choice intentions by providing unique and credible information (McManus and Baratta 1992, Vecchio 1995) and enhancing perceived personorganization fit (Chapman et al. 2005). We interpret social network behaviors by coaches and athletes during recruitment using Spence's (1973) signaling theory, and are informed by previous work studying the information asymmetries in football recruiting (Bricker and Hanson 2012) and the intersection between signaling and decision analysis (Cobb and Basuchoudhary 2009).

### 2. Methods

This research analyzes the decision-making of college football recruits, specifically investigating the role of objective, cost/benefit factors, heuristics, and signals of interest communicated by colleges to athletes via social media. Additionally, we seek to produce predictions of school choice that may assist coaches in crafting recruiting strategies. We explore how athletes' connections and behaviors on social media can provide timely information about their college preferences. We measure the value added by social media features to school choice predictions by developing multiple predictive models and comparing their performance. We describe the dataset and methods in the remainder of Section 2.

2.1. Data

We scraped data on 2,644 high school football athletes in the 2016 recruiting class from the recruiting database of 247Sports.com (247Sports). For each individual athlete, we collected personal information, e.g., height, weight, hometown, position, as well as recruiting information, e.g. scholarship offers, visits, commitments, and decommitments. We also obtained basic information about the recruiting schools, including location, academic ranking, and football team ranking. Many 247Sports profiles contained embedded Twitter timelines; 1,629 Twitter IDs for recruits in the class of 2016, 466 IDs for Division I coaches, and 2,225 IDs for currently college athletes were retrieved from the site. We conducted a manual search of the remaining athletes in the class of 2016, locating 700 additional Twitter IDs. In full, 2,329 recruits in the dataset (88%) possess public Twitter accounts. Twitter information for these individuals was collected using the Twitter REST API (Twitter). Profile information, friend and follower lists, and tweets were gathered monthly between September 2015 and March 2016—the most active period of recruitment for the class of 2016. Social media data was collected over time in order to observe the changes in athletes' online social networks prior to commitment.

For the purposes of this project we consider both verbal commitments and signings. We identify commitments from athletes' recruiting data in two ways: (1) commitment announcements tracked in a player's 247Sports timeline, and (2) NLI signings. 2,277 unique athletes (86%) committed at some point during recruitment. This study seeks to predict which school an athlete will commit to among those that have offered scholarships, so we exclude

10

athletes without a commitment from our data. Although an athlete can verbally commit at any time during recruitment, we consider only commitments occurring between October 2015 and February 2016, approximately 43% of commitments for the class of 2016. This time range is selected so that at least one month of retrospective social media data is available for each player. There were 25 commitments after March 1, 2016 (1% of all commitments) that we do not include in this study. Late commitments are fairly uncommon, and are most likely to occur in instances where academic eligibility or oversigning—when a team signs more than 25 NLIs and has to revoke scholarship offers—are an issue. Thus, the school choice decision of late commits is likely to be significantly different than athletes who commit before National Signing Day. We eliminate 705 athletes without multiple offers from our data; 93% of athletes who received only one offer committed to that school.

Because our models consider athletes' social media data, we also exclude athletes without public social media profiles. We contend that this does not introduce bias into our models because there is no evidence of significant differences between athletes with and without social media. Chi-squared tests fail to reject the null hypothesis of independence between possession of a public Twitter account and star rating (p = 0.9912) as well as the number of offers received (p = 0.394).

These steps yield a dataset with 573 athletes who selected a school from among 8 scholarship offers on average. For each athlete in the data, we create an instance for every offering school, ultimately resulting in 4,408 "athlete-school" pairs. Therefore, our prediction dataset has multiple instances corresponding to the same athlete. For each athlete-school pair in our data, we measure features relative to the "prediction school" and model the likelihood of the athlete selecting that school.

### 2.2. Feature engineering

The performance of a predictive model depends on the features it considers. To determine the value added by considering athletes' social media data, we compare one group of features constructed from recruiting data with four groups of social media features corresponding to different aspects of athletes' social media profiles: in-links, out-links, interactions with other (mentions, replies, retweets, quotes), and tweet content.

## **2.2.1. Offline features**

We construct a set of "offline" features from the 247Sports recruiting data and school data. We draw upon the work of Dumond et al. (2008), who identified economic capital, athletic capital, and human capital objectives in school choice decisions. As these fundamental objectives may be difficult to express and measure, we construct features relating to means objectives. For instance, a school's *U.S. News & World Report* ranking may impact the fundamental objective of increasing human capital through quality of education. We expect that features that increase the benefits associated with attendance at a given school will also increase likelihood of commitment. We expand on previous work by considering these objectives in relation to alternative options in the athlete's choice set. For example, an athlete's likelihood of selecting the prediction school may be influenced not only by that school's geographic proximity, but also by the number of other schools recruiting him that are closer. We include data about "offline" recruiting activities that demonstrate affinity between a college and athlete, e.g. offers and visits. In light of the constraints on time, information, and cognitive resources faced by athletes, we also consider the role of heuristics in school choice and track the sequence of these recruiting events.

In total, we construct and test 26 offline features (Table 1). We use time-consistent data, meaning that we exclude events that occurred after the commitment decision. For example, to

12

predict which school an athlete will commit to in January, we count only official visits that occurred before January 1. We assume that the month of commitment is known, as our focus is predicting *where* and athlete will commit, rather than *when*.

