

Summer 2018

From hashtags to Heismans: social media and networks in college football recruiting

Kristina Gavin Bigsby
University of Iowa

Copyright © 2018 Kristina Gavin Bigsby

This dissertation is available at Iowa Research Online: <https://ir.uiowa.edu/etd/6371>

Recommended Citation

Bigsby, Kristina Gavin. "From hashtags to Heismans: social media and networks in college football recruiting." PhD (Doctor of Philosophy) thesis, University of Iowa, 2018.
<https://doi.org/10.17077/etd.zm7ol4dj>

Follow this and additional works at: <https://ir.uiowa.edu/etd>

FROM HASHTAGS TO HEISMANS:
SOCIAL MEDIA AND NETWORKS IN COLLEGE FOOTBALL RECRUITING

by

Kristina Gavin Bigsby

A thesis submitted in partial fulfillment
of the requirements for the Doctor of Philosophy
degree in Informatics (Information Science) in the
Graduate College of
The University of Iowa

August 2018

Thesis Supervisors: Associate Professor Jeffrey W. Ohlmann
Assistant Professor Kang Zhao

Copyright by
KRISTINA GAVIN BIGSBY
2018
All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the PhD. thesis of

Kristina Gavin Bigsby

has been approved by the Examining Committee for
the thesis requirement for the Doctor of Philosophy degree
in Informatics (Information Science) at the August 2018 graduation.

Thesis Committee:

Jeffrey W. Ohlmann, Thesis Supervisor

Kang Zhao, Thesis Supervisor

Kajsa Dalrymple

Padmini Srinivasan

Nick Street

ACKNOWLEDGMENTS

I would like to thank the community of family, friends, and University of Iowa faculty and staff that have supported me while I completed this work.

Thank you to my advisors, Jeffrey Ohlmann and Kang Zhao. It was my honor to be the recipient of your patience, wisdom, and guidance for the past four years.

Gratitude to my committee, whose mentorship and support has helped me improve as a student and scholar.

Many thanks to the Interdisciplinary Graduate Program in Informatics and the Graduate College for funding assistance during my graduate education.

I am eternally grateful to my parents, who instilled in me a love of learning. This achievement is as much yours as mine.

And a very special thanks to my husband, Patrick, for providing moral support and (most importantly) vast quantities of coffee during the research and writing process.

ABSTRACT

Social media has changed the way that we create, use, and disseminate information and presents an unparalleled opportunity to gather large-scale data on the networks, behaviors, and opinions of individuals. This dissertation focuses on the role of social media and social networks in recruitment, examining the complex interactions between offline recruiting activities, online social media, and recruiting outcomes. Specifically, it explores how the information college football recruits reveal about themselves online is related to their decisions as well as how this information can diffuse and influence the decisions of others.

Recruitment occurs in many contexts, and this research draws comparisons between college football and personnel recruiting. This work is one of the first large-scale studies of social media in college football recruiting, and uses a unique dataset that is both broad and deep, capturing information about 2,644 recruits, 682 schools, 764 coaches, and 2,397 current college football players and tracking offline and online behavior over six months. This dissertation comprises three case studies corresponding to the major decisions in the football recruiting cycle—the coach’s decision to make a scholarship offer, the athlete’s decision to commit, and the athlete’s decision to decommit.

The first study investigates the relationship between a recruit’s social media use and his recruiting success. Informed by previous work on impression management in personnel recruitment, I construct logistic classifiers to identify self-promotion and ingratiation in 5.5 million tweets and use regression analysis to model the relationship between tweets and scholarship offers over time. The results indicate that tweet content

predicts whether an athlete will receive a new offer in the next month. Furthermore, the level of Twitter activity is strongly related to recruiting success, suggesting that simply possessing a social media account may offer a significant advantage in terms of attracting coaches' attention and earning scholarship offers. These findings underscore the critical role of social media in athletic recruitment and may benefit recruits by informing their branding and communication strategies.

The second study examines whether a recruit's social media activity presages his college preferences. I combine data on recruits' college options, recruiting activities, Twitter connections, and Twitter content to construct a logistic classifier predicting which school a recruit will select out of those that have offered him a scholarship. My results highlight the value of social media data—especially the hashtags posted by the athlete and his online social network connections—for predicting his commitment decision. These findings may prove useful for college coaches seeking innovative methods to compete for elite talent, as well as assisting them in allocating recruiting resources.

The third study focuses on athletic turnover, i.e., decommitments. I construct a logistic classifier to predict the occurrence of decommitments over time based on recruits' college choices, recruiting activities, online social networks, and the decommitment behavior of their peers. The results further underscore the power of online social networks for predicting offline recruiting outcomes, giving coaches the tools to better identify vulnerable commitments.

PUBLIC ABSTRACT

Social media has changed the way that we create, use, and disseminate information. This dissertation explores how the information that individuals reveal about themselves through connections, content, and interactions on social media can be analyzed to understand and predict their offline decisions as well as how this information can diffuse and influence the decisions of others. Focusing on an application to college football recruitment, this research comprises three studies analyzing and predicting the coach's decision to offer a scholarship, the recruit's decision to commit, and the recruit's decision to decommit based on recruiting and Twitter data. As one of the first large-scale studies of social media in college football recruiting, this work may benefit athletes and coaches by informing their communication and networking strategies during recruitment. Additionally, my approach to using online social media data to predict offline outcomes will also be of interest outside of the sports world, highlighting the transformative role of social media on recruitment processes in other domains.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 OFFERS.....	14
2.1 Related Research.....	17
2.2 Methods.....	23
2.2.1 Data	23
2.2.2 Defining Impression Management.....	27
2.2.3 Classifying Impression Management.....	30
2.2.4 Offer Analysis.....	34
2.3 Results.....	38
2.3.1 Tweet Classification.....	38
2.3.2 Modeling Offers.....	39
2.3.3 Predicting Offers.....	45
2.4 Discussion.....	46
CHAPTER 3 COMMITMENTS	52
3.1 Related Research.....	54
3.2 Methods.....	60
3.2.1 Data	60
3.2.2 Modeling the School Choice Decision	62
3.2.3 Predicting School Choice.....	69
3.3 Results.....	70

3.3.1 Factors Related to School Choice	71
3.3.2 Predictive Performance	76
3.3.3 Application to Recruiting Resource Allocation	79
3.4 Discussion	81
CHAPTER 4 DECOMMITMENTS	84
4.1 Related Research	87
4.2 Methods	92
4.2.1 Data	93
4.2.2 Modeling Decommitments	94
4.2.3 Predicting Decommitments	99
4.3 Results	103
4.3.1 Factors Related to Decommitment	103
4.3.2 Predictive Performance	108
4.3.3 Application to Recruiting Decision Support	111
4.4 Discussion	115
CHAPTER 5 CONCLUSIONS AND FUTURE WORK	118
REFERENCES	126

LIST OF TABLES

Table 1. Summary statistics of recruit Twitter activity (September 2015).....	25
Table 2. Overview of tweet dataset.....	26
Table 3. Labeled tweet data	31
Table 4. Tweet classifier features	32
Table 5. Offer analysis features	35
Table 6. Tweet classifier testing results.....	38
Table 7. Fitted ZINB regressions.....	40
Table 8. Offer classifier results	45
Table 9. Baseline commitment features.....	64
Table 10. Social media commitment features.....	66
Table 11. Fitted commitment regressions.....	72
Table 12. Top 10 Iowa prospects (January 2016).....	80
Table 13. Baseline decommitment features.....	96
Table 14. Social media decommitment features.....	98
Table 15. Decommitment classifier testing results.....	101
Table 16. Fitted decommitment regressions.....	104
Table 17. Decommitment predictions for Utah recruits (November 2015)	111

LIST OF FIGURES

Figure 1. Factors, information sources, and constraints in recruiting decisions	10
Figure 2. Total number of offers received by month (June 2012-March 2016).....	34
Figure 3. Distribution of offers received per month	37
Figure 4. Performance of commitment predictions	77
Figure 5. Performance of school choice rankings.....	79
Figure 6. Performance of decommitment predictions	109
Figure 7. Trends in predicted probability (October 2015-February 2016).....	114

CHAPTER 1 INTRODUCTION

Social media has changed the way that individuals create, use, and disseminate information and presents the new opportunities for researchers to gather large-scale data on the networks, behaviors, and opinions of individuals. Actions on social media, whether communicative (e.g., posting a meme) or connective (e.g., following), constitute information artifacts that, like their archaeological analogues, give insight into the culture and beliefs of their creators. Moreover, the social network aspect of sites like Facebook and Twitter accelerates the process of information dissemination and amplifies the influence of these artifacts on other users.

This dissertation explores how the information people reveal about themselves on social media can be analyzed to understand and predict their offline decisions as well as how this information can diffuse and influence the decisions of others. The predictive power of social media data has been supported in the context of box office success (e.g., Asur & Huberman, 2010; Ding, Chen, Duan, & Jin, 2016), event attendance (e.g., Bogaert, Ballings, & Van den Poel, 2015; Zhang, Zhao, & Cao, 2015), and political elections (e.g., Franch, 2012; Kristensen et al., 2017), among others. Focusing on an application to college football recruitment, my work examines the complex interactions between offline recruiting activities, online social media, and recruiting outcomes.

Recruitment—the process of attracting, selecting, and obtaining suitable candidates for positions in an organization—occurs in many contexts such as religious groups, higher education, fraternities and sororities, and the military. This research specifically draws comparisons between human resources (HR) and college football recruitment. While personnel recruitment has been extensively studied in the

management, sociology, and psychology literatures, athletic recruitment has garnered relatively little research attention, and I investigate the applicability of HR theories and constructs to the athletic domain. I contend that analyzing the wealth of data related to college football recruiting can not only help us better understand context-specific processes and predictors of football recruiting outcomes, but inform generalizable conclusions regarding the value of social media data for forecasting other types of recruiting decisions.

Human resources are commonly understood to be among a company's "most important assets and...may offer the only non-imitative competitive edge" (Singh & Finn, 2003, p. 395). How to attract and retain valuable employees is an age-old question; Julius Caesar instituted what was potentially the first referral program for Roman army recruitment in 55 B.C. (Behera & Pathy, 2013), and the modern roots of HR can be traced back to the middle of the 20th century (e.g., Malm, 1954; Parnes, 1954). HR recruiting and retention is a high-stakes issue, with good employees adding value to an organization through their work product, knowledge, and even their professional and personal networks. However, in a recent Towers Watson survey (2014) of 321 U.S. and Canadian companies, 48% reported difficulty attracting top-performing candidates, with 35% having trouble retaining their most valuable employees. At the same time, a Deloitte research team found that companies spent an average of approximately \$4,000 per hire in 2014, an increase of 7% on average over the previous year (Krider, O'Leonard, & Erickson, 2015). Furthermore, companies reported an average of 52 days to fill an open position, an increase of 6 days. It is clear that improved recruiting processes are necessary to attract and retain the best talent in the ever-changing world of business.

Employers have long looked to the power of networks in their quest to reduce recruitment costs and increase benefits. In the 1993 National Organizations Study, more than one third of organizations reported using employee referrals as part of their recruitment strategy (Marsden, 1994). Montgomery (1991) explored the role of social network homophily during the screening process, positing that employers use the characteristics of the referrer as a signal of the referral's quality, and referrals have been linked better-informed candidates (Quagliari, 1982), enhanced applicant quality (Kirnan, Farley, & Geisinger, 1989), and improved retention (McManus & Baratta, 1992). A meta-analysis by Zottoli and Wanous (2000) linked referrals to an 18% gain in job survival and a slight, but significant performance advantage.

While the role of offline social networks in job-seeking and recruiting has been widely studied (e.g., Granovetter, 1973; Bian, 1997; Han, 2009), the introduction of information technology into the recruiting process has redefined the concept of "social recruitment," suggesting the need for more work on this topic. The proliferation of social media sites like Twitter, Facebook, and LinkedIn offers both candidates and organizations new avenues to engage in networking and communication, as well as additional sources of information to use when making recruiting decisions. Broughton, Foley, Ledermaier, and Cox (2013) found that in 2011 over half of British job-seekers utilized social media in their job search, including 18% on Facebook and 31% using LinkedIn. For employers, social media has been described as a "real time search engine" (Penttila, 2009, p. 2), with 64% of companies using social media to inform hiring decisions (HRreview, 2012). In recent years, companies have tripled their investment in professional social networking sites like LinkedIn (Kridler et al., 2015). Yet while we

know that organizations are utilizing social media, the ramifications have not yet been thoroughly studied in the research literature. As stated in a review of prior work on social media and recruitment by Roth, Bobko, Iddekinge, & Thatcher, (2016):

Organizational practice has greatly outpaced the scientific study of SM assessments in an area that has important consequences for individuals (e.g., being selected for work), organizations (e.g., whether this information helps predict job performance or withdrawal), and society (e.g., adverse impact, diversity). (p. 288)

This research takes a small step toward answering questions about these offline consequences of online social media. For instance, has social media become a necessary professional tool, i.e., do users have a significant advantage over non-users in terms of attracting offers? Are individuals more likely to accept an offer if an organization connects with them on social media? Are an individual's online friendships correlated to turnover intentions?

However, obtaining empirical data on HR recruiting is often difficult. The majority of previous studies examining social networks in recruitment have relied on survey/questionnaire methods. For example, Granovetter's (1973) seminal paper on the "strength of weak ties" surveyed Boston area job-seekers about job information sources. While questionnaires are a practical method for gathering data from a large sample, they are vulnerable to self-reporting bias, i.e., respondents may not remember accurately or may actively misreport information about their social networks, job-seeking activities, or turnover intentions. Additionally, because they often depend on simply asking individuals who their social ties are, questionnaires best support egocentric network analysis (Butts, 2008) and may not capture the complex interpersonal and inter-firm dynamics of

recruiting networks. Alternatively, other researchers have worked with companies to gain access to private personnel data (e.g., Fernandez, Castilla, & Moore, 2000). Proprietary data offers the advantage of being very deep, with the possibility of tracking many different aspects of an employee's tenure with the company including hiring, performance ratings, compensation history, and separation. However, such data may not possess great breadth at either the individual—lacking data about other employment options or networks outside of the workplace—or market level—giving a detailed snapshot of only one organization.

This dissertation takes a novel approach to exploring the complex interactions between online social media and offline recruiting outcomes by focusing on an understudied recruitment context, college athletics. Indeed, a recent meta-analysis of the employee turnover literature has called for researchers to “further study underexplored cultures and occupations,” noting a high degree of occupational homogeneity in previous HR studies, with the majority focused on white-collar or healthcare workers (Rubenstein, Eberly, Lee, & Mitchell, 2018, p. 55). Several features of athletic recruiting make it a promising area for research on social networks and social media in recruitment.