# **Table 1 Offline features**

Category	Feature	Description				
Objectives	prediction_distance	Numeric; distance between recruit hometown and school being predicted				
	prediction_inState	Binary; recruit hometown in same state as prediction school				
	prediction_usNews	Binary; prediction school is included in 2015 U.S. News academic rankings (published September 2014)				
	prediction_AP	Binary; prediction school ranked in top 25 of AP poll (end of 2014-2015 season)				
	prediction_division	Categorical; athletic division of prediction school				
	prediction_conference	Categorical; athletic conference of prediction school				
	prediction_sanction	Binary; football program under active NCAA sanction or probation				
	prediction_coachChange	Binary; head coach change at prediction school during 2015-2016 season				
	other_offer	Numeric; number of offers from other schools				
	other_closer	Numeric; number of other schools closer to recruit hometown				
	other_higherAP	Numeric; number of other schools with higher AP ranking				
	other_higherUsNews	Numeric; number of other schools with higher U.S. News academic ranking				
Affinity	prediction_unofficial	Binary; unofficial visit to prediction school				
	prediction_coachVisit	Binary; coach visit from prediction school				
	prediction_official	Binary; official visit to prediction school				
	other_unofficial	Numeric; number of unofficial visits to other schools				
	other_coachVisit	Numeric; number of coach visits from other schools				
	other_official	Numeric; number of official visits to other schools				
Heuristics	prediction_firstOffer	Binary; first offer from prediction school				
	prediction_lastOffer	Binary; last offer from prediction school				
	prediction_firstUnofficial	Binary; first unofficial visit to prediction school				
	prediction_lastUnofficial	Binary; last unofficial visit to prediction school				
	prediction_firstCoach	Binary; first coach visit from prediction school				
	prediction_lastCoach	Binary; last coach visit from prediction school				
	prediction_firstOfficial	Binary; first official visit to prediction school				
	prediction_lastOfficial	Binary; last official visit to prediction school				

# **2.2.2.** Followers of athletes

The behavior and decisions of individuals are often related to their social network neighbors, and the next group of features focuses on in-links from other Twitter users, i.e. "followers." We treat these features as signals of interest from the school to the athlete and expect that the likelihood of commitment to a given school will increase as the number of followers from that school increases. We determine the number of new followers in the month prior to commitment by comparing the Twitter IDs of an athlete's followers on the first days of the commitment month and the previous month and calculating the difference. For example, if predicting where an athlete will commit in January, the followers lists retrieved January 1 and December 1 will be compared to determine the number of new followers during December. We also track the type (recruits, current college athletes, and coaches) and school affiliation of these followers. We construct 6 features based upon athletes' in-links (Table 2).

Category	Feature	Description				
In-links	coach_followers_prediction	Numeric; coaches from prediction school following user over prior month				
	2016_followers_prediction	Numeric; 2016 recruits committed to prediction school following user over prior month				
	current_followers_prediction	Numeric; current athletes at prediction school following user over prior month				
	coach_followers_other	Numeric; coaches from other schools following user over prior month				
	2016_followers_other	Numeric; 2016 recruits committed to other schools following user over prior month				
	current_followers_other	Numeric; current athletes at other schools following user over prior month				
Out-links	coach_friends_prediction	Numeric; coaches from prediction school followed by user over prior month				
	2016_friends_prediction	Numeric; 2016 recruits committed to prediction school followed by user over prior month				
	current_friends_prediction	Numeric; current athletes at prediction school followed by user over prior month				
	coach_friends_other	Numeric; coaches from other schools followed by user over prior month				
	2016_friends_other	Numeric; 2016 recruits committed to other schools followed by user over prior month				
	current_friends_other	Numeric; current athletes at other schools followed by user over prior month				
Interactions	interaction_prediction	Binary; athlete has posted a retweet, reply, quote, or mention of users associated with prediction school over prior month				
	interaction_other	Binary; athlete has posted a retweet, reply, quote, or mention of users associated with other schools over prior month				
Content	hashtag_prediction	Binary; athlete has posted a hashtag associated with prediction school over prior month				
	hashtag_other	Binary; athlete has posted a hashtag associated with other schools over prior month				

 Table 2 Social media features

## **2.2.3.** Users followed by athletes

In addition to analyzing factors impacting the school choice decision, we consider how athletes' online behaviors may be interpreted as early signals of their school preferences. The second set of features track users followed by the athlete on social media during the month prior to commitment, i.e. "friends" or out-links. We expect that athletes intending to commit to a certain school will add friends from that school. This interpretation is supported by social network theory; network realignment holds that overlap in members' respective social networks will increase with the intensity of a dyadic relationship (Jowett and Timson-Katchis 2005). Indeed, 62% of Division I athletes report building friendships with their future teammates during recruitment (Sander 2008). Similar to the features tracking in-links, friends lists retrieved on the first days of the commitment month and previous month are compared to determine the number and affiliation of new friends in the month prior to commitment. We construct 6 features related to Twitter out-links (Table 2).