First, college football is often a significant part of the public face of an institution. In 2017 alone, over 42 million fans attended Division I football games (NCAA, 2018). Successful programs stand to make money via ticket sales, merchandise, postseason bonuses, and television contracts. Texas A&M University had the most profitable athletic department for the 2015 fiscal year at over \$57 million (Berkowitz et al., 2016). Athletic success has been linked to trends in general student body enrollment (Goff, 2000; Mixon & Yu, 1994; Pope & Pope, 2009), institutional reputation (Goidel & Hamilton, 2006),

and donor behaviors (Baade & Sundberg, 1996; Humphreys & Mondello, 2007; Koo & Dittmore, 2014; Meer & Rosen, 2009; Smith, 2009; Stinson & Howard, 2007). For example, Toma and Cross (1998) analyzed undergraduate enrollment data for 24 universities before and after winning a national championship in football or men's basketball, finding that the number of applications increased after championship seasons both in absolute terms and relative to their non-winning peers.

Yet these benefits do not come without major investments on the part of the university. Oregon's record 2015 profits came after spending over \$137 million (Berkowitz et al., 2016). In college sports, on-field success is highly dependent on recruiting (Caro, 2012; Dronyk-Trosper & Stitzel, 2015; Langelett, 2003). Bergman and Logan (2014) performed a regression analysis on data from 2002-2012, finding that individual recruit quality is a significant predictor of team success, with each additional five-star recruit increasing the number of wins by 0.306. The worth of a premium college football player in terms of marginal contribution to athletic department revenue has been estimated between \$500,000 and \$2.3 million per season (Brown, 1993; Hunsberger & Gitter, 2015). Thus, it is unsurprising that college football teams expend upwards of \$1 million per year on recruiting (Brady, Kelly, & Berkowitz, 2015), not to mention the time and energy of coaches and recruiting staff. *Meat Market*, a profile of then-head coach at Ole Miss, Ed Orgeron, details a seemingly never-ending cycle of staff meetings; video review; grueling travel for evaluation, combines, games, clinics, and visits; and endless phone calls to recruits, parents, and high school coaches (Feldman, 2007). Like their peers in the corporate world, major athletic departments are engaged in an "arms race"

(Jones, 2013, p. 601), seeking improved and innovative recruiting strategies in order to compete for elite talent.

Second, the “sports analytics revolution,” popularized by works like *Moneyball* (Lewis, 2003) and *The Signal and the Noise* (Silver, 2012), has increased interest in applying data-driven methods to athletics. As in the HR domain, it is clear that improved recruiting processes are necessary for college athletic teams to attract and retain the best talent. College athletic recruiting is a logical extension of the analytics-based personnel management and player evaluation that is already being implemented in professional sports. However, very little research has focused on the role of social networks and social media in athletic recruitment, and my work represents a significant addition to this rapidly growing discipline.

Third, social media is highly utilized by high school and college athletes—a recent Pew report on social media finds that 89% of young adult Internet users are on social media (2014). As stated by Ryan Gunderson, former director of player personnel for the University of Nebraska, “Nothing has impacted recruiting more in the last 20 years than social media” (as cited in Crabtree, 2016). While sports media outlets like *ESPN* and *SB Nation* have widely reported on coaches’ use of social media to monitor players’ behavior and the proliferation of peer recruiting (e.g., Caldwell, 2015; Crabtree, 2016; Myerberg, 2015), the evidence is largely anecdotal. One exception is Cornerstone Reputation’s survey of 477 Division I-III college coaches across 19 sports regarding about their utilization of online information resources when evaluating recruits (Safian, 2016). Of the 85% who reported searching for information about recruits online, 87% used Facebook, 79% used Twitter, 65% used Instagram, figures that are significantly

higher than the proportion of HR managers using social media to screen job candidates (Broughton et al., 2013). This dissertation is the first study specifically focusing on football recruits' use of social media and makes a unique contribution to the fields of sports analytics and organizational management by exploring the complex interactions between online social media and offline recruiting outcomes.

Most importantly, more than any other type of recruiting, college athletics comes with large amounts of publicly available data. Online recruiting databases contain information on thousands of high school athletes, with profiles tracking basic demographic information, performance statistics, recruiting activities, and recruiting outcomes. Furthermore, because of the degree of public interest in college sports, athletes and coaches acquire near-celebrity status, making it easier to identify and track the social media activities of individuals in the sphere of college football than it would be in virtually any other recruiting context. In contrast to previous work on recruiting networks in the HR domain, this dissertation utilizes publicly available data on the recruiting activities and online social networks of college football recruits in the class of 2016. The data sources selected offer both depth—tracking each athlete's online and offline behavior over the course of six months—and breadth—capturing information about all of the athletes and schools in the football labor market. This enables a holistic approach to studying the role of networks in recruiting, with detailed data on both intra- and inter-firm connections. While there are still many factors involved in the recruiting process that are not publicly observable, such as recruits' academic performance, disciplinary records, or private communications with coaches, I contend that using publicly available data from

247Sports and Twitter is an advantage because it lends transparency to my work, with results that can be verified and replicated by other researchers.

This dissertation focuses on relating connections and activities in online social networks with offline recruiting outcomes. I have three major goals in this work. First, I seek to describe the use of social media by college football recruits and coaches.

Although anecdotal evidence (e.g., Crabtree, 2016) and surveys of coaches (Safian, 2016) suggest that social media plays an increasingly large role in athletic recruitment, little research has focused on the social media usage by college athletes or recruits, and such work has been fairly limited in scope (e.g., Browning & Sanderson, 2012; Kian & Sanderson, 2014). This dissertation represents the one of the first large-scale, empirical studies of social media in college football recruiting, representing a direct contribution to the growing discipline of sports analytics as well as a unique addition to the field of information science by documenting the type of information created by college football recruits on social media and its relationship to their decisions and the decisions of others. Second, I investigate the extensibility of established theories and frameworks from the HR recruiting, considering the similarities and differences between these two domains. Borrowing from the more established literature surrounding personnel recruitment informs a hybrid data- and theory-driven approach to studying the college football recruiting process. Additionally, college athletic recruiting offers the unique opportunity to take a “big picture” view of a labor market and yield significant findings about the role of networks and technology in recruiting that may be generalizable to other contexts. Third, I investigate the value of social media data for predicting recruiting decisions. While previous studies have constructed predictive models for the number of scholarship

offers (Pitts & Rezek, 2012), school choice (e.g., Dumond, Lynch, & Platania, 2008; Mirabile & Witte, 2015), and future performance (Peng et al., 2018) in college football, none have incorporated social media information.

This dissertation comprises three case studies that correspond to the major decisions of the athletic recruiting cycle—the coach’s decision to make a scholarship offer, the recruit’s decision to commit, and the recruit’s decision to decommit—and examines how the information created on Twitter by recruits via their connections and content can be used to understand and predict these decisions. Figure 1 displays the three stages, the major factors that impact each decision, the information sources consulted, and known constraints on the information-seeking and decision-making processes.

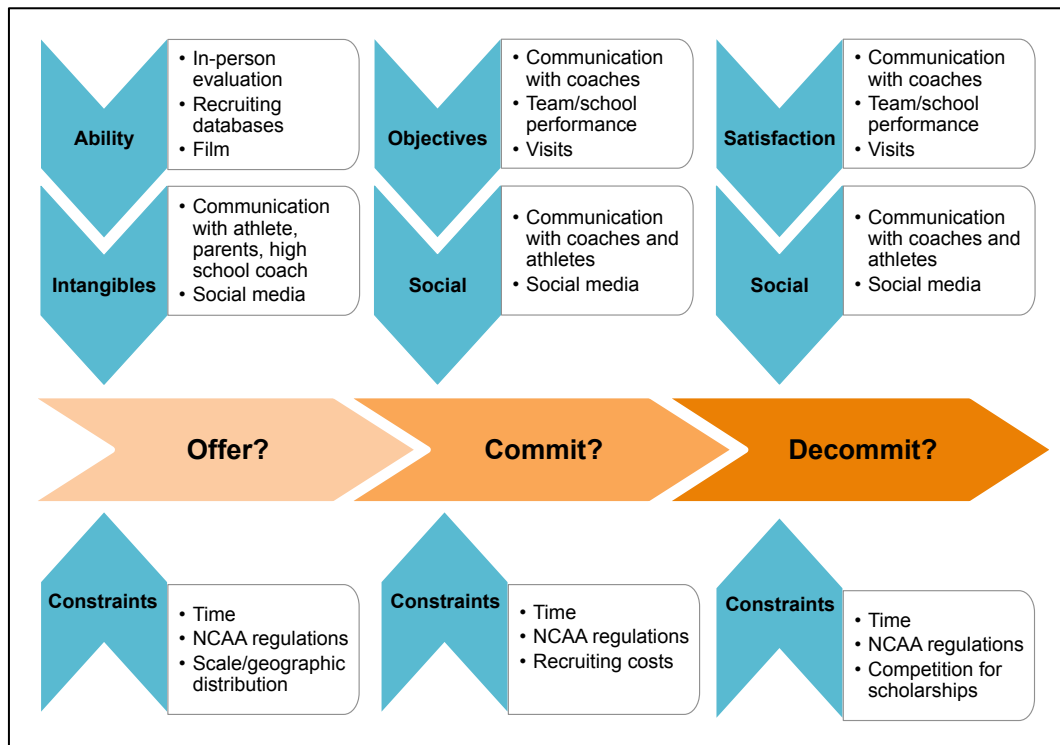


Figure 1 Factors, information sources, and constraints in recruiting decisions

The first study (Chapter 2), focuses on the screening and selection phase of college football recruiting, resulting in the coach’s decision to extend a scholarship offer.

Similar to the HR context, where organizations vet candidates with consideration of their application materials, work histories, and interviews, coaches obtain information on recruits' athletic abilities from recruiting databases, game film, and in-person observations. In both domains, these traditional sources of information may not be sufficient to properly gauge "intangible" qualities like personality, character, and maturity. Evidence suggests that both hiring managers and college coaches are increasingly turning to social media when screening prospects (Broughton et al., 2013; Safian, 2016). This study analyzes the tweets posted by recruits in the class of 2016 from the perspective of impression management theory. Specifically, I identify instances of self-promotion and ingratiation in recruits' tweets and model the relationship between Twitter content and scholarship offers over time. This study represents the first empirical research relating athletes' online behavior with their offline recruiting outcomes. My results indicate that positive self-representation online is correlated to recruiting success, and that recruits with social media may have a significant advantage over their peers without an online presence. These findings highlight the role of social media as an indispensable recruiting tool, and may benefit college recruits by informing their communication and branding efforts during recruitment.

The second study (Chapter 3), examines the question of school choice in college football recruiting, which is analogous to job choice in the HR domain. In both settings, a rational decision-making model predicts that candidates will select the organization that maximizes their expected utility, with consideration of objective factors like compensation and benefits. Yet human decision-makers often deviate from rationality, and reviews of the personnel psychology literature have called for further research on the

impact of cognitive biases, heuristics, and social factors in recruiting decisions (Rynes, Heneman, & Schwab, 1980). Moreover, surveys of prospective college athletes indicate that they tend to be ill-informed of their college options and choices under significant time constraints (Lujan, 2010; Sander, 2008), enhancing the impacts of social networks on the decision-making process. This study combines data on recruits' college options, recruiting activities, and social media to predict which school a recruit will select out of those that have offered him a scholarship. I investigate how the connective and communicative actions taken by recruits and others on Twitter can predict commitment decisions. This study makes a significant contribution to the management and personnel recruiting literatures by using social media data to track connections between candidates and organizations. In addition, it represents a novel addition to the athletic recruiting literature as the first work to investigate heuristics in school choice decisions. My findings—that a combination of cost/benefit factors, online social network data, and Twitter hashtags posted by the recruit are useful for predicting his impending commitment—may assist coaches in formulating recruiting strategies and allocating recruiting resources.

The third study (Chapter 4), addresses the problem of athletic decommitments, a phenomenon that mirrors employee turnover. While predictors of turnover intention have been well-studied in the HR domain (Griffeth, Hom, & Gaertner, 2000; Rubenstein et al., 2018), this work is the first to model decommitments. Specifically, I construct a logistic classifier to predict the occurrence of decommitments over time based on institutional data about the recruit's current school and other schools that are recruiting him, his recruiting activities, and social media. In addition to the practical benefit of offering early

warning for coaches seeking to salvage a vulnerable commitment or recruit a replacement, this research makes a novel contribution to the HR domain by demonstrating the utility of online social network data for predicting offline turnover. This work adds to the information science literature by exploring how recruits' online social network ties provide insight into their turnover intentions as well as how the behavior of their peers can influence decommitment decisions.

Chapter 5 contains final conclusions and outlines opportunities for future work. I believe that this dissertation presents a clear value to college football programs, but will also be of interest outside the sports world, due to its overlap with HR recruiting theories and concepts. Specifically, this research represents one of the first attempts to use online social media data to predict offline recruiting outcomes, with direct parallels to employee screening and selection, job choice, and turnover. This dissertation contributes to the larger conversation in the field of information science on the impacts of social media, how individuals create information online, and how that information can be used to understand and predict offline behaviors and decisions.

CHAPTER 2 OFFERS¹

The proliferation of social media sites has given individuals the power to craft personas and influence how others perceive them in a virtual setting. But questions remain as to whether these online identities have measurable offline impacts. In this study, I explore the issue of online self-presentation in recruitment from the perspective of impression management theory, which describes individuals' conscious and subconscious attempts to influence the perceptions of others. While there is a large body of literature studying the role of in-person impression management during the employment interview (e.g., Kacmar, Delery, & Ferris, 1992; Tsai, Chen, & Chiu, 2005), there exists a need for more work on the impacts of online impression management on offline recruiting outcomes.

I utilize a novel approach by focusing on an application to football recruiting and taking advantage of the wealth of publicly available data on college athletics. In addition to better understanding the role of social media and self-presentation in football recruiting, I use sports as a natural laboratory to explore the relationship between social media, impression management, and recruiting in a broader sense. Social media has had a transformative effect on the college football recruiting process, and mass media outlets have widely reported on coaches' strategic use of Twitter to engage recruits (Crabtree, 2016), the practice of peer recruiting on social media (Myerberg, 2015), the positive and negative impacts of online interactions with fans (Davenport, 2015), and the growing role of social media as a screening tool (Caldwell, 2015). In the words of Oregon State

¹ A version of this study has been submitted for publication in the Special Issue on Sports Analytics of *Journal of Big Data*. This study was also presented at the 2018 Jakobsen Memorial Conference, Iowa City, IA and received second place in the Social Sciences and Education division.

football's director of player personnel, Darrick Yray, "Every single school does it. You have to, especially since you're investing almost \$500,000 in a player's development over a four- or five-year period" (as cited in Crabtree, 2016). However, there is currently no empirical research focusing on the recruits' use of social media during recruitment. In this work, I make a significant contribution to the sports analytics and athletic recruiting literatures by examining the relationship between college football recruits' behavior on Twitter and the number of scholarship offers they receive. I propose two main research questions:

Q1: What type of impression management strategies do athletes use during recruitment?

This study is the first large-scale examination of college football recruits' Twitter activity. The information science literature is replete with surveys of social media usage by academics (Nentwich & Konig, 2013), non-profits (Lovejoy & Saxton, 2012), the elderly (Xie, Huang, & Watkins, 2012), and millennials (S-O'Brien, Read, Woolcott, & Shah, 2011), among others, but there is very little work examining the use of social media by college athletes. This study represents a unique addition to the field by documenting how athletes create information and craft identities online during the recruiting process. In contrast to previous mass media reports focused on extreme, isolated cases of negative behavior on social media, including "racist, sexist, vulgar or profane posts" (Caldwell, 2015), I explore examples of positive self-representation by athletes. I seek to quantify the prevalence of different types of impression management—self-promotion and ingratiation—in the Twitter timelines of college football recruits.

Q2: What is the relationship between impression management and scholarship offers?

In other words, does engaging in certain behaviors online 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 recruits over time. Specifically, I model the number of new offers in the next month as a function of the athlete's personal characteristics, recruiting activities, and Twitter content up to that point. In addition, I build a logistic classifier predicting whether the athlete will receive at least one offer in the next month. This work makes a unique contribution to the information science literature by analyzing the relationship between athletes' social media content and coaches' scholarship decisions. Previous information science research has examined how public perception of political candidates (Bhattacharya, Yang, Srinivasan, & Boynton, 2015) and companies (Bella, Colladon, Battistoni, Castellan, & Francucci, 2018) can be assessed from general social media chatter, but have not thoroughly explored how the information presented online by these figures can influence others' opinions and decisions.

Overall, my analysis of the Twitter timelines of the class of 2016 shows that self-promotion—posting positive content about the athlete's own athletic prowess, academic abilities, and recruiting achievements—occurs more often than ingratiation—posting positive content about other recruits, college teams, or current college athletes. Furthermore, the explanatory and predictive models give strong evidence of a relationship between social media and recruiting success. Athletes with any level of Twitter activity, including a protected account, were significantly more likely to receive an offer in the next month as well as garnering more offers. Models incorporating social media data were a significantly better fit for the data than the baseline model using only

recruiting data. The results also indicate that self-promotion may be more effective than ingratiation when it comes to attracting scholarship offers. Model 1, which considered the athlete's self-promotion efforts, was a better fit than the baseline incorporating only personal characteristics and recruiting activities and Model 2, which focused on ingratiation. These findings could be useful to athletes, informing their online communication strategies during recruitment. In addition, as one of the first studies to link online content to offline recruiting outcomes, this work can also be generalized to other recruitment settings. It may be of special interest to job-seekers, helping them effectively utilize social media.

2.1 Related Research

There is a large body of literature studying how hiring managers process information to evaluate candidates during the screening process (e.g., Spence, 1973) as well as specifically focusing on the role of impression management in employment interviews. In a laboratory study, Kacmar, Delery, and Ferris (1992) found 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. The importance of impression management was also highlighted in research using real-world hiring data. Stevens and Kristof (1995) conducted a field study, finding that interviewees used self-promotion tactics more often than ingratiation, and that ingratiation did not significantly impact perceptions of suitability or the likelihood of receiving a job offer. Tsai, Chen, and Chiu (2005) found that verbal self-promotion significantly influenced applicant ratings. This effect was moderated by both job characteristics, with self-promotion weighed more heavily for customer-facing positions,

and interview length, with self-promotion having a larger effect during short interviews. In a rare longitudinal study examining the impacts of impression management on job performance ratings (Wayne & Liden, 1995), the authors found that supervisor-focused impression management was positively related to both perceived similarity and supervisor ratings.

However, the literature on online impression management and its impact on recruiting outcomes remains sparse. Social media has changed the landscape of recruiting, at the same time offering individuals new avenues to engage in impression management activities and presenting companies with new information sources to evaluate potential employees. Social media differs from traditional sources (e.g., resumes, interviews) in that it does not require face-to-face interaction, is not limited by time or physical location, and can be accessed at any time. Evidence suggests that social media may be especially useful for obtaining what Rees (1966) referred to as “intensive” information, subjective data that may be difficult to observe through formal methods. Schawbel (2012) highlighted the value of social media for evaluating candidates’ professionalism and communication skills, with HR managers particularly attentive to grammar and spelling. Caers and Castelyns (2010) found that HR professionals used applicants’ profile photographs on LinkedIn and Facebook to assess traits of extraversion and maturity.

While these surveys provide evidence that hiring managers are increasingly consulting social media when making recruiting decisions, there is little research quantifying the impact of online behaviors on offline recruiting outcomes. In a 2013 survey, 35% of HR managers reported finding material on social media that caused them

not to hire a job candidate, especially “provocative” photographs (Broughton et al., 2013). Most relevant to the current work, an experimental study by Bohnert and Ross (2010) confirmed that in addition to resume quality, behavior and preferences indicated on social media profiles of applicants had a significant impact on the overall evaluation of job candidates. Participants rated candidates with family- or professional-oriented Facebook profiles higher than candidates with alcohol-oriented profiles. In addition, participants offered an average of \$2,400 more to candidates who exhibited a professional online presence. My study extends this prior research by utilizing observational data, tracking the recruiting activities, social media content, and scholarship offers of real college football recruits in the class of 2016. Outside of the laboratory, individuals are more likely to construct multi-faceted personas on social media—not easily reducible to discrete categories like “alcohol” or “family.” Instead, I focus on identifying occurrences of self-promotion and ingratiation in individual tweets, then examining the relationship between the proportion of each type of impression management in recruits’ Twitter profiles and their recruiting success. Like job candidates, college football recruits can be both self-promoters and ingratiators, and my analysis attempts to reflect the complexity of online self-presentation.

In the only previous research study examining the offer process in college football recruiting, Pitts and Rezek (2012) analyzed data on members of the class of 2008, finding that the physical attributes of a recruit were the most significant predictors of number of scholarship offers. An increase of 10 pounds of weight was associated with 0.15 additional scholarship offers and an extra inch of height led to 0.604 additional offers. While the authors focused on the relationship between static, demographic characteristics

and the total number of offers at the end of recruitment, my work uses a dynamic approach, modeling the number of new offers in the next month as a function of an athlete's personal characteristics, recruiting activities, and Twitter content to-date.

In addition, this previous work prioritized physical features, yet college coaches also consider an athlete's "intangibles" like maturity, teamwork, and "field sense" (USA Football, 2001). For coaches, the process of deciding which athletes to extend scholarship offers to is complicated by issues of scale, time, and policy. In 2014, coaches at 250 Division I programs recruited from a pool of more than 250,000 high school seniors, with only 2.5% advancing to play Division I football (NCAA, 2015). While evaluation of recruits in other collegiate sports is facilitated by participation in elite summer leagues (e.g., AAU basketball), football recruiters' main source of information on high school players comes in the form of film study, personal interactions, and performance statistics (Feldman, 2007). Social media has become a vital recruiting tool, in no small part due to the ban on text messaging between college coaches and recruits that existed from 2007 until April 2016 (USA Today High School Sports, 2016). As with HR professionals, social media presents coaches with a convenient opportunity to gather nuanced information on recruits' personality and character. Indeed, 99% of Division I-III coaches rated character as either important or very important, and 85% believed that an athlete's online presence gave a "better sense of a recruit's character and personality" (Safian, 2016, p. 3). Mass media reports suggest that coaches are especially attentive to red flags, indications that an athlete might not be a good academic or cultural fit with their institution (Crabtree, 2016). As stated by Herb Hand, then-offensive line coach at Penn

State, “Dropped another prospect this AM due to his social media presence ... Actually glad I got to see the 'Real' person before we offered him” (as cited in Caldwell, 2015).

Yet there are indications that such outcry may be outsized. While 83% of coaches surveyed by Cornerstone Reputation reported finding something negative in a recruit’s online presence, only 19% had ever rescinded an offer based on online content (Safian, 2016). Despite cautionary tales about online faux pas and the adoption of restrictive social media policies by college athletic departments (Sanderson, 2011), anecdotal evidence suggests that social media misbehavior is actually decreasing over time. As stated by Joe Dooley, then-head coach of men’s basketball at Florida Gulf Coast University, “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” (as cited in Caldwell, 2015). Networking and communicating on social media may actually benefit recruits. 90% of coaches surveyed by Cornerstone Reputation saw something online that gave a positive impression, and 82% believed that athlete with strong and positive online presence had an advantage over other recruits (Safian, 2016). Low-rated or unrated recruits must work harder to get noticed by coaches (Horn, 2015), and Sander (2008) noted an increase in recruits initiating contact with coaches—37% of Division I athletes surveyed reported reaching out to coaches during recruitment via email, mail, or sending a video highlight. Social media offers a convenient medium to engage in this type of self-promotion; recruits can instantly connect with coaches on Twitter and share information about themselves by replying, mentioning, and/or including URLs linking to news coverage or their Hudl highlights in their tweets.

While numerous anecdotal accounts suggest the importance of social media in college football recruitment, there exists almost no research on this topic. Focusing on current college athletes, a survey of 202 student-athletes and 419 non-athletes found that, while athletes did not spend significantly more time on Facebook, they did tend to have more friends (Lampe & Ellison, 2010). Browning and Sanderson performed qualitative interviews exploring student-athletes' motivations for using Twitter, noting that "Twitter possesses tremendous connective and identity-building capabilities." (2012, p. 505). The authors touched briefly on the role of social media during recruitment, with one athlete recounting criticism by fans disappointed in his college choice. Previous work on professional athletes has investigated the types of impression management employed across different sports (Pegoraro, 2010; Sanderson, 2013) and differences by gender (Lebel, 2013). This study is the first to investigate the prevalence of different impression management strategies in college football recruits' social media profiles, and represents a significant addition to athletic recruiting literature.

Specifically considering the offline impacts of online content, Pegoraro and Jinnah (2012) discussed four case studies of professional athletes' social media in relation to their sponsorship opportunities, and Lebel (2013) surveyed golf fans about public perception of different impression management tactics utilized by professional golfers on Twitter. In the only study to focus on coaches' use of social media as an evaluation tool, Cornerstone Reputation surveyed college coaches across 19 sports, finding that the number who reported searching for information about recruits online and on social media increased 5% between 2014 and 2015 (Safian, 2016). In contrast, my study uses observational data to examine the connection between recruits' social media activity and

the number of scholarship offers received over time. Also relevant to the current work, Kian and Sanderson (2014) investigated the impact of college football recruits' offline actions on their online popularity, tracking the daily number of Twitter followers for 20 recruits for three weeks before and after National Signing Day. The authors hypothesized that late commitments might enhance "personal brand," with uncommitted recruits attracting more followers in the period before signing. Ultimately, they found no significant difference in followers between uncommitted and committed recruits. While my study shares the same broad question regarding the relationship between information created on recruits' social media and the opinions and decisions of others, I undertake a deeper analysis of athletes' individual characteristics, recruiting activities, and online content, focusing on their relationship to offline scholarship offers. In addition, I utilize a much larger dataset, including 2,644 athletes in the class of 2016.

2.2 Methods

In this study, I begin by identifying instances of impression management in the Twitter accounts of college football recruits. Secondly, I use this data in addition to information about athletes' personal characteristics and recruiting activities to build a statistical model explaining the number of offers received by college football recruits over time. I perform a follow-up analysis predicting whether a recruit will receive at least one offer in the next month, based on his offline and online activities up to that point.

2.2.1 Data

All three studies in this dissertation use data on the class of 2016, starting with recruiting profiles obtained from the sports media website 247Sports.com (CBS Sports, n.d.). 247Sports is a premier source for information on football and basketball at both the

college and professional levels. Their recruiting database contains data on prospective college football players reaching as far back as 1999 and looking forward to the recruitment of current high school freshmen (class of 2022). At the time of data collection (August 2015), there were 2,644 recruits listed in the database for the class of 2016. For each individual, I collected basic information (e.g., height, weight, hometown, position, rating) and timelines of recruiting events (e.g., scholarship offers, visits, and commitments) using the Selenium package for Python (Muthukadan, 2011). The 2,644 profiles were visited again in March 2016 to retrieve final recruiting events and commitments. I also obtained basic information about the schools recruiting these athletes, including location, academic ranking (U.S. News and World Report, 2014), football team ranking (SB Nation College News, 2014), and NCAA disciplinary record (NCAA, n.d.).

In addition to providing a comprehensive college football recruiting data source, 247Sports was selected because a majority of the profiles contained links to the athletes' Twitter profiles. 1,629 of the pages for the class of 2016 had Twitter timelines and IDs embedded. A manual search was conducted for the remaining recruits, and 700 additional Twitter profiles were located. In full, 2,329 recruits in the dataset (88.1%) were linked to public Twitter accounts, while 160 (6.1%) possessed protected accounts. 155 recruits (5.9%) did not have Twitter. Using 247Sports, I was also able to gather 764 Twitter IDs for Division I coaches and 2,397 IDs for current college football players.

Detailed Twitter information for the class of 2016 was collected using the Twitter REST API (Twitter, 2016) and the Scrapy package for Python (Scrapy, 2008). In September 2015, I gathered profile information, friend and follower lists for each recruit

with a public Twitter account. Table 1 displays basic information about the Twitter activities of the recruits as of September 2015.

Table 1 Summary statistics of recruit Twitter activity (September 2015)

	Mean	SD	Min	Max
Account Age (days)	921	466.3	6	2,361
Tweets	4,594	7,422.2	0	64,311
Friends	658	448.3	0	3,841
Followers	1,195	1,184.6	0	14,284
Reciprocated	493	340.5	0	2,821

This summary data suggests that recruits tended to adopt Twitter at a fairly young age. In September 2015, the mean account age was 2.5 years, with a few recruits (1.1%) having joined Twitter as early as 2009, when they would have been only 11 or 12 years old. They were also prolific; on average, recruits in the class of 2016 posted about five tweets per day, though this distribution is highly skewed. While 28.9% of recruits posted fewer than one tweet per day, 13.0% posted more than 10, and 3.4% posted more than 20.

While only one recruit had no friends or followers, the majority were active networkers. Recruits with public Twitter accounts followed 658 other users on average, approximately 11.9% of whom were coaches, current college football players, or other recruits in the class of 2016. Some recruits appeared to use Twitter almost exclusively as a professional networking tool, with up to 77.1% of their friends being college football-related. Only 1.2% of recruit in the data had no out-links to coaches, current college football players, or other recruits. These figures suggest that social media may play an even larger role in athletic recruitment as compared to HR recruitment. While only 35% of job-seekers report using Twitter to connect with potential employers (Westfall, 2017), 96.2% of recruits followed a college football coach on Twitter. On average, recruits had

1,195 followers, approximately 5.9% of whom were coaches, current college football players, or other recruits. Only 0.8% of athletes had no football-related followers.

The Twitter REST API returns a list of a user’s current friends and followers, but does not yield information on when a connection was formed. Therefore, it was important to collect data over time to observe the creation of new ties. I gathered friend and follower lists monthly between October 2015 and March 2016.

In addition to network data, also collected information on each recruit’s tweets. I gathered retrospective tweets (up to 3,200 historic tweets for each individual) in September 2015 as well as new tweets each month from October 2015 to March 2016. This resulted in a dataset of 5.5 million tweets (Table 2).