#### 2.2.4. Social network interactions

The next group of features examines the actions of athletes on social media. Twitter allows users to interact in several ways: replying to other users' posts, copying posts, forwarding posts, and mentioning other users. We interpret these actions as signals of an athlete's preferences and hypothesize that greater numbers of interactions with recruits, coaches, and current athletes associated with the prediction school will be associated with increased likelihood of commitment. As NCAA policies prohibited college coaches and athletic department staff from mentioning, quoting, retweeting, or replying to high school athletes during the period of data collection (Elliot and Kirshner 2016), we only track the social network interactions initiated by athletes. We retrieve the replies, quotes, retweets, and mentions posted on each athlete's Twitter timeline during the month prior to commitment. Because some types of interactions are more common than others, we create an aggregate measure. For each athlete, we track the affiliation of users interacted with and combine these counts into an "interactions" feature (Table 2).

## 2.2.5. Hashtag Content

Social media also offers a rich source of text data from users' posts. Thus we investigate whether the content posted by an athlete on social media is predictive of school choice. We analyze the hashtags posted by athletes in the month prior to commitment, and expect that posting hashtags relevant to a school will be associated with increased likelihood of commitment. As free text data, the topic of a hashtag is not always evident. We use a two-step information retrieval process to determine the likely topic of each hashtag:

- (1) For each of the 682 schools in our data we generate a set of positive query terms (terms likely to retrieve documents relevant to the school), denoted *T*. These query terms are substrings constructed manually based on the school name, team name, nickname, abbreviation, coach name, and/or location of each school. For instance, the query terms for the University of Utah are  $T=\{\text{'utah', 'utes', 'utenati'}\}$ . Each athlete's hashtags are treated as a set of documents *D*. Initial queries are constructed using the Boolean OR operator and a hashtag is considered to be potentially relevant to a school and included in subset *S1* if it contains at least one positive term.
- (2) We also construct a list of negative terms NT for each school, or substrings that should be disallowed in relevant hashtags. For the University of Utah, NT={'utahst'}, effectively excluding references to Utah State. S1 is then queried using the NOT operator. Of all potentially relevant documents retrieved by the first query, we include in the final set S2 only those that do not match any of the negative terms. We then create two features

tracking the number of hashtags relevant to the prediction school and relevant to other schools that have offered scholarships to the athlete (Table 2).

#### 2.3. Models and evaluation

As opposed to consulting decision-makers to estimate the value placed on different objectives, we infer the feature weights through logistic regression. We select logistic regression for this study because of its interpretability and performance with non-normally distributed response variables. We divide the dataset into 3,072 training instances (409 commitments) and a hold-out set of 1,336 testing instances (179 commitments). As our data contains multiple instances corresponding each individual athlete, we keep such instances together to avoid training and testing on the same athlete.

To evaluate the contributions of different groups of features, we implement the following 6 models with different combinations of feature groups. Model 0 (the baseline model) uses offline, recruiting features only. Because the features we derive may be highly correlated to each other, we perform feature selection using a Lasso regression with L1 penalty (C=0.1) to construct the baseline model. We remove predictors whose weight reduces to 0 as well manually eliminating non-significant predictors. For consistent comparison to the other models, we re-fit the baseline using logistic regression without regularized maximum likelihood or penalty. Model 1 adds to the baseline the features related to an athlete's in-links, i.e., the Twitter users that followed the athlete in the month prior to commitment. Model 2 focuses on the "friends" in an athlete's online social network, adding features tracking the number and affiliation of out-links to the baseline. Model 3 adds the features tracking social media interactions to the baseline model, and Model 4 adds the features derived from hashtag content to the baseline. Model 5 combines all features from Models 0-4. As the issue of collinearity arises again when combining all social

media features, we apply Lasso regression to construct Model 5. As with the baseline, we subsequently re-fit Model 5 using logistic regression without regularized maximum likelihood or penalty.

We evaluate the predictive performance of each model on the hold-out set. We consult standard classification metrics: precision (ratio of true positives to predicted positive instances), recall (ratio of true positives to actual positive instances), balanced F score (harmonic mean of precision and recall), and area under the receiver operating characteristic curve (AUC), which measures the probability of ranking a randomly chosen positive instance higher than a randomly chosen negative instance. Precision, recall, and F score are measured in regards to performance on the positive class. Because only 17% of instances in our data correspond to commitments, we do not use overall accuracy (proportion of correctly predicted instances), which is not robust to class imbalance.

Logistic regression yields a predicted probability of commitment for each athlete-school pair, marking each instance where the predicted probability is greater 50% as a commit. Thus, it is possible that our models will predict more than one commitment per athlete. To account for this, we rank each athlete's college options according to the predicted probability and assess the school choice prediction as a ranking. We use normalized discounted cumulative gain (NDCG), which measures the quality of a ranking based on relevance and position in the results list. Because the number of predictions for each athlete in our data varies based on the number of scholarship offers received (but each athlete has at least two offers), we calculate NDCG using only the top two predictions for each athlete.

# 3. Results

This section illustrates and discusses the results of the six different logistic regression models and compares their predictive performance on the hold-out training data.

# 3.1. Factors Related to School Choice

Logistic regression is commonly used in situations where the response variable Y is binary. In our case, the outcome we wish to predict is whether an athlete will commit to a specific school, and for each athlete-school pair, our models calculate the predicted probability that Y = 1. Different than linear regression, where the coefficients of each independent variable Xcan be expressed as the rate of change in Y as X changes, model coefficients in logistic regression measure the rate of change in the log odds. Thus, by applying the exponential function to the coefficients, we can quantify how each variable impacts the likelihood of commitment. Model results are contained in Table 3.