Table 2 Overview of tweet dataset

Category	Description	Count (%)
Tweets	All statuses posted on recruits’ timelines	5,547,230
Original	Statuses authored by the posting user	2,932,122 (52.9%)
Retweets	Statuses authored by other users and reposted on recruits’ timelines	2,615,108 (47.1%)
Replies	Statuses responding to another user	1,332,622 (24.0%)
Mentions	Statuses referring to a user by screen name with @	3,751,757 (67.6%)
Hashtags	Statuses containing a topic keyword with #	734,213 (13.2%)
URLs	Statuses containing a hyperlink	751,650 (13.6%)
Media	Statuses containing an image or video	983,759 (17.7%)

The full tweet dataset includes both original tweets and retweets; 47.1% of recruits’ tweets were reposting content by other users. Only 7.4% of these retweets were of college coaches, current college football players, or other recruits in the class of 2016. Replies, tweets directed at other users or in response to other users’ tweets, constituted 24.0% of all tweets. 9.0% of replies were directed to coaches, college football players, or other recruits. Mentions, references to other users, were common, occurring in 67.6% of tweets, and 27.5% of mentions were coaches, college football players, or recruits. This figure includes 627,789 self mentions, tweets or retweets that reference the athlete

himself. 13.2% of tweets contained hashtags, words or phrases used to denote tweet topics. Hashtags were cross-referenced with a hand-curated dictionary to determine whether they were relevant to college football. Approximately 7.3% of hashtags referenced college football topics. Users can also post hyperlinks, and 13.6% of tweets contained URLs. 11.8% of these linked to recruiting or sports media websites, such as Hudl, Rivals, or ESPN. Further analyzing the sports-related URLs, I was able to determine whether each 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 I did not analyze the content of media (e.g., images, gifs, videos) posted in recruits' tweets, 17.7% of all tweets contained media.

2.2.2 Defining Impression Management

First introduced by Irving Goffman (1959), impression management theory describes individuals' efforts to influence their perception by others. Over time, several taxonomies for different impression management strategies have been proposed. Goffman's (1959) original dramaturgical metaphor delineated "front stage" and "back stage" performances, while other researchers have categorized strategies as verbal vs. non-verbal or self-focused vs. other-focused (Kacmar et al., 1992). Jones and Pittman (1982) proposed the most commonly used taxonomy of impression management techniques. The authors describe five primary tactics: self-promotion, ingratiation, exemplification, intimidation, and supplication.

Individuals who engage in self-promotion are attempting to enhance perceptions of competence. Self-promotion most often takes the form of publicizing real, measurable accomplishments, rather than outlandish bragging (Jones & Pittman, 1982). Of course,

exceptions do exist (e.g., resume cheating). For a college football recruit, self-promotion on Twitter is likely to take the form of posting tweets publicizing his athletic ability and recruitment activities. Recruits 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 tweeted, “@JeffFaris Hey Coach, can you follow back please. Thanks.” to a Duke assistant coach.

Ingratiation is intended to produce the “attribution of likability” (Jones & Pittman, 1982, p. 235) and may take the form of opinion conformity or favor doing. Ingratiation may be targeted—focused on increasing one’s likability in the eyes of a specific person—or intended for a broad audience. Applied to college football recruiting and Twitter, ingratiation is likely to take the form of mentioning coaches, current college football players, and other recruits or posting positive tweets about other athletes and teams. Khris Pam, a 3-star cornerback, tweeted congratulations to another class of 2016 recruit from his home state of South Carolina, “@ShaedonMeadors Glad Y'all Beat Them Boys Bro.”

The motivation for exemplification is to project “moral worthiness” (Jones & Pittman, 1982, p. 245). Exemplification often takes the form of non-verbal behavior intended to demonstrate discipline and selflessness, such as charitable giving. Exemplification is likely to be difficult to measure in the specific context of college football recruiting, and the relationship between morality-enhancing tactics and recruiting success is more tenuous than ingratiation or self-promotion.

Intimidation is meant to enhance perceptions of danger and fear of negative consequences. While intimidation seems apropos in the competitive atmosphere of sports—mass media contains many accounts of professional athlete “beefs” on Twitter

(e.g., Wong, 2014)—it may not be the wisest tactic to employ during college football recruiting, where the primary goal is to favorably impress coaches.

In contrast, supplication is intended to produce an impression of weakness and dependence. Supplication occurs most often in situations of power imbalance, targeted from a less-powerful individual to a more powerful individual or group (Jones & Pittman, 1982). Because of the physical and competitive nature of college football, it is unlikely that athletes will choose to advertise physical or mental weakness, or that such a strategy would lead to recruiting success.

In this work, I focus on the first two strategies (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 (e.g., Stevens & Kristof, 1995), while the implications of the latter three tactics are not well-understood. My study assumes that these strategies are non-exclusive—that a tweet may be simultaneously self-promoting and ingratiation. For example, Bryce Huff, a 2-star linebacker recruit, tweeted “I'm blessed to say that I've received an offer from Troy University! <http://t.co/3anrze96ss>.” This tweet could be interpreted as self-promotion, with Huff publicizing his recruiting success, or as ingratiating, intended to compliment Troy by expressing Huff's appreciation at being recruited by the program. This approach is intended to reflect the complexity of online self-presentation, in contrast to previous analyses of social media in employment screening (e.g., Bohnert & Ross, 2010), which placed social media profiles into discrete categories.

2.2.3 Classifying Impression Management

Manually identifying occurrences of self-promotion and ingratiation in a dataset of this size would be infeasible. Therefore, I construct supervised classifiers to learn the features that accompany each impression management strategy and can be applied to the full set of 5.5 million tweets.

Due to the fact that almost half of all tweets posted on athletes' timeline were retweets, I elected to consider self-promotion and ingratiation in both original content and retweets. For example, Michel Boykin, a 3-star defensive end, reposted a tweet by another user promoting his own recruiting success and containing a mention of his Twitter screen name, “#Cincinnati offers 2016 Stud DL Mike Boykin (Carrolton HS, GA) @mikebfeb_13 @VarsityPreps @SouthRecruit1.” Although Boykin was not the original author of the tweet, posting it on his timeline clearly presented an example of self-promotion. Mass media reports indicate that coaches also pay attention to retweets, in addition to other online behaviors such as liking and following (Crabtree, 2016). This sentiment was repeated by a coach responding to the Cornerstone Reputation survey, “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” (as cited in Safian, 2016, p. 12).

7,000 tweets (both original and retweets) were randomly selected for hand-labeling as self-promotion and/or ingratiation. A second annotator labeled 100 tweets so that the robustness of the class definitions could be verified. For self-promotion, the annotators agreed on 83% of examples ($\kappa=0.82$). For ingratiation, the annotators agreed on 87% of examples ($\kappa=0.87$). The nuanced differences between the two categories,

hinging on whether a tweet was about the athlete or about another individual, made up a significant portion of the discrepancy between annotators. Because of the high level of agreement, it was determined that the first annotator’s labels could be used. Table 3 displays the number of tweets belonging to each class. While tweeting about football in general was common, self-promotion and ingratiation were more rare, constituting 21.9% and 14.4% of all tweets, respectively.

Table 3 Labeled tweet data

Class	Description	Count (%)
Self-promotion	Tweet or retweet that promotes the athlete’s recruiting success or athletic ability	1,535 (21.9%)
Ingratiation	Tweet promoting or praising another athlete, coach, team, or school	1,005 (14.4%)

Both structural and text features are used when building the classifiers (Table 4). Tweet text was preprocessed using the NLTK package for Python (Loper & Bird, 2002), including removing non-ASCII characters (e.g., symbols, emojis), hashtags, and standard English stop words. In addition, I replaced entities (e.g., mentions, URLs) 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 OtherURL.” I utilized a simple bag of words model to represent the tweet text, where each document is transformed into a vector containing the number of occurrences of each word, out of all words in the corpus. More information on bag of words and other methods of text representation can be found in *Introduction to Information Retrieval* (Manning, Raghavan & Schutze, 2008). In order to reduce the size and sparsity of the feature space, a Porter stemmer was applied and the 0.1% most- and

least-common words in the corpus were excluded. This reduced the number of unique words to approximately 720.

Table 4 Tweet classifier features

Type	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 another 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 another recruit, 0 otherwise
	Mention	1 if tweet contains mention, 0 otherwise
	Coach mentions	Number of coach mentions
	College mentions	Number of college player mentions
	Recruit mentions	Number of mentions of other recruit
	Self mention	1 if tweet mentions posting athlete, 0 otherwise
Content	Sports hashtags	Number of hashtags referencing college football
	URL	1 if tweet contains URL, 0 otherwise
	Self URLs	Number of football-related URLs about posting athlete
	Other URLs	Number of football-related URLs about others
Text	Media	1 if tweet contains media, 0 otherwise
	Words	Bag of words representation of processed tweet text

As different classification methods are suited to different types of problems, I test several methods from the scikit-learn package for Python (Pedregosa et al., 2011): logistic regression, decision tree, naïve Bayes, support vector machines (SVM), artificial neural networks (ANN), and random forest. More details about each classification method can be found in *Introduction to Data Mining* (Tan, Steinbach & Kumar, 2006). Each classifier is trained to identify instances of self-promotion, and a grid search was used 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.

- Naïve Bayes: In order to avoid the problem of zero estimates for unknown terms and zero, an additive smoothing factor of 0.25 was used.
- Support vector machine (SVM): I implemented a linear kernel function with a penalty weight of 1.
- Artificial Neural Network (ANN): The rectified linear unit function was used as the transfer function between the 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, with consideration of a subset of $\sqrt{n_features}$.

To assess the performance of each classification method, I use stratified Monte Carlo cross validation. Over 100 independent trials, I randomly split the full set of 7,000 labeled tweets into two equal subsets, each with the same proportion of positive instances. Because self-promotion is relatively rare, oversampling is applied to the training set to reduce bias. I randomly select and copy positive examples until the training set contains 50% self-promoting tweets. For each individual trial, each classification method is trained on the oversampled data with consideration of the same set of features (Table 4), then applied to the hold-out testing set.

I evaluate performance 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), which measures the probability of ranking a randomly chosen positive instance higher than a randomly chosen negative instance.

2.2.4 Offer Analysis

The task of modeling and predicting scholarship offers is complicated by the fact that offers are made over a long period of time. Although coaches cannot begin directly contacting high school athletes until the start of their junior year (NCAA, 2016), it is not unheard-of to make scholarship offers (sent to the recruit's high school coach or school) as early as freshman and sophomore year. However, a college coach's selection process is almost certainly different when deciding whether to extend an offer to a freshman vs. a senior, especially considering the quantity and quality of information available in each case. Figure 2 displays the number of scholarship offers received each month by recruits in the class of 2016.

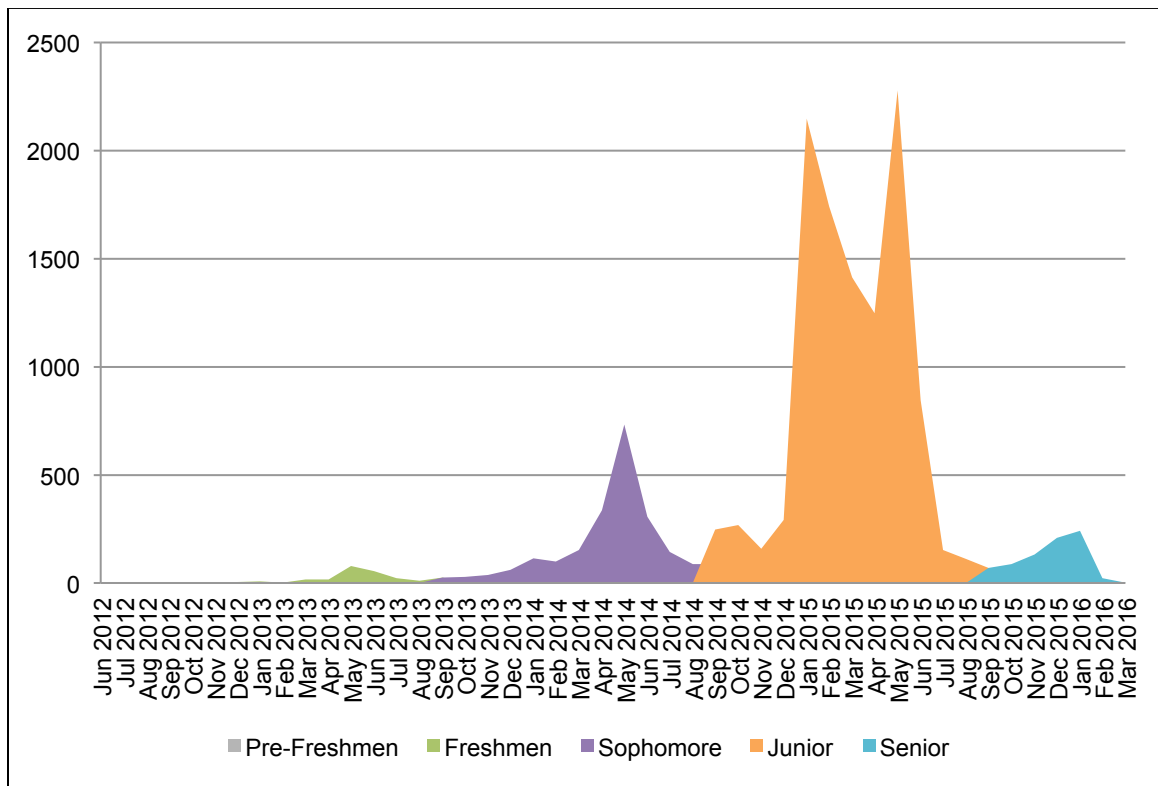


Figure 2 Total number of offers received by month (June 2012-March 2016)

78% of all scholarship offers were received during the recruit’s junior year of high school and following summer, and I elect to focus on this time period. In addition, though the REST API allows retrieval of up to 3,200 retrospective tweets, because some athletes tweet more or less frequently, this figure covers different periods of time.

Narrowing my analyses ensures fairly consistent coverage across all athletes.

To build an explanatory model of the number of new offers received over time, I create a set of “athlete-month” instances. For each recruit in the data, an instance represents a month during his junior year and following summer (9/1/2014-8/1/2015). This yields 26,576 instances, each with features tracking the athlete’s personal characteristics and recruiting activities up to that point in time as well as his social media activity level (categorizing athletes as None, Protected, or into one of four equally-sized groups by tweet frequency) and social media impression management during the prior month, and an outcome variable counting new offers received by the recruit in the next month (Table 5).

Table 5 Offer analysis features

Type	Feature	Description
<i>Personal</i>	star	Recruit star rating (0, 2, 3, 4, 5)
	height	Height (inches)
	weight	Weight (pounds)
	BMI	Body mass index ($weight/height^2$) \times 703
	position	Recruit position (ATH, DB, DL, LB, OB, QB, RB, ST, WR)
	region	U.S. Census region of recruit (INT, MW, NE, S, W)
<i>Recruiting</i>	updates	Number of news updates on 247Sports.com
	camp	Number of college camps attended
	unofficial	Number of unofficial visits
	coach	Number of coach visits
	offers	Number of offers received
	FBS	Proportion of offers received from Division I-FBS school
	verbal	1 if recruit is currently committed, 0 otherwise
	months	Calendar month (9/1/14-8/1/15)
<i>Social Media</i>	activity	Level of Twitter activity based on account status (None, Protected) or number of tweets in prior month (Low=1-18, Mid=19-52, High=53-134, Highest=135+)
	self_promotion	Proportion of self-promoting tweets posted in previous month
	ingratiation	Proportion of ingratiating tweets posted in previous month

In order to examine the relationship of impression management to number of scholarship offers, I compare a series of statistical models. First, the baseline (Model 0) models the number of new offers as a function of the athlete's personal characteristics and recruiting activities to-date. Model 1 incorporates the proportion of self-promoting tweets posted by the recruit. Because Twitter usage varies widely by individual—12.5% of recruits posted 0 tweets during their junior year, while 11.7% averaged 300 or more per month—I also include variables representing the online activity level during the prior month. Model 2 considers the proportion of ingratiating tweets and the level of Twitter activity. Model 3 combines all of the features of the previous models.

I utilize zero-inflated negative binomial regression (ZINB), implemented with the `pscl` package (Zeileis, Kleiber, & Jackman, 2008) for R (R Development Core Team, 2014). First introduced by Greene (1994), zero-inflated models are appropriate for count data that contain excess zeros. Indeed, 81% of athlete-month instances represent months where the recruit received zero offers (Figure 3). ZINB accounts for excess zeros by treating the outcome as subject to two distinct processes. First, an instance is classified as either zero or non-zero using a logistic regression. Second, for instances determined to be non-zero, the number of occurrences is modeled using a negative binomial regression. Because the data is also overdispersed—the variance in offers per month ($\sigma^2 = 1.11$) is much larger than the mean ($\mu = 0.36$)—I elect to use ZINB as opposed to zero-inflated poisson regression. While zero-inflated models can treat the outcome as dependent on two completely separate processes, my analysis assumes that the same features determine the likelihood of receiving an offer and the number of offers received, and uses the same set of variables in the count and zero-inflation portions of the model.

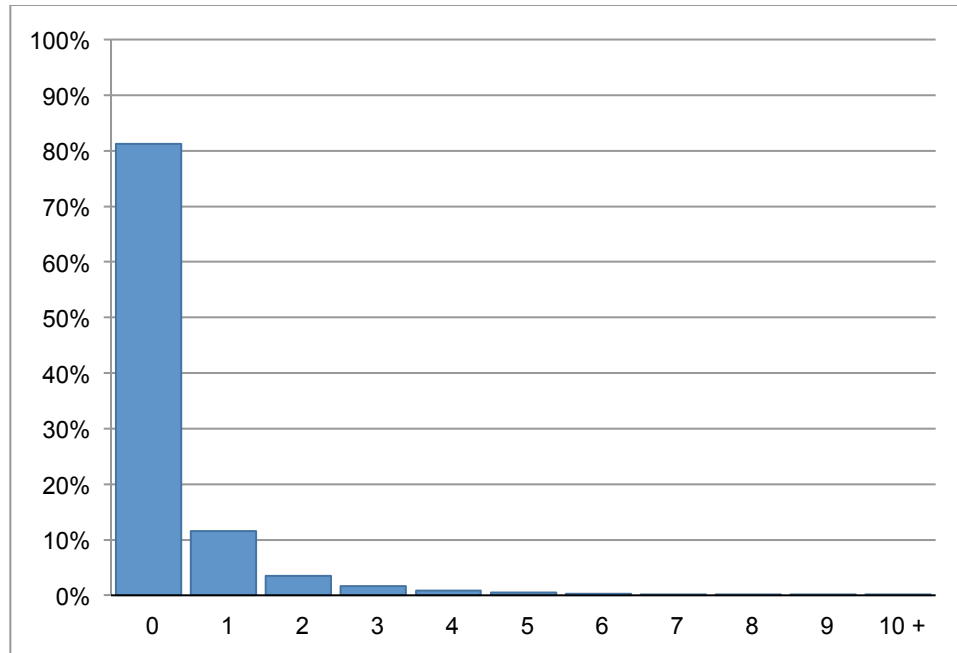


Figure 3 Distribution of offers received per month

In comparison to other types of recruitment, college football recruits are much more likely to receive multiple, simultaneous offers. Indeed, it would be difficult to imagine a job-seeker receiving thirty offers in the same month, which was the case for a few lucky 2016 recruits. Thus, in an effort to generalize this model, I perform a follow-up analysis shifting the problem from explaining the number of new offers to predicting whether an athlete will receive at least one new offer in the next month. To assess the value added by social media data for predicting offers, I construct two models, a baseline containing only recruiting data from 247Sports.com and a full model including the recruit’s level of Twitter activity, self-promotion, and ingratiation. I evaluate the models using stratified Monte Carlo cross validation. The data is randomly divided into two equal subsets with the same proportion of positive and negative instances, and I apply oversampling to balance the training data. The model is then tested on the hold-out data, tracking accuracy, AUC, precision, recall, and F1 scores over 100 trials.

2.3 Results

This study is intended to document the type of online impression management strategies employed by athletes during recruitment and examine the relationship between online impression management and offline recruiting outcomes. The following subsections report the results of applying the tweet classifiers to the full set of 5.5 million tweets, the fitted regressions modeling the number of tweets received over time, and the performance of the logistic classifier predicting whether an athlete will receive an offer in the next month.

2.3.1 Tweet Classification

As a preliminary test, I evaluate six different classification methods for the task of identifying self-promoting tweets. Table 6 displays the comparative performance of each classifier, averaged over 100 trials using different train/test partitions. Standard deviations are shown in italics next to each column.

Table 6 Tweet classifier testing results

Method	Accuracy		AUC		Precision		Recall		F1	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Logistic	0.866	<i>0.005</i>	0.792	<i>0.007</i>	0.711	<i>0.017</i>	0.660	<i>0.014</i>	0.684	<i>0.010</i>
Decision tree	0.859	<i>0.006</i>	0.762	<i>0.011</i>	0.719	<i>0.023</i>	0.588	<i>0.024</i>	0.647	<i>0.017</i>
Naïve Bayes	0.844	<i>0.005</i>	0.790	<i>0.007</i>	0.630	<i>0.015</i>	0.695	<i>0.015</i>	0.661	<i>0.010</i>
SVM	0.858	<i>0.006</i>	0.783	<i>0.007</i>	0.687	<i>0.019</i>	0.649	<i>0.015</i>	0.667	<i>0.011</i>
ANN	0.847	<i>0.008</i>	0.761	<i>0.012</i>	0.665	<i>0.028</i>	0.610	<i>0.028</i>	0.635	<i>0.016</i>
Random forest	0.873	<i>0.006</i>	0.774	<i>0.010</i>	0.773	<i>0.030</i>	0.598	<i>0.025</i>	0.673	<i>0.014</i>

These results indicate that logistic regression has the best overall performance, achieving the highest AUC and F1 scores. T-tests demonstrate that these differences are significant. The logistic classifier has a significantly AUC ($p < 2.2 \times 10^{-16}$ for decision tree, SVM, ANN, and random forest, $p = 0.02771$ for naïve Bayes) and significantly higher F1 score ($p < 2.2 \times 10^{-16}$ for decision tree, naïve Bayes, SVM, and ANN,

$p = 1.54 \times 10^{-9}$ for random forest) than all of the other methods. Additionally, it has the second-highest scores for accuracy and recall, and the third-highest precision. For this reason, I elect to use logistic regression to identify occurrences of self-promotion and ingratiation in the full set of 5.5 million tweets.

After applying the trained logistic classifier to all 5.5 million tweets, 1,065,735 (19.2%) are classified as self-promotion and 772,099 (13.9%) as ingratiation. These figures are consistent with the proportions of self-promoting and ingratiating tweets observed in the training set, 21.9% and 14.4% respectively. This initial analysis demonstrates that self-promotion—tweeting positively about one’s own athletic achievements, recruitment activities, or academic ability—is more prevalent than ingratiation—praising another athlete, coach, team, or school—in the tweets of high school football recruits.

2.3.2 Modeling Offers

In order to examine the relationship between number of offers received in the next month and the athlete’s personal characteristics, recruiting activities, and Twitter content to-date, I constructed four ZINB regression models trained on the full set of 26,576 athlete-month instances. Table 7 lists the regression coefficients and significance for each feature across the three models. Note that each model contains both a zero-inflation (logistic) and count (negative binomial) process, and that the zero-inflation process estimates the odds that an athlete will have zero offers in the next month. I also report the mean-squared error (MSE) and pseudo R-squared (McFadden, 1974).

Table 7 Fitted ZINB regressions

Feature	Model 0			Model 1			Model 2			Model 3						
	Zero-inflation		Count	Zero-inflation		Count	Zero-inflation		Count	Zero-inflation		Count				
intercept	-2.3477		-6.4151	◦	0.2292	-6.6281	◦	0.4807	-6.7479	◦	0.4331	-6.8612	◦			
2-star	-0.7638	**	0.3521	**	-0.7075	**	0.3355	*	-0.7074	**	0.3139	*	-0.6988	**	0.3321	*
3-star	-1.7799	***	0.5584	***	-1.7233	***	0.5131	***	-1.7224	***	0.5110	***	-1.7100	***	0.5149	***
4-star	-4.2098	***	0.8486	***	-4.1403	***	0.8168	***	-4.1928	***	0.8163	***	-4.1296	***	0.8304	***
5-star	-5.7239	**	1.0228	***	-5.3881	**	0.9511	***	-5.3378	**	1.0083	***	-5.3585	**	0.9904	***
height	0.1096		0.0684		0.0930		0.0641		0.0897		0.0679		0.0895		0.0673	
weight	-0.0275		-0.0061		-0.0264		-0.0046		-0.0255		-0.0051		-0.0258		-0.0051	
BMI	0.2327		0.0458		0.2225		0.0347		0.2159		0.0384		0.2177		0.0383	
MW	-1.6460		0.1552		-1.7105		0.3454		-1.7519		0.2044		-1.6852		0.3895	
NE	-1.3316		0.0539		-1.3731		0.2285		-1.3879		0.0931		-1.3408		0.2749	
S	-1.8624	◦	0.1940		-1.9352	◦	0.3604		-1.9622	◦	0.2330		-1.9007		0.4056	
W	-1.1037		-0.1621		-1.2049		0.0062		-1.2369		-0.1499		-1.1692		0.0348	
QB	-0.0895		-0.3524	***	-0.0245		-0.3635	***	-0.0241		-0.3632	***	-0.0212		-0.3675	***
updates	-0.1334		-0.1103	***	-0.1849		-0.1146	***	-0.1806		-0.1089	***	-0.1867		-0.1137	***
campus	0.0420		-0.0318	*	0.0273		-0.0314	*	0.0370		-0.0335	*	0.0311		-0.0313	*
unofficial	-0.5436	***	-0.0325	***	-0.4834	***	-0.0338	***	-0.4996	***	-0.0339	***	-0.4766	***	-0.0348	***
coach	-0.9021	**	-0.0634	***	-0.8292	**	-0.0600	**	-0.8607	**	-0.0654	***	-0.8426	**	-0.0635	***
offers	-0.7122	***	0.0193	***	-0.6439	***	0.0135	***	-0.6594	***	0.0158	***	-0.6399	***	0.0126	***
FBS	0.0068	***	0.0015	**	0.0055	***	0.0011	*	0.0057	***	0.0011	*	0.0054	***	0.0010	*
verbal	0.6984	**	-1.0169	***	0.6684	**	-1.0035	***	0.6344	*	-1.0273	***	0.6649	**	-1.0082	***
October	-0.6738	*	-0.1877		-0.6916	*	-0.2160		-0.6962	*	-0.2196		-0.6985	*	-0.2332	
November	-0.3938		-0.5271	**	-0.4373		-0.5688	***	-0.4452		-0.5764	***	-0.4504		-0.5889	***
December	-1.2896	***	-0.3387	*	-1.3132	***	-0.3867	*	-1.3170	***	-0.3865	*	-1.3107	***	-0.4004	**
January	-21.665	***	0.6100	***	-22.218	***	0.5848	***	-21.998	***	0.5775	***	-22.284	***	0.5708	***
February	-2.1916	***	1.0199	***	-2.2536	***	0.9689	***	-2.2423	***	0.9517	***	-2.2519	***	0.9453	***
March	-2.2492	***	0.7001	***	-2.3458	***	0.6137	***	-2.3245	***	0.6137	***	-2.3450	***	0.5855	***
April	-2.3551	***	0.5063	***	-2.4508	***	0.4254	***	-2.4406	***	0.4234	**	-2.4508	***	0.4003	**
May	-2.7365	***	1.0531	***	-2.8406	***	0.9866	***	-2.8346	***	0.9685	***	-2.8323	***	0.9612	***
June	-3.4121	***	-0.0590	***	-3.5956	***	-0.1339	***	-3.6044	***	-0.1582	***	-3.6003	***	-0.1678	***
July	-4.0958	***	-1.6409	***	-4.2839	***	-1.6840	***	-4.3188	***	-1.7233	***	-4.2567	***	-1.7128	***
August	-1.0976	*	-1.4158	***	-1.2122	*	-1.4309	***	-1.2116	*	-1.4640	***	-1.2082	*	-1.4391	***
low				***	-0.9688	***	0.0782	***	-1.0344	***	0.1498	***	-0.9427	***	-0.0199	***
mid				***	-1.1912	***	0.2580	*	-1.2220	***	0.3206	**	-1.1753	***	0.1773	***
high				***	-1.3014	***	0.2987	**	-1.3118	***	0.3535	**	-1.2843	***	0.2339	*
highest				***	-1.2759	***	0.4854	***	-1.2964	***	0.5277	***	-1.2610	***	0.4336	***
protected				***	-1.4818	***	0.3055	*	-1.4878	***	0.2960	*	-1.4706	***	0.3080	**
self-promotion				◦	-0.0046	◦	0.0073	***			-0.0040	***			0.0061	***
ingratiation								-0.0034		0.0076	***	-0.0018		0.0054	***	
Pseudo R-squared	0.183				0.190			0.189				0.191				
MSE	0.906				0.890			0.892				0.888				

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ◦ $p < 0.1$

Examining the zero-inflation portion of Model 0, we can estimate the marginal effect of each feature on the odds of receiving zero offers by applying the exponential function. Compared to recruits with 0 stars, the odds of receiving zero offers decrease by 53%, 83%, 99%, and 99.7% for 2-, 3-, 4-, and 5-star recruits, respectively. There is also an observable regional effect; compared to international recruits, athletes from the southern United States have 84% lower odds of receiving zero offers in the next month. Traditionally, the South believed to be a recruiting “hotbed,” and a previous analysis of scholarship offers supports provided evidence that athletes from Florida and Texas receive significantly more offers than athletes in other states (Pitts & Rezek, 2012). The athlete’s recruiting activities to-date were also related to the odds of receiving an offer. Each unofficial visit is associated with a 42% decrease in the odds of receiving zero offers and each coach visit with a 59% decrease. Interestingly, while the each previous offer is associated with a 51% lower odds of earning zero offers, the proportion of offers from FBS teams has the opposite impact. For each 1% increase in proportion, the odds of receiving zero new offers in the next month increase 0.7%. This result may be due to the competitive nature of recruitment. Intuitively, the more attention that a recruit receives from top programs, the less likely that he will consider a lower-rated program. Having a current verbal commitment doubles the odds that an athlete will receive zero offers in the next month. This model also accounts for the fixed effect of time. Compared to September of their junior year, recruits are much more likely to receive at least one offer during the months of December through July, and almost all recruits received at an offer in January.