	N. 110		N 111		N 110		NC 112		N 114		Nr. 1.1.7	
	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5	
constant	-0.9904	***	-1.0086	***	-1.0408	***	-0.8405	***	-0.9475	***	-0.9665	***
prediction_inState	0.7478	***	0.7680	***	0.7758	***	0.7126	***	0.7053	***	0.7466	***
other_closer	-0.0607	*	-0.0455		-0.0443	n.s.	-0.0534	*	-0.0408	n.s.	-0.0254	n.s.
other_higherUsNews	-0.3229	*	-0.2805		-0.2362	n.s.	-0.3316	*	-0.2432	n.s.	-0.1471	n.s.
other_offer	-0.1173	***	-0.1248	***	-0.1257	***	-0.1190	***	-0.1100	***	-0.1212	***
prediction_official	2.6427	***	2.3059	***	2.3423	***	2.5211	***	2.2794	***	2.0916	***
other_unofficial	-0.0407		-0.0447		-0.0442		0.0400		-0.0326	n.s.	-0.0405	n.s.
other_official	-0.3198	***	-0.3386	***	-0.3566	***	-0.3470	***	-0.3251	***	-0.3257	***
prediction_firstOffer	0.6084	***	0.5631	***	0.5607	***	0.6250	***	0.5975	***	0.5578	**
prediction_lastOffer	-0.7238	***	-0.7458	***	-0.7455	***	-0.7337	***	-0.8779	***	-0.8868	***
prediction_lastCoach	0.4997		0.4602	n.s.	0.3409	n.s.	0.4305	n.s.	0.3664	n.s.	0.2985	n.s.
prediction_lastOfficial	1.3497	***	1.4265	***	1.4181	***	1.3205	***	1.5529	***	1.5433	**
coach_followers_prediction			0.3365									
2016_followers_prediction			0.4088	***								
current_followers_prediction			0.3782								0.2411	n.s.
coach_followers_other			-0.0509	n.s.							-0.0355	n.s.
2016_followers_other			-0.0131	n.s.								
current_followers_other			-0.2210	*							-0.1872	
coach_friends_prediction					0.3868	***					0.3258	
2016_friends_prediction					0.4855						0.3987	***
current_friends_prediction					0.1682	n.s.						
coach_friends_other					-0.0070	n.s						
2016 friends other					-0.0270	*					-0.0193	n.s.
current_friends_other					-0.0652	n.s.						
interaction_prediction							0.6156	***				
interaction_other							-0.2173	n.s.				
hashtag_prediction									1.3986	***	1.2342	***
hashtag_other									-1.0366	***	-1.0106	***

#### **Table 3 Logistic regression results**

## \*\*\* : p < 0.001, \*\* : p < 0.01, \* : p < 0.05, . : p < 0.1, n.s. : not significant

The features in the baseline model (Model 0) relate to cost/benefit factors influencing school choice, comparisons to alternatives, athlete-school affinity demonstrated by recruiting activities, and decision-making heuristics. Applying Lasso regression reduces the size of the baseline model from 26 to 11 features. Intuitively, we find that features that decrease costs of attendance for athletes are associated with increased likelihood of commitment. Athletes are approximately 111% more likely to commit to a college in their home state (prediction\_inState). Attendance at an in-state school is linked not only to decreased travel costs, but also an increased sense of satisfaction and fit (Barden et al. 2013). Considering alternative options in the athlete's choice set, likelihood of commitment to the prediction school decreases 11% for each offer from another school. The relative costs of these alternatives also factor into the commitment decision. An athlete's likelihood of choosing the prediction school decreases 6% for each offering school that is geographically closer (other\_closer) and 28% for each school that has a higher academic ranking (other\_higherUsNews). We also find that offline recruiting activities are strong predictors of school choice. Completing an official visit to the prediction school (prediction\_official) increases likelihood of commitment by over 1300%, while each official visit to another school decreases likelihood of commitment by 27%, with a 4% decrease for unofficial visit to another school (other\_unofficial).

The sequence variables support the use of the availability heuristic in athletes' commitment decision-making, suggesting that recency and vividness may impact athletes' evaluation of schools. Athletes are 286% more likely to select a school when it was their most recent official visit, and likelihood of commitment increases 65% when the last coach to visit the athlete was from the prediction school. Athletes are 84% more likely to commit to a school when it was the first to offer them a scholarship, and 52% less likely to select the last school to offer

them. Anecdotal evidence suggests that athletes may attribute more emotional weight to the first offer. For example, Shaquille Quarterman, a 2016 linebacker commit, stated on Twitter, "First coach to ever offer me face to face was Richt, told me he didn't offer 16 year olds but knew dawgs when he saw them. We canes now."

Model 1 adds features corresponding to the signals of interest communicated by colleges to recruits, specifically the number of Twitter users associated from the prediction school and other schools following the athlete in the month prior to commitment. As hypothesized, an increase in social media followers from the prediction school is associated with greater likelihood of commitment. Each coach from the prediction school following an athlete increases likelihood of commitment by 40% (coach\_followers\_prediction), each recruit from the class of 2016 increases likelihood of commitment by 51% (2016\_followers\_prediction), and each current athlete at the prediction school following the athlete in the prior month increases likelihood selecting that school by 46% (current\_followers\_prediction). As recruitment occurs in a competitive environment, we consider signals of interest from other schools recruiting the athletes. We find that each current college athlete attending another school following the recruit in the previous month decreases likelihood of selecting the prediction school by 20% (current\_followers\_other). These results suggest that the signals of interest communicated by schools on social media are related to school choice.