The coefficients of the count model estimate the marginal effect of each feature on the log of the number of offers. The log number of offers received in the next month increase by star rating; 0.35, 0.55, 0.85, and 1.02 for 2-, 3-, 4-, and 5-star recruits, respectively. The log number of offers month decreases 0.35 when the recruit is a quarterback. This position effect is likely due to the fact that colleges tend to offer and secure commitments from quarterbacks early in the recruitment process. Notably, increased numbers of recruiting events are associated with receiving fewer offers. The log number of offers decreases 0.11, 0.03, 0.03, and 0.06 for each news update, camp, unofficial visit, and coach visit, respectively. For each current offer, the log number of offers increases 0.01, with an increase of 0.002 for each 1% increase in proportion of total offers from FBS teams. Having a current verbal commitment has a large effect on the number of offers received in the next month, decreasing the log number of offers by 1.02. As with the zero-inflation portion of the model, a significant time effect is observed. Compared to September 2014, the log number of offers increases from January 2015 through March 2015.

Model 1 incorporates data about the athlete's online self-promotion into the zero-inflation and count processes. Compared to recruits without any Twitter account, having a protected account is associated with a 78% decrease in the odds of receiving zero offers in the next month. Recruit with low (1-18 tweets), mid (19-52 tweets), high (53-134 tweets) and very high (>135 tweets) activity during the prior month had 62%-73% lower odds. Each 1% increase in proportion of self-promoting tweets posted in the prior month decreases the odds of receiving zero offers by 0.5%. Examining the effect on number of offers, mid, high, and very high levels of Twitter activity increases the log number of

offers by 0.26, 0.30, and 0.49, respectively. Having a protected account is associated with 0.31 more log offers than having no Twitter at all. Each 1% increase in proportion of self-promoting tweets is associated with a 0.007 additional log offers. These results suggest that any level of Twitter activity offers a significant advantage in terms of gaining coaches' attention and earning scholarship offers over, and that self-promotion is positively related to recruiting success.

Model 2 considers the impact of ingratiating behavior on the likelihood of receiving an offer in the next month and the number of offers. Again, the odds of receiving zero offers are much lower for recruits with public (64-73% less) or protected Twitter accounts (77% less), as opposed to those without Twitter. The effect of ingratiation was not significant on the zero-inflation process, although its coefficient shows the expected sign. Compared to recruits without Twitter, having a protected Twitter account is associated with a 0.30 increase in the log number of offers. Recruits with different levels of public Twitter activity received between 0.15 and 0.53 additional log offers. Each 1% increase in proportion of ingratiating tweets posted in the prior month is associated with 0.008 additional log offers.

In the combined model (Model 3), there is again a significant effect for Twitter activity. In comparison to recruits without a Twitter presence, the odds of receiving zero offers are much lower for recruits with any level of public activity (61-72% less) or protected Twitter accounts (77% less). While neither self-promotion nor ingratiation have a significant effect on the zero-inflation process, both have positive and significant effects on the number of offers received in the next month. A 1% increase in self-promoting tweets is associated with receiving 0.0061 additional log offers. The marginal effect of

ingratiation was slightly smaller; each 1% increase in ingratiating tweets posted in the prior month increases the log number of offers received by 0.0054. This result further suggests that self-promotion may have a stronger relationship to recruiting success than ingratiation. While the marginal effects on the count process are very small for the two impression management strategies, both self-promotion and ingratiation are highly significant.

I also compare the models based on measures of fit. While differences in pseudo R-squared are small, all of the models incorporating social media data achieve higher pseudo R-squared values than the baseline containing only recruiting data. Furthermore, the self-promotion model (Model 1) has a slightly higher pseudo R-squared value than the ingratiation model (Model 2), suggesting that self-promotion may explain more of the variation in number of offers over time than ingratiation, though Model 3 has the highest pseudo R-squared value overall. The combined model also achieves the lowest the mean-squared error (MSE), 0.888. The distinguishability and goodness of fit of the four models are also compared using a Vuong test (Vuong, 1989). This test calculates whether models equally approximate the true data generating process, against the alternative that one model is closer. Comparing Model 0 and Model 1, the Vuong test determines that Model 1 is significantly closer ($p < 2.22 \times 10^{-16}$), indicating that adding features tracking online self-promotion improves the fit of the model. Similar for Model 0 vs. Model 2, considering ingratiation improves the model ($p < 4.30 \times 10^{-15}$). Testing Model 1 against Model 2, there is a marginally significant difference in favor of Model 1 ($p = 0.06$), further suggesting that self-promotion is more valuable to an athlete's recruiting prospects than ingratiation. Additionally, the test indicates that a model incorporating

both self-promotion and ingratiation is a significant improvement on the baseline containing only recruiting information ($p = 2.22 \times 10^{-16}$) as well as Model 1, including just self-promotion ($p < 0.008$), and Model 2, containing only ingratiation ($p = 5.16 \times 10^{-5}$).

2.3.3 Predicting Offers

In an effort to generalize the dynamic offer model to the job-seeking context, I constructed a logistic classifier that predicts whether an athlete will receive at least one offer in the next month based on personal characteristics, recruiting activities, and online impression management up to that point in time. Table 8 displays the predictive performance of the baseline and full models, averaged over 100 trials. Standard deviations are shown in italics next to each column.

Table 8 Offer classifier results

Model	Accuracy		AUC		Precision		Recall		F1	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Baseline	0.730	<i>0.004</i>	0.749	<i>0.004</i>	0.399	<i>0.005</i>	0.780	<i>0.010</i>	0.528	<i>0.005</i>
Full	0.741	<i>0.004</i>	0.756	<i>0.003</i>	0.411	<i>0.005</i>	0.782	<i>0.008</i>	0.539	<i>0.004</i>

When evaluated on a hold-out testing set, a logistic classifier that includes social media data can predict whether a recruit will receive an offer in the next month with 74% accuracy and achieves an AUC score of 0.756. Indeed, the full model consistently outperforms the baseline containing only personal characteristics and recruiting data. While the improvements are not large in practice (differences of 0.3% to 2.1%), paired T-tests demonstrate that considering Twitter activity and impression management adds significant value to the predictions in terms of accuracy ($p < 2.2 \times 10^{-16}$), AUC ($p < 2.2 \times 10^{-16}$), recall ($p < 2.71 \times 10^{-7}$), precision ($p < 2.2 \times 10^{-16}$), and F1 score

($p < 2.2 \times 10^{-16}$). These results indicate that the athlete's behavior on Twitter during recruitment can be useful for predicting the scholarship decisions of college coaches.

2.4 Discussion

This study examines the use of online social media by college football recruits as a platform to engage in impression management and analyzes the relationship between self-presentation on Twitter and the number of scholarship offers received 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 and to use observational data to relate online impression management to offline recruiting outcomes. My work has significant implications for college football recruits, and can inform their communications and branding efforts during recruitment.

I find that athletes utilize different impression management strategies during recruitment, but tend to engage in more self-promotion than ingratiation. In both the hand-labeled subset (7,000 tweets) and the full set of tweets (5.5 million tweets), recruits posted approximately 50% more self-promoting than ingratiating tweets. This finding is consistent with previous studies examining impression management during in-person employment interviews (e.g., Stevens & Kristof, 1995), supporting the application of established theory from personnel recruitment to the athletic context. As the first large-scale study to document the use of social media by college football recruits, this work contributes to the information science literature by describing this population's use of social media for information creation, dissemination, and identity-formation.

Extending previous research on scholarship offers in college football (Pitts & Rezek, 2012), I model the relationship between the number of offers received in the next month and the athlete's personal characteristics and recruiting activities up to that point in time, as well as Twitter activity and online impression management during the prior month. I find that both self-promotion and ingratiation have positive and significant effects on recruiting success. Both Models 1 and 2, tracking self-promotion and ingratiation, respectively, were a significantly better fit than the baseline Model 0, containing only demographic and recruiting data from 247Sports. This result indicates that considering impression management on Twitter adds value to a model of scholarship offers, and suggests that athletes' online actions may have offline consequences during recruitment. I also performed a follow-up analysis, constructing a logistic classifier to predicting whether an athlete would receive any offers in the next month. The full model incorporating social media data achieved an AUC of 0.756, a significant improvement over the baseline with only recruiting data, and both self-promotion and ingratiation were significant predictors.

Yet the two impression management strategies do not necessarily have the same impact on offers. Model 1, which tracks self-promotion, had a lower MSE and was a better fit for the data than Model 2, focusing on ingratiation. Furthermore, the regression coefficients for the combined model (Model 3) suggest that the proportion of self-promoting tweets posted during the prior month may have a larger effect on the number of new offers than ingratiating tweets. Previous studies of impression management in HR recruiting have found similar results; self-focused tactics resulted in higher interviewer ratings (Stevens & Kristof, 1995) and fewer rejections (Kacmar et al., 1992) than other-

focused tactics. The larger marginal effect of self-promotion on scholarship offers may be related to the fact that coaches tend to prioritize characteristics that are closely related to on-field success, such as physical features and performance statistics, over those that reflect academic ability or good character (Pitts & Rezek, 2012). Because self-promotion is intended to increase perceptions of competence, it may be weighed more heavily by coaches focused on future performance outcomes. Conversely, ingratiation, which is meant to increase liking, is likely to have little impact on coaches' perceptions of athletes' abilities and thus, their likelihood of extending a scholarship offer.

These findings are interesting in light of evidence that self-promotion is increasing during the recruiting process (Sander, 2008). Football has the highest number of participants of any high school sport (NCAA, 2015), and even though recruits' performance statistics are readily available on platforms like 247Sports and Rivals, social media offers a means to create and share information that helps them stand out from the crowd. Another interesting feature of self-promotion is that its effects tend to be fairly universal. That is, posting tweets boasting about one's achievements may impress many coaches at once, while posting positive material about a specific college team would most likely only increase the chances of getting an offer from that team. This is not to say that athletes should stop engaging in ingratiation through social media. Ingratiation—which I defined broadly as content promoting or praising any other athlete, coach, team, or school—had a positive and significant effect. Ingratiation often signals the presence of other positive traits, such as maturity and good character. Furthermore, because direct contact between college coaches and high school players is highly regulated (NCAA,

2016), it may be difficult for coaches to gauge such “intangibles,” and social media can provide a vital source for such information.

It should be noted that this study focused on forms of positive self-presentation on social media, while mass media reports and surveys of college coaches (Caldwell, 2015; Crabtree, 2016; Safian, 2016) have almost exclusively highlighted cases of online misbehavior and its negative impacts on offline recruiting outcomes. It is possible that exhibiting bad character traits through racist, sexist, vulgar, or aggressive content could have a much larger effect on scholarship offers than the good character traits (e.g., humility, sportsmanship) suggested by ingratiation. As stated by longtime Florida Gulf Coast men’s basketball coach, Michael Fly, “We’ve actually stopped recruiting kids in the past because of something they put on Twitter. I thought they were really good players, but...we shut it down” (as cited in Caldwell, 2015). However, both anecdotal evidence from interviews with college coaches and my own analysis, where only 261 (3.7%) of the 7,000 hand-labeled tweets were determined to be potentially inappropriate or offensive, suggest that athletes are becoming increasingly adept at using social media as a tool for branding and positive self-presentation during recruitment.

Perhaps more interesting than the differential effects of self-promotion and ingratiation was the overall impact of social media usage on recruiting success. Athletes’ levels of social media engagement varied widely; from those with no Twitter accounts, to athletes who posted rarely (24% of recruits with public accounts posted fewer than 10 times per month during junior year), to those who posted almost incessantly (6% of recruits with public accounts posted more than 10 times per day during junior year), to those with protected accounts whose tweets could not be collected using the Twitter API.

To account for this variation, I also considered the effect of each athlete's Twitter status (none, protected) and level of activity based on number of tweets per month (low, mid, high, highest). ZINB regression coefficients indicate that a possessing public social media presence is a significant improvement over no Twitter account. Averaged across Models 1-3, having low Twitter activity (1-18 tweets per month) was associated with 63% lower odds of receiving zero offers in the next month, with a 70% decrease for recruits with a moderate level of activity (19-52 tweets). A high level of Twitter activity (53-134 tweets per month) was associated with a 72% decrease on average and the highest activity level (135 or more tweets) with a 73% decrease in the odds of receiving zero offers. Similar results were achieved with the predictive model, where, in general, more Twitter activity was associated with greater chances of receiving an offer.

Furthermore, even a protected (non-public) Twitter account represented a significant advantage in terms of both likelihood of receiving an offer and the number of offers. Averaging the ZINB regression coefficients across Models 1-3, having a protected account was associated with a 77% decrease in the odds of receiving zero offers in the next month and 0.3 additional log offers in the next month. Having a protected account was a significant factor in the predictive model as well. These findings are notable, as they contradict previous statements by coaches expressing suspicion in regards to protected social media accounts, "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!" (as cited in Safian, 2016, p. 8). At least in the context of college football, where one of the primary benefits of Twitter is the ability to use direct messages to communicate privately without violating NCAA regulations, it is

almost a foregone conclusion that coaches and recruits would mutually follow each other, and thus, be able to view protected tweets. My results indicate that having any type of online presence represents a significant benefit to athletes during recruitment. Furthermore, the features tracking account status and activity level explained approximately 8% of the variation in number of offers, more than the proportion of self-promoting tweets and ingratiating tweets combined.

The results of this study are noteworthy, as they signal that the information created and shared by recruits on social media may influence the scholarship offer decisions of college football coaches. In contrast, Chapters 3 and 4 discuss how recruits' online connections and content can be used to forecast their own commitment and decommitment decisions. This work represents a promising next step in explaining and predicting scholarship offers in college football and provides a framework for using candidates' online behaviors to predict offline recruiting outcomes in other contexts. Overall, this study makes a unique contribution to the larger conversation in the field of information science about the uses and impacts of social media.

CHAPTER 3 COMMITMENTS²

The explosive growth of social media over the last decade has resulted in the ready availability of data about individuals' connections and communications in online social networks, providing new opportunities to study how people make decisions. In this study, I explore how the information created on Twitter by college football recruits can provide insight into their commitment decisions.