In Model 2, we consider how an athlete's online social network out-links can provide insight into his college preferences. We find that following accounts associated with the prediction school in the previous month increases the likelihood of commitment—47% for each coach (coach\_friends\_prediction) and 62% for each committed athlete from the class of 2016 (2016\_friends\_prediction). Conversely, following accounts associated with schools other than

21

the prediction school in the previous month is associated with decreased likelihood of commitment. An athlete's likelihood of commitment to the prediction school decreases 3% for each new friend from the class of 2016 committed to another school (current\_friends\_other). These findings suggest that social media connections can be interpreted as indications of an athlete's school choice intentions. In Model 3, we use Twitter "interactions" (mentions, replies, quotes, and retweets posted by the recruit) as predictors of school choice. Interacting with the prediction school is associated with an 85% increase in likelihood of selecting that school (interaction\_prediction). This result suggests that athletes will invest effort into building relationships with their preferred schools, and that this information can be used when predicting school choice.

Model 4 investigates the use of text data in predicting school choice, specifically the hashtags used by an athlete in the month prior to commitment. Making a reference to the prediction school via hashtag increases likelihood of commitment by 305% (hashtag\_prediction), and making a reference to a competing school decreases likelihood of commitment by 65% (hashtag\_other). These results suggest that athletes may communicate their preference for a given school via content posted on social media.

Model 5 adds to the offline features of the baseline model all of the social media features tested in Models 1-4. We apply Lasso regression again to correct for potential collinearity issues. This process results in a final model with 19 features, 11 offline recruiting features and 8 social media features. Each current college player from the prediction school following the athlete is associated with a 27% increase in likelihood of commitment, while each player from another school in the athlete's choice set decreases likelihood of commitment by 17%. While not causal, these results suggest that colleges may be able to leverage peer recruiting to influence athletes'

school choice decisions. We see that athletes' out-links are strong predictors of their preferences, with likelihood of commitment increasing 39% and 49% for each coach and recruit from the prediction school, respectively. While interactions were eliminated from the final model, hashtag content can be interpreted as communicating athletes' relative interest in recruiting schools. Posting a hashtag associated with the prediction school increases likelihood of commitment by 244%, and posting a hashtag associated with another offering school decreases likelihood of commitment by 64%. Overall, these results demonstrate the utility of considering actions and connections in both offline and offline environments when predicting school choice decisions.

# **3.2. Predictive Performance**

To evaluate the contribution of social media data to prediction, we apply the models constructed with training data to hold-out testing data. We compare the performance of the baseline model containing only offline recruiting data (Model 0) with models incorporating social media features (Models 1-4).

Evaluating each model based upon standard metrics for classifier performance (AUC, precision, recall, and F score), we see that models incorporating social media features consistently add value over the baseline model. For instance, Models 1-5 achieve 0-8% improvement in AUC, 0-3% improvement in precision, 0-30% improvement in recall, and 0-19% improvement in F score over the baseline model (Fig. 1).



Figure 1 Performance of Models 0-5 (AUC, precision, recall, F score)

These metrics suggest that of each set of social media features (Models 1-4), those related to network centrality (in-links and out-links) and content show the largest gains in performance over the baseline model with only recruiting data. Features related to social media interactions (mentions, replies, retweets, quotes) display the smallest gains over the baseline and both of the "interaction" features are eliminated from the combined model via Lasso regression. Additionally, the final model (Model 5) outperforms all other models tested and achieves an AUC of 0.720.

Similar results are achieved when evaluating the school choice predictions as a ranking problem. Figure 2 displays the NDCG@2 scores for each model averaged over all players in the testing set. Again, models considering signals communicated via the athlete's social media

consistently outperform the baseline, with Models 1-5 achieving gains of 0-13% over the baseline NDCG score.



#### Figure 2 Performance of Models 0-5 (NDCG@2)

Like the evaluation under standard classifier metrics, Models 1 and 2 show the largest gains in NDCG over baseline out of any the individual sets of social media features (Model 1-4). Model 5 achieves the highest average NDCG score (0.755), indicating that combining features related the estimated cost/benefit of attendance, comparisons to alternative options, athlete-school affinity, decision-making heuristics, , signals of interest communicated via social media, and early indications of athletes' preferences on social media can yield accurate school choice predictions. We can also measure differences in NDCG across the same athlete, and Model 5 also has the highest average gain in NDCG over the baseline model..

#### 4. Discussion

This work represents a novel addition to research examining the intersection of information exchanges on social media and individual decision-making. We extend previous research predicting school choice (Dumond et al. 2008, Mirabile and Witte 2015) by considering decision-making heuristics and social media data. Over all tests performed, the combined model with both recruiting features and social media features (Model 5) is the highest performer, with

7% improvement in AUC and 18% improvement in F score over the baseline model. This result suggests that a model of school choice incorporating information gleaned from offline and online behaviors would be most effective. Both Dumond et al. (2008) and Mirabile and Witte (2015) use predictive accuracy to evaluate their models, with the former achieving 71% accuracy in predicting the school choices of the 2005 Rivals top 100 recruits and the latter achieving 65% accuracy over 19,815 recruits in ten recruiting classes. While we elect to use metrics that are robust to class imbalance (e.g. AUC) and cannot directly compare our results to previous works, our findings suggest that it is worthwhile to include consideration of decision-making heuristics and social factors into a predictive model of school choice.

Additionally, we find that among the four types of social media features we explore, the models focusing on structural features (Models 1 and 2) contribute more to predictive performance than those based on social media interactions or hashtag content. These results suggest that connective behaviors on social media (e.g. "following") function as clearer signals of interest than other relationship-building behaviors. These findings can inform the recruiting strategies employed in both athletics and HR, with signals of interest directed from the organization to the candidate positively related to likelihood of commitment. We also find that athletes' behavior on social media is strongly predictive of school choice, and contend that recruiters in athletics and other domains should consider the preferences communicated by candidates on social media when deciding how to allocate recruiting resources.

We note some limitations of this work. First, the accuracy of the recruiting data in the study is dependent on its source. 247Sports incorporates both user-supplied and expert data, which we believe makes it both a comprehensive and up-to-date source of recruiting information. Second, the scope of this project is limited to athletes with public Twitter profiles, and predicting

school choice for athletes without a presence on social media may yield different results. However, initial data analysis suggests no significant differences between athletes with and without Twitter in terms of recruiting activities. Third, the project uses only one year of recruiting data (information on the class of 2016 obtained from 247Sports.com) and 6 months of social media data (Twitter information gathered from 1 September 2016 to 1 March 2016). Therefore, these results may not be generalizable to other recruiting classes. Finally, as this is not an experimental study, no causality can be inferred from our results. Our findings indicate that features derived from the social media profiles of American college football recruits may be useful for predicting which school an athlete will commit to out of those that have offered him a scholarship. While our models are consistent with existing theories, we cannot state that social media variables cause commitments.

### **4.1. Practical Application**

In addition to analyzing the decision-making of college football recruits, our work seeks to provide practical insights for coaches and recruiters to use when crafting recruiting strategies. Therefore, we produce a sample report that a school might consider when making decisions on how to allocate recruiting resources for the remainder of the recruiting cycle. In Table 4, we list the predicted probability of ten high school athletes that received offers from the University of Iowa, but remained uncommitted as of January 2016.

Name	Star	Position	Hometown	Iowa	<b>Top Prediction</b>		
Alaric Jackson	3	Offensive tackle	Detroit, MI	90%	Iowa		
Matt Farniok	4	Offensive tackle	Sioux Falls, SD	60%	Iowa		
Obi Obialo	3	Wide receiver	Coppell, TX	29%	Iowa		
Tyler Johnson	3	Quarterback	Minneapolis, MN	22%	Minnesota (48%)		
Kene Nwangwu	3	Running back	Frisco, TX	14%	Iowa State (39%)		
Tyquan Statham	3	Athlete	Oakwood, GA	8%	Cincinnati (63%)		
KJ Gray	3	Athlete	Jersey City, NJ	6%	Rutgers (95%)		
Izon Pulley	3	Defensive end	Olney, MD	5%	Illinois (36%)		
Terrance Landers, Jr.	3	Wide receiver	Dayton, OH	4%	Purdue (50%)		

 Table 4 Top 10 Iowa Prospects (January 2016)

Jerrion Nelson	3	Defensive end	Columbia, MO	3%	Syracuse (70%)

According to Model 5 (which combines both recruiting data and features from social media), Alaric Jackson had a 90% chance of selecting Iowa. He had made an official visit and received signals of interest from Iowa (2 coaches and 5 committed recruits followed him on Twitter). Additionally, he showed his preference for Iowa by following 2 coaches, 6 committed recruits, and posting 6 Iowa hashtags. Ranked by predicted probability of attendance, Iowa's next prospect was Matt Farniok. He had made an official visit and was followed by 2 coaches and 2 committed recruits from Iowa. Unlike Jackson, had did not follow any Iowa accounts and did not post any Iowa hashtags. A coach interpreting this prediction might safely consider Jackson a strong prospect. Given the fact that both recruits were offensive tackles, the coach might decide to focus his efforts on maintaining a relationship with Jackson during the final weeks of recruitment. However, the coach might also decide to take an action to try to increase the probability of Farniok choosing Iowa. In this case, our results suggest that having current athletes from Iowa follow Farniok on Twitter could increase his likelihood of commitment. Interestingly, Iowa football has a "no Twitter" policy for current athletes, missing out on the benefits of peer recruiting. Ultimately, Jackson did commit to Iowa, while Farniok selected Nebraska.

Our model, which produces a predicted probability of commitment for each school that has offered an athlete, may also be useful for coaches seeking to identify which prospects are most likely to select a competitor. For example, KJ Gray is predicted to have a 6% chance of choosing Iowa, compared to a 95% chance of committing to Rutgers. While Gray was followed by two Iowa recruits and followed them back, he was followed by three Rutgers recruits and two current athletes and followed them back, in addition to being in-state. A coach interpreting these results could realistically assume that expending additional time and effort recruiting Gray would be unlikely to pay off in a commitment. Indeed, Gray did commit to Rutgers.