Human decisions are rarely entirely logical. However, both job choice and school choice have largely been approached using rational decision-making models, presuming that candidates will select the organization based on objective factors that maximize their expected utility. Because these decisions occur under significant constraints in terms of time, information, and cognitive resources, a quasi-rational or bounded rationality model including social factors and heuristics may be more appropriate (Highhouse, 1997). Social media presents a unique opportunity to study how people make decisions in light of these considerations. In addition to providing new avenues for users to network, influence, and be influenced by others, it enables researchers to gather large-scale, detailed data about individuals' behaviors, preferences, and online connections. While a few previous studies have examined the impact of candidates' offline friends and coworkers on job choice behavior (e.g., Kilduff, 1990; Tenbrunsel & Diekmann, 2002), there remains a need for more research on the complex interactions between online social media and offline recruiting outcomes.

² A version of this study was selected for the Special Issue on Decision Analysis and Social Media of *Decision Analysis* and published as 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.

While obtaining private personnel data is often difficult, and may not paint a full picture of trends across an entire labor market, I take a unique approach to studying the relationship between connections and content in online social networks and offline recruiting outcomes by using college football recruiting data. I propose two main research questions:

Q1: What institutional characteristics and recruiting activities predict school choice?

I extend previous work on athletic recruiting (Dumond et al., 2008) that identified rational, cost/benefit factors in school choice by examining the role of decision-making heuristics. This study is the first to investigate how the sequence of recruiting activities influences school choice via availability and saliency. I combine data about the individual recruit, his choice set (the colleges that have offered a scholarship), and his recruiting activities to construct a baseline model predicting which college a recruit will commit to from among those that have offered him a scholarship.

Q2: What value does social media data add to school choice predictions?

While both anecdotal and empirical evidence indicates that parents, high school coaches, and other athletes play an important role in school choice decisions (e.g., Croft, 2008), previous work predicting college football commitments has either ignored these social factors (Dumond et al., 2008; Klenosky, Templin, & Troutman, 2001) or focused only on family connections (Mirabile & Witte, 2015). I consider how the recruit's online social network reflects the closeness of the athlete-school relationship (Jowett & Timson-Katchis, 2005) as well as how the information created online by athletes and schools can be interpreted as signals of interest (Spence, 1973). This work is the first to incorporate social media data into a model of school choice. In order to evaluate the utility of

considering Twitter data, I build several models using different sets of features and compare their performance to the baseline using only recruiting data.

Overall, my analysis of the commitment decisions of recruits in the class of 2016 indicate that athletes weigh a variety of factors when making their college choice. While objective, cost/benefit factors, including proximity, academic ranking, and the availability of alternatives, do a fair job explaining and predicting school choice, the results indicate that a quasi-rational model considering cognitive biases and social factors may be more appropriate. Model 0, which includes institutional data, recruiting activities, and features related to the availability heuristic achieves a 3% improvement in AUC over the purely rational model. This study also provides evidence supporting the utility of social media data for predicting commitments. Model 5, which combines cost/benefit factors, the sequence of recruiting events, and social media features tracking the recruit's online social network and Twitter hashtags, achieves the highest predictive performance. This research can help coaches effectively allocate recruiting resources by identifying and prioritizing athletes who are most likely to commit, and the findings suggest that online social media data may be useful for predicting offline recruiting outcomes in other contexts.

3.1 Related Research

This study is the first to examine the use of social media data for predicting school choice in college football. In addition to contributing to the growing body of sports analytics research, this work has more general implications about the role of social factors and heuristics in recruiting decisions, and I review related literature from the fields of organizational management, personnel psychology, and information science.

There is a large body of literature devoted to studying predictors of candidate attitudes and job intentions, including organizational and job characteristics (e.g., Chapman, Uggerslev, Carroll, Piasentin, & Jones, 2005) and corporate reputation (e.g., Cable & Turban, 2003). These works suggest that choosing to accept a job offer is a complex decision, with multiple, conflicting objectives. Auger, Devinney, Dowling, Eckert, and Lin (2013) conceptualized the decision-making process using a compensatory model (e.g., an attractive salary may compensate for an unattractive location). In an experiment requiring subjects to make “trade offs,” the authors found that salary, time demands, and promotion opportunities were the most influential features for job choice among MBA students.

Like most human decisions, selecting which job offer to accept is unlikely to be an entirely rational process. Yet the extant literature has primarily applied rational decision models, overlooking heuristics and biases (Rynes et al., 1980). Of the studies relating judgment and decision-making theory to recruitment outcomes, most have focused on the employer perspective, examining the effects of cognitive biases in selection (Highhouse, 1997) and performance evaluations (Wong & Kwong, 2005). Looking at the candidate decision-making process, Crossley and Highhouse (2004) surveyed 204 employees and 503 undergraduate students about their job search techniques, choice process, and level of satisfaction. The authors found that engaging in a focused, as opposed to haphazard, search process led to higher job satisfaction, but that there was no difference in satisfaction between individuals who utilized a rational decision-making process and those who relied on intuition or gut feeling. While their findings support the use of a quasi-rational approach to model candidate decision-

making, there are no previous empirical studies applying these models to predict actual job choices.

Individuals lacking time, information, and cognitive resources may also turn to their social networks when making job decisions (Parnes, 1954; Rynes et al., 1980). The HR research literature has paid a significant amount of attention to social networks in job-seeking and job choice (e.g., Granovetter, 1973). Social ties between a candidate and organization can increase applicant attraction and job choice intentions by providing unique and credible information (McManus & Baratta 1992; Vecchio, 1995) and enhancing perceived person-organization fit (Chapman et al., 2005). Other studies have examined the effects of recruiter personality (Alderfer & McCord, 1970; Hilgert & Eason, 1968), perceived similarity (Wyse, 1972), and behavior during the interview process (Downs, 1969; Schmitt & Coyle, 1976) on job choice intentions. Proposed by Leon Festinger (1954), social comparison theory holds that, in the absence of an objective standard, decision-making is informed by the context of “similar others.” Applying this theory to the job-seeking process, Kilduff (1990) studied of the interview bidding process of 209 MBA students. The author found that there was significant overlap in bidding behavior between students with perceived similarity. While these works highlight the potential influence of recruiters and colleagues on candidates’ decisions during recruitment, the impact of online social connections remains to be explored. My study extends this prior research in the HR domain by utilizing social media data, predicting the commitment decisions of college football recruits based on their recruiting activities, online social networks, and Twitter content.

Like job candidates, student-athletes weigh many factors when selecting a college. Surveys of college athletes have identified scholarship amounts (Doyle & Gaeth, 1990), geographic proximity (Barden, Bluhm, Mitchell, & Lee, 2013; Lujan, 2010), facility quality (Croft, 2008), academic rigor (Popp, Pierce, & Hums, 2011), team performance (Sander, 2008), and opportunities for postseason play (Croft, 2008) as highly influential factors in commitment decisions. Previous work analyzing and predicting school choice has focused on rational models, expecting that an athlete will seek to maximize the expected utility of attendance with respect to school choice (Dumond et al., 2008; Klenosky et al., 2001).

Recruits make college choices under significant constraints in terms of time, information, and cognitive resources. Because colleges can only award a maximum of 25 scholarships to incoming freshmen (NCAA, 2016), recruits 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. The father of a class of 2016 quarterback estimated spending \$40,000 on travel expenses for camps, combines, and unofficial visits, and the guardian of a 360-pound, 4-star defensive lineman described \$700 monthly grocery bills (Elliott, 2015). Visits represent one of the most important sources of information for athletes selecting a college; 46% of respondents to Sander's (2008) survey went on at least three unofficial visits. Although the NCAA allows unlimited unofficial visits, the financial burden on the recruit's family may make it impossible to visit each college that is recruiting him. Athletes can also take official visits paid for by the college, but are only allowed a maximum of five (NCAA, 2016).

Additionally, athletes in the class of 2016 could not take official visits until fall of their senior year, a rule that was recently relaxed to allow spring visits during junior year.

Gathering information about college options by communicating with coaches is also fraught with difficulties due to NCAA regulations on the timing, frequency, and medium of contact between college coaches and recruits. The NCAA divides the recruiting calendar into alternating periods of “quiet,” when recruits may make visits to colleges, but coaches may not visit recruits off-campus, “contact,” when coaches are allowed to visit with recruits and their families, “evaluation,” when only visits to a recruit’s high school are permitted, and “dead,” when only telephone and written communication are allowed. Because of these restrictions, it is possible for athletes to receive (and accept) scholarship offers without ever meeting their prospective coach. In a study of 98 high school athletes, only 18.4% of those intending 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 to recruits, 65% of high school athletes reporting an intention to play in college had spent little or no time researching colleges (Lujan, 2010).

While coaches may begin sending materials to a recruit during his junior year of high school, the first “contact” period does not occur until November of the athlete’s senior year (NCAA, 2016). Thus, the most intense periods of recruitment directly conflict with the school year and high school football season, negatively impacting the cognitive resources available to the recruit. Furthermore, a large body of literature explores age-specific differences in psychology and decision-making (e.g., Steinberg & Cauffman, 1996), suggesting that adolescents are likely to deviate from rational decision-making.

Despite these constraints, no previous research on athletic recruiting has examined the role of cognitive biases and heuristics in the school choice process.

Anecdotal and empirical evidence indicates that parents (Croft, 2008), high school coaches (Prunty, 2014), and other recruits (Myerberg, 2015) play an important role in school choice decisions. Relationships with college coaches are also significant (Forsythe, 2015; Treadway, Adams, Hanes, Perrewe, Magnusen, & Ferris, 2014), with 42% of respondents in a survey of 300 Division I athletes rating the reputation of the head coach “extremely important” (Sander, 2008). As stated in a study of commitment timing by Bricker and Hanson, “Once a coach identifies a talented player, they will try to develop a relationship with the player in order to entice them to accept a scholarship offer. The coach will make frequent phone calls and will travel to talk to the family, friends, and coaches of the recruit” (2013, p. 972). However, only one previous study has used athletes’ social networks to predict their school choice decisions. Mirabile & Witte (2015) identified family connections between recruits and colleges, finding that having a family member who played or coached at a college increased the likelihood of commitment between 96% and 253%. My work represents a unique addition to the athletic recruiting literature, and is the first to utilize social media data to observe recruits’ ties to the schools that are recruiting them.

Previous research has demonstrated the power of online social network data to predict offline outcomes. For example, Bogaert et al. (2015) found that the behavior of one’s Facebook friends strongly predicted event attendance. The information created by individuals through other types of social media interactions can provide insight into their preferences and decisions. Kristensen et al.’s (2017) analysis of Facebook data found that

likes mirrored actual voting intentions in a Danish parliamentary election. My study makes a novel contribution to the field of information science by examining how the information that individuals create via their online connections, interactions, and content reveals can be used to understand and predict their decisions and the decisions of others.

3.2 Methods

In this study, I seek to predict which college a recruit will commit to out of all those that have offered him a scholarship. Focusing on the question of *where* the recruit will commit, rather than *when*, I assume that the month of commitment is known, and build a static model that considers recruiting and social media data up to that point.

3.2.1 Data

As described in Chapter 2, I began with the full set of 2,644 recruits in the class of 2016 whose recruiting data was obtained from 247Sports.com. For the purposes of this project, I considered both verbal commitments and signings. I identified commitments from athletes' recruiting data in two ways: (1) from commitment announcements tracked in a player's 247Sports timeline, and (2) from NLI signings (as of 13 March 2016). Out of the full dataset, 2,277 unique athletes (86%) committed at some point during recruitment. Since a recruit may commit and decommit multiple times, some athletes in the class of 2016 had commitments to more than one college.

Although an athlete can verbally commit at any time during recruitment, the analysis includes only athletes that committed 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 recruit. There were 25 commitments after March 1, 2016 (1.1% of all commitments) that were

not included 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 decision-making process of recruits with late commitments is likely to be different than those who commit before National Signing Day. Because the focus of this work is predicting which college a recruit will commit to out of those that have offered scholarships, I also eliminated 705 individuals without multiple offers. 93% of recruits with a single offer ultimately committed to that college. Furthermore, my analysis was limited to athletes with public Twitter accounts. I contend that this does not introduce bias because there is no evidence of significant differences between athletes with and without social media. Though Chapter 2 demonstrates that an athlete’s level of Twitter activity significantly predicts whether he will receive an offer in the next month, Chi-squared tests fail to reject the null hypothesis of independence between simply possessing a public Twitter account and star rating ($p = 0.9912$) or the total number of offers received ($p = 0.394$).

These steps reduce the dataset to 573 recruits who selected a college from among eight scholarship offers on average. I created a set of 4,408 “athlete-school” pairs; for each recruit, there is an instance for every college that has offered him a scholarship. Therefore, the prediction dataset has multiple instances corresponding to the same recruit. For each athlete-school pair in the data, I measured features relative to the focal school (termed the “prediction school”) in comparison to the others that have offered a scholarship.

3.2.2 Modeling the School Choice Decision

College football recruiting can be categorized into two sequential decision-making stages. In the first stage, college coaches identify and evaluate potential recruits and decide whether to extend a scholarship offer. In the second, athletes select from among the scholarship offers and announce a commitment. This study focuses on this latter stage, building an explanatory model to identify factors predicting school choice decisions and applying the model to predict commitments in the college football recruiting class of 2016.

The decision of which college to attend is an important one, affecting more than the next four to five years of the athlete's life. 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, 2015). Of those who do advance to the college level, only 1.6% will be drafted into the NFL (NCAA, 2015). School choice also has major impacts on the recruit's career prospects outside of professional athletics, his social life, and overall wellbeing. Based on previous research on the school choice process of student-athletes (e.g., Croft, 2008; Popp et al., 2011) and the reality that recruits make commitment decisions under significant constraints, I propose a quasi-rational approach. I assume that recruits' school choices will be influenced by cognitive biases and social networks in addition to objective, cost/benefit factors. Thus, the proposed model of school choice includes several sets of features.

First, I construct a set of baseline features derived from the 247Sports recruiting data and institutional data. Because the school choice process may vary by individual (Doyle & Gaeth, 1990; Gabert, Hale, & Montalvo, 1999; Huffman & Cooper, 2012), I

create features tracking the athlete's demographic characteristics, including home state and star rating. Previous studies of college football commitments have identified economic capital, athletic capital, and human capital objectives in the school choice process (Dumond et al., 2008), but these fundamental objectives may be difficult to express and measure. I construct baseline features to represent means objectives, i.e., observable, intervening factors that are related to the fundamental objective (Huynh & Simon, 2016). For instance, a college's *U.S. News & World Report* ranking may impact the fundamental objective of increasing human capital through perceived educational quality. I expect that features that increase the benefits associated with attendance at a given college will also increase likelihood of commitment. I extend previous work by considering these objectives in comparison alternative options in the athlete's choice set. An athlete's likelihood of selecting the prediction school may be influenced not only by its geographic proximity, but also by the number of other colleges recruiting him that are closer. I consider offline recruiting activities that demonstrate affinity between a college and recruit (e.g., visits).

In light of the constraints on time, information, and cognitive resources faced by recruits, I also consider the role of the availability heuristic in school choice. The availability heuristic holds that decision-makers will select the most memorable option (Tversky & Kahnemann, 1973). Availability may be influenced by several factors, including sequence, frequency, and vividness. Because the recruiting data from 247Sports includes timelines of recruiting events, I focus on the relationship between school choice and sequence, tracking the first and last events in each series (e.g., first offer, last offer).