Although this example describes the recruiting prospects of only one team, we believe that it shows the potential for our model to be applied in real-world settings and inform the strategies of college football coaches.

## **5.** Conclusion

In this research, we analyze the school choice decisions of college football recruits. Previous work on school choice has focused on applying rational decision-making models, and our results support the importance of cost/benefit factors, comparisons to other schools in the athlete's choice set, recruiting activities, and signals of interest communicated from schools to athletes via social media.

As school choice decisions occur under significant constraints in terms of time, information, and cognitive, we contend that a bounded rationality model may be more appropriate. This research makes a unique contribution to the literature by investigating the role of heuristics in school choice. We find that features tracking the sequence of recruiting events produce predictions consistent with the availability heuristic and that the addition of these features to the baseline model yields a 4% increase in AUC compared the same model constructed without heuristic features. We believe that this finding may be generalized to other recruiting contexts where optimal decision-making is impacted by lack of time and information.

In addition to better understanding the school choice decision process of athletes, this work endeavors to produce an accurate predictive model that can inform the recruiting strategies of college coaches. We also consider how the behaviors and actions of athletes on social media can act as early signs of school preference. While athletic commitments receive a great deal of attention in the mass media, this research is the first study incorporating information from athletes' social media into an empirical predictive model. Our results demonstrate that social media features consistently add value to predictive models. The success of the combined model (Model 5) suggests that incorporating information from both recruiting data and social media would be most effective. Among the four types of social media features tested, the models focusing on structural features (in-links and out-links) and hashtags contribute more to predictive performance than those based on social media interactions. These findings indicate that coaches looking for early indications of athletes' school choice preferences should focus on out-links and content posted on social media. While our results are not causal, they also suggest that schools may be able to positively impact the probability that a recruit will commit by connecting with them on Twitter.

There are several interesting directions for future work. While this project focuses specifically on predicting where an athlete will commit rather than when, predicting the timing of commitments can be further explored. In addition, differences in decision-making among groups of recruits should be investigated. For instance, Mirabile and Witte (2015) analyzed the performance of their predictive model on athletes by star rating, as top recruits may value different school choice factors compared to lower-ranked recruits, and Popp et al. (2011) compared the college selection process for international and domestic student-athletes. Looking at differences in school choice decisions between early commits and late commits as well as between athletes at different positions (quarterback, kicker, etc.) would be a promising extension of this work. As our project provides preliminary evidence for the value of text data in predicting school choice, deeper analysis of preferences communicated via social media posts, such as topic models and sentiment analysis, present possible extensions to this work. Overall, this study

represents both a promising first step in analyzing and predicting school choice in college football using social media data and using social media predictors in other recruitment contexts.

# References

- 1. 247Sports (2005) FB Recruiting. Accessed June 1, 2015, http://247sports.com.
- 2. Barden JQ, Bluhm DJ, Mitchell TR, Lee TW (2013) Hometown proximity, coaching change, and the success of college basketball recruits. *J Sport Manage* 27:230-246.
- Bricker J and Hanson A (2013) The impact of early commitment on games played: evidence from college football recruiting. *S Econ J* 79(4):971-983, http://dx.doi.org/10.4284/0038-4038-2010.119.
- Broughton A, Foley B, Ledermaier S and Cox A (2013) The use of social media in the recruitment process. *ACAS*, http://apo.org.au/files/Resource/ACAS\_social-media-in-therecruitment-process\_2013.pdf.
- Chapman DS, Uggerslev KL, Carroll SA, Piasentin KA, Jones DA (2005) Applicant attraction to organizations and job choice: a meta-analytic review of the correlates of recruiting outcomes. *J Appl Psychol* 90(5):928-944, http://dx.doi.org/10.1037/0021-9010.90.5.928.
- Cobb BR, Basuchoudary A (2009) A decision analysis approach to solving the signaling game. *Decis Anal* 6(4):239 - 255, http://dx.doi.org/10.1287/deca.1090.0148.
- Crabtree J (2016) The 'social' science of recruiting. *ESPN* (January 26), http://espn.go.com/college-football/recruiting/story/\_/id/14646545/social-media-becomespowerful-aide-dangerous-connection-recruiting.
- 8. Croft C (2008) Factors influencing Big 12 Conference college basketball male studentathletes' selection of a university. Dissertation, The University of Texas at El Paso.

- Davenport R (2015) The recruiting guy: Under Armour All American survey. Whole Hog Sports (December 29), http://www.wholehogsports.com/news/2015/dec/29/under-armour-allamerican-survey/?f=recruiting.
- Dees RA, Nestler ST, Kewley R (2013). WholeSoldier performance appraisal to support mentoring and personnel decisions. *Decis Anal* 10(1):82-97, http://dx.doi.org/10.1287/deca.1120.0263.
- 11. Dumond JM, Lynch AK, Platania J (2008) An economic model of the college football recruiting process. *J Sport Econ* 9(1):67-87, http://dx.doi.org/10.1177/1527002506298125.
- Elliot R, Maguire J (2008) Getting caught in the net: examining the recruitment of Canadian athletes in British professional ice hockey. *J Sport Soc Issue* 32(2):158-176, http://dx.doi.org/10.1177/0193723507313927.
- Elliott B (2015) 8 pieces of advice from parents of star college football recruits. SB Nation (August 4), http://www.sbnation.com/college-football-recruiting/2015/8/4/9060917/ncaafootball-recruit-parents.
- 14. Elliott B, Kirshner A (2016) Players like it, but the NCAA's new social media rule could be chaos for coaches. SBNation (August 1), http://www.sbnation.com/college-footballrecruiting/2016/4/14/11429908/coaches-recruits-twitter-facebook-social-media-rules.
- 15. Granovetter M (1973) The strength of weak ties. Am J Sociol 78:1360-80.
- 16. Highhouse S, Dalal RS, Salas E (Eds.) (2014) Judgment and decision making at work (Routledge, New York).
- 17. Hogarth RM, Karelaia N (2006) Regions of rationality: Maps for bounded agents. *Decis Anal* 3(3):124-144, http://dx.doi.org/10.1287/deca.1060.0063

- Huyn CH, Simon J (2016) Using means objectives to present risk information. *Decis Anal* 13(2):117-127, http://dx.doi.org/10.1287/deca.2015.0328.
- 19. Jowett S, Timson-Katchis M (2005) Social networks in sport: parental influence on the coach-athlete relationship. *Sport Psychol* 19:267-287.
- 20. Lujan C (2010) How well informed are high school student-athletes about post-secondary options? Thesis, University of Arizona.
- 21. McManus MA, Baratta JE (1992) The relationship of recruiting source to performance and survival. Paper presented at the Meeting of the Society for Industrial and Organizational Psychology, Montreal, Canada.
- 22. Mirabile MP, Witte MD (2015) A discrete-choice model of college football recruit's program selection decision. *J Sport Econ*, http://dx.doi.org/10.1177/1527002514566278.
- Montgomery JD (1991) Social networks and labor-market outcomes: toward an economic analysis. *Am Econ Rev* 81(5):1408-1418, http://www.jstor.org/stable/2006929.
- 24. Myerberg P (2015) The best quarterback recruits become the best recruiters. USA Today (July 10), http://www.usatoday.com/story/sports/ncaaf/2015/07/09/college-footballrecruiting-quarterbacks-dwayne-haskins-coaches/29920671/.
- 25. National Collegiate Athletic Association (2015a) Recruiting calendars. Accessed January 1, 2016, http://www.ncaa.org/student-athletes/resources/recruiting-calendars/2015-16-division-i-and-ii-recruiting-calendars.
- National Collegiate Athletic Association (2015b) Division I manual. Accessed January 1, 2016,

http://www.ncaapublications.com/DownloadPublication.aspx?download=D116APR.pdf.

- 27. National Collegiate Athletic Association (2015c). Estimated probability of competing in college football. Accessed January 1, 2016, http://www.ncaa.org/about/resources/research/football.
- 28. National Collegiate Athletic Association (2016) Major infractions database. Accessed June 1, 2016, https://web1.ncaa.org/LSDBi/exec/miSearch. Pitts JD, Rezek JP (2012) Athletic scholarships in intercollegiate football. *J Sport Econ* 13(5):515-535, http://dx.doi.org/10.1177/1527002511409239.
- 29. Popp N, Pierce D, Hums MA (2011) Comparison of the college selection process for international and domestic student-athletes at NCAA Division I universities. *Sports Management Review* 14:176-187, http://dx.doi.org/10.1016/j.smr.2010.08.003.
- 30. Rutter E (2015) 3-Star LB Andre Polk talks Minnesota commit and peer recruiting. *Today's U Sports* (October 8), http://www.todaysu.com/recruiting/cfb-recruiting/3-star-lb-andre-polk-chooses-minnesota/.
- Sander L (2008) For college athletes, recruiting is a fair (but flawed) game. *Chron High Educ* (December 19), http://chronicle.com/article/For-College-Athletes/8066.
- 32. SB Nation College News (2014) The regular season's last football rankings: playoff, polls, and computers. SB Nation (December 7)m http://www.sbnation.com/college-football/2014/12/7/7347525/college-football-rankings-top-25-playoff-ap-coaches.
- 33. Schawbel D (2012). How recruiters use social networks to make hiring decisions now. *TIME* (July 9), http://business.time.com/2012/07/09/how-recruiters-use-social-networks-to-makehiring-decisions-now/.
- 34. Sherman M (2012) Balancing the recruiting budget. ESPN (June 12), http://espn.go.com/college-sports/recruiting/football/story/\_/id/8041461/the-cost-recruiting.

- 35. Spence AM (1973) Job market signaling. Q J Econ 87:355–374.
- 36. Steinberg L, Cauffman E (1996) Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law Hum Behav* 20(3): 249-272. doi: 10.1007/BF01499023.
- 37. Treadway DC, Adams G, Hanes TJ, Perrewe PL, Magnusen MJ, Ferris GR (2014) The roles of recruiter political skill and performance resource leveraging in NCAA football recruitment effectiveness. *J Manage* 40(6):1607-1626, http://dx.doi.org/10.1177/0149206312441836.
- Tversky A, Kahnemann D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.
- 39. Twitter (2015) REST APIs. Accessed August 1, 2015, https://dev.twitter.com/rest/public.
- 40. U.S. News Education (2014) National universities rankings. US News World Rep, http://colleges.usnews.rankingsandreviews.com/best-colleges.
- 41. Vecchio RP (1995) The impact of referral sources on employee attitudes: Evidence from a national sample. *J Manage* 21: 953-965.