In total, I create and test 44 baseline features (Table 9). These features are time-consistent, including only events that occurred prior to the commitment decision. For example, to predict which college a recruit will commit to in January, I count only official visits that occurred before January 1.

Table 9 Baseline commitment features

Type	Feature	Description
<i>Demographic</i>	star	Recruit star rating (0, 2, 3, 4, 5)
	height	Height (inches)
	weight	Weight (pounds)
	BMI	Body mass index ($weight/height^2$) $\times 703$
	position	Recruit position (ATH, DB, DL, LB, OB, QB, RB, ST, WR)
	region	U.S. Census region of recruit (INT, MW, NE, S, W)
	hotbed	1 if hometown located in recruiting hotbed states of CA, FL, or TX, 0 otherwise
<i>Objectives</i>	distance	Distance between recruit hometown and school being predicted (miles)
	in_state	1 if recruit hometown in same state as prediction school, 0 otherwise
	type	Institutional type of prediction school (military, private, public)
	us_news	Prediction school academic ranking in U.S. News
	division	Athletic division of prediction school (FBS, FCS, II, III, JUCO)
	power	1 if prediction school is a member of Power 5 conference, 0 otherwise
	AP	1 if prediction school ranked in top 25 of AP poll, 0 otherwise
	postseason	1 if prediction school played in bowl game during 2014 season, 0 otherwise
	historic_win	Winning percentage of prediction school over past five seasons (2010-2014)
	current_win	Winning percentage of prediction school during 2015 season
	commits	Number of other recruits verbally committed to prediction school
	coach_change	1 if head coach change at prediction school during 2015 season, 0 otherwise
	sanctions	1 if football program under active NCAA sanction or probation, 0 otherwise
	other_offer	Number of offers from schools other than prediction school
	other_closer	Number of offering schools closer than prediction school
	other_us_news	Number of offering schools with higher U.S. News ranking
	other_power	Number of offering schools from Power 5 conferences
	other_AP	Number of offering schools with higher AP poll ranking
	other_postseason	Number of offering schools that played in bowl game during 2014 season
	other_historic_win	Number of offering schools with greater winning percentage over last five seasons
other_current_win	Number of offering schools with greater winning percentage during 2015 season	
other_coach_change	Number of offering schools with head coach change during 2015 season	
other_sanctions	Number of offering schools under NCAA sanction or probation	

Table 9 (cont.)

<i>Affinity</i>	unofficial	Number of unofficial visits to prediction school
	coach_visit	Number of coach visits from prediction school
	official	1 if recruit has taken official visit to prediction school, 0 otherwise
<i>Heuristics</i>	other_unofficial	Number of unofficial visits to other schools
	other_coach_visit	Number of coach visits from other schools
	other_official	Number of official visits to other schools
	first_offer	1 if first offer from prediction school, 0 otherwise
	last_offer	1 if last offer from prediction school, 0 otherwise
	first_unofficial	1 if first unofficial visit to prediction school, 0 otherwise
	last_unofficial	1 if last unofficial visit to prediction school, 0 otherwise
	first_coach	1 if first coach visit from prediction school, 0 otherwise
	last_coach	1 if last coach visit from prediction school, 0 otherwise
	first_official	1 if first official visit to prediction school, 0 otherwise
last_official	1 if last official visit to prediction school, 0 otherwise	

This study also explores how the recruit’s social media activities reveal information about his preferences that can be used to predict school choice. The next set of features tracks the users followed by the recruit on social media during the month prior to commitment, i.e., “friends” or out-links. I expect that recruits intending to commit to a given college will add friends from that college. This hypothesis 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 & Timson-Katchis, 2005). Indeed, 62% of Division I athletes report building friendships with their future teammates during recruitment (Sander, 2008). Because the Twitter REST API (Twitter, 2016) does not return information on the date that a connection is formed, friend lists retrieved on the first days of the commitment month and previous month are compared to determine the number, type (coach, recruit, college player), and affiliation of new friends. I construct 6 features related to Twitter out-links (Table 10).

Table 10 Social media commitment features

Type	Feature	Description
<i>Out-links</i>	coach_friends	Number of coaches from prediction school followed by user over prior month
	recruit_friends	Number of 2016 recruits committed to prediction school followed by user over prior month
	college_friends	Number of current football players at prediction school followed by user over prior month
	coach_friends_other	Number of coaches from other schools followed by user over prior month
	recruit_friends_other	Number of 2016 recruits committed to other schools followed by user over prior month
	college_friends_other	Number of current players at other schools followed by user over prior month
<i>In-links</i>	coach_followers	Number of coaches from prediction school following user over prior month
	recruit_followers	Number of 2016 recruits committed to prediction school following user over prior month
	college_followers	Number of current players at prediction school following user over prior month
	coach_followers_other	Number of coaches from other schools following user over prior month
	recruit_followers_other	Number of 2016 recruits committed to other schools following user over prior month
	college_followers_other	Number of current players at other schools following user over prior month
<i>Interactions</i>	interaction	1 if user posted retweet, reply, quote, or mention associated with prediction school over prior month, 0 otherwise
	interaction_other	1 if user posted retweet, reply, quote, or mention associated with other schools over prior month, 0 otherwise
<i>Content</i>	hashtag	1 if user posted hashtag associated with prediction school over prior month, 0 otherwise
	hashtag_other	1 if user posted hashtag associated with other schools over prior month, 0 otherwise

The next group of features focuses on in-links from other Twitter users, i.e., “followers.” Mass media reports indicate that college teams actively utilize social media to cement connections with recruits (Myerberg, 2015; Rutter, 2015). Conversely, the team risks turning off athletes by not investing in relationship-building. 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” (as cited in Davenport, 2015). In addition, NCAA limits on athlete-coach communication during recruitment have created significant information asymmetries (Bricker & Hanson, 2013). Thus, connections on

social media may take on an even greater importance. Signaling theory (Spence, 1973) predicts that recruits and coaches will engage in activities intended to communicate their relative levels of interest, and I interpret these in-links as signals of the college's interest in the athlete. Similar to the features tracking Twitter friends, I compare follower lists over time to determine the number and affiliation of new followers in the month prior to commitment. I construct 6 features based upon athletes' in-links (Table 10).

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. In addition to friends and followers, I consider interactions between recruits and colleges on social media. Twitter allows users to interact in several ways: replying to other users' posts, copying posts, forwarding posts, and mentioning other users. I hypothesize that interactions with recruits, coaches, and current athletes associated with the prediction school signal the athlete's college preferences and will increase the likelihood of commitment. Because 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 (NCAA, 2016), I only track the social network interactions initiated by recruits, i.e., the replies, quotes, retweets, and mentions posted on each athlete's Twitter timeline during the month prior to commitment. I create an aggregate measure, combining these counts into a binary "interaction" feature (Table 10).

Social media also offers a rich source of text data from users' posts. As demonstrated in Chapter 2, the content posted by an athlete online provides information about his opinions and beliefs that can have significant impacts on his offline recruiting

success. The final set of features investigates whether the content posted by a recruit on social media is predictive of school choice, specifically the hashtags posted in the month prior to commitment. I expect that using hashtags relevant to a given school will be associated with increased likelihood of commitment. As free text data, the topic of a hashtag is not always evident, and I use a two-step information retrieval process to determine the likely topic of each hashtag:

1. For each of the 682 schools in the data, I generate a set of positive query terms (terms likely to retrieve documents relevant to the school) denoted T . These query terms are substrings 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 $T=\{\text{'utah'}, \text{'utes'}, \text{'utenati'}\}$. Each recruit's hashtags are treated as a set of documents D . Initial queries are constructed using the Boolean OR operator, where 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. I 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, $NT=\{\text{'utahst'}\}$, effectively excluding references to their in-state rival Utah State. $S1$ is then queried using the NOT operator. Of all potentially relevant documents retrieved by the first query, I include in the final set $S2$ only those that do not match any of the negative terms. I then create two binary features tracking whether the recruit posted a hashtag relevant to the prediction school or to another college that has offered a scholarship during the prior month (Table 10).

3.2.3 Predicting School Choice

In addition to constructing explanatory models that estimate the value placed on different objectives when selecting a college, I seek to predict which school a recruit will choose out of those that have offered him a scholarship, based on school characteristics, recruiting activities, and social media data. I elect to use logistic regression because of its interpretability and performance with non-normally distributed response variables.

To evaluate the contributions of each set of features, I implement the following 6 models with different combinations of feature groups. Model 0 (the baseline model) uses recruiting and institutional data only. Because some of the proposed features may be highly correlated to each other, I perform feature selection using a lasso regression with L1 penalty ($C=0.1$) to construct the baseline model. For consistent comparison to the other models, I manually remove predictors whose weight was reduced to 0 and re-fit the baseline using logistic regression without regularization penalty. Model 1 adds to the baseline the features related to a recruit's Twitter "friends," i.e., users followed by the recruit in the month prior to commitment. Model 2 focuses on followers in the online social network, adding features tracking the number and affiliation of in-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. Because the issue of collinearity arises again when combining all social media features, I apply lasso regression to perform feature selection when constructing Model 5. As with the baseline, I subsequently re-fit Model 5 using logistic regression without regularization penalty.

Stratified Monte Carlo cross validation is used to assess the performance of each model. Over 100 trials, the full data is randomly split into two equal subsets, each with the same proportion of positive and negative instances. I ensure that athlete-month instances corresponding to the same athlete are kept together in order to avoid training and testing on the same athlete. I evaluate the predictive power of each model on the hold-out testing set based on standard classification metrics: AUC, precision, recall, and F1 score.

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 the proposed models will predict more than one commitment per recruit. To account for this, I also evaluate the school choice prediction as a ranking. I rank each recruit's college options according to predicted probability, and I 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), I calculate NDCG@2 using only the top two predictions for each recruit. Simply, the model receives a perfect score if the recruit commits to the top-ranked prediction, a lower score if he commits to the second prediction, and no points if he commits elsewhere.

3.3 Results

This study analyzes the factors at play in college football recruits' school choice decisions and investigates the value of social media data in predicting commitments. The following sub-sections report the results of the fitted logistic regression models, compare

their predictive performance on the hold-out training data, and present an example application for school choice predictions.

3.3.1 Factors Related to School Choice

Logistic regression is commonly used in situations where the outcome variable Y is binary. In this study, the outcome records whether a recruit will commit to a specific college, calculating the predicted probability that $Y=1$ for each athlete-school pair. As opposed to linear regression, where the coefficients of each independent variable X can 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. I begin by fitting the five models on the full set of 4,408 instances. Model results are contained in Table 11.

Table 11 Fitted commitment regressions

Feature	Model 0		Model 1		Model 2		Model 3		Model 4		Model 5	
intercept	-1.1837	***	-1.1910	***	-1.1769	***	-1.0584	***	-1.1441	***	-1.1202	***
in_state	0.7450	***	0.7825	***	0.7784	***	0.7103	***	0.7180	***	0.7609	***
other_offer	-0.1019	***	-0.1122	***	-0.1093	***	-0.1048	***	-0.0914	***	-0.1010	***
other_closer	-0.0669	**	-0.0553	*	-0.0548	*	-0.0590	*	-0.0513	*	-0.0431	°
other_us_news	-0.0695	**	-0.0529	*	-0.0553	*	-0.0686	**	-0.0646	*	-0.0497	°
official	2.5224	***	2.1941	***	2.1985	***	2.3675	***	2.1733	***	1.9720	***
other_unofficial	-0.0364	°	-0.0412	°	-0.0421	°	-0.0351		-0.0286		-0.0377	°
other_official	-0.2952	***	-0.3242	***	-0.3020	***	-0.3260	***	-0.2922	***	-0.2860	***
first_offer	0.6459	***	0.6240	***	0.6287	***	0.6541	***	0.6608	***	0.6540	***
last_offer	-0.5206	**	-0.5872	***	-0.5839	***	-0.5264	**	-0.6081	***	-0.6658	***
last_coach	0.6443	**	0.4488	°	0.5449	*	0.5653	*	0.4943	*	0.3932	
last_official	1.3289	***	1.4002	***	1.3747	***	1.3457	***	1.4800	***	1.4853	***
coach_friends			0.4800	**							0.4449	**
recruit_friends			0.4519	***							0.3800	***
college_friends			0.3422	*								
coach_friends_other			-0.0027									
recruit_friends_other			-0.0259	**							-0.0241	**
college_friends_other			-0.1060	°							0.2768	
coach_followers					0.5218	**						
recruit_followers					0.3861	***						
college_followers					0.4378	*						
coach_followers_other					-0.0606	*					-0.0477	
recruit_followers_other					-0.0276	°						
college_followers_other					-0.2172	*					-0.2168	*
interaction							0.6619	***				
interaction_other							-0.1909					
hashtag									1.2733	***	1.1031	***
hashtag_other									-0.9099	***	-0.8630	***
Pseudo R-squared	0.3423		0.3741		0.3659		0.3478		0.3758		0.4005	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ° $p < 0.1$

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 44 to 11 features. As hypothesized, features that decrease costs of attendance for athletes are associated with increased likelihood of commitment. A recruit's odds of committing increase 111% for a college in their home state. This result agrees with previous work on college athletic recruitment, where attendance at an in-state school has been linked to decreased travel costs as well as increased sense of satisfaction and fit (Barden et al., 2013). Considering alternative options in the recruit's choice set, the odds of commitment to the prediction school decrease 10% for each offer from another school. The relative costs of these alternatives also factor into the commitment decision. The odds of choosing the prediction school decrease 6% for each offering school that is geographically closer and 7% for each school that has a higher academic ranking. Recruiting activities are also strong predictors of college choice. Each official visit to the prediction school is associated with a 1100% increases in the odds of commitment, while each official visit to another school decreases the odds of commitment by 26%. Each unofficial visit to another college is associated with a 4% decrease. The individual characteristics of recruits were not statistically significant.

The sequence variables support the use of the availability heuristic in recruits' decision-making, suggesting that recency and vividness may impact their evaluation of college options. A recruit has 91% greater odds of committing to a school when it was the first to offer him scholarship, with 41% lower odds for the last school to offer.

Anecdotal evidence suggests that recruits may attribute more emotional weight to their first scholarship offer. Shaquille Quarterman, a class of 2016 linebacker, expressed his sense of loyalty in a tweet about then-head coach at the University of Georgia, Mark Richt, “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.” The odds of choosing the prediction school increase 278% when it was the athlete’s most recent official visit and 90% when the last coach visit was from the prediction school.

Model 1 adds features tracking the recruit’s online social network out-links. Following accounts associated with the prediction school in the previous month is associated with increasing odds of commitment—62% for each coach, 57% for each committed athlete from the class of 2016, and 41% for each current player. A recruit’s odds of selecting the prediction school decrease 3% for each new friend from the class of 2016 committed to another school and 10% for each new friend currently playing football at another school. These results indicate that the recruit’s online connections may convey information about his college preferences. While the feature tracking the recruit’s out-links to coaches from other colleges were not significant, the coefficient shows the expected sign.

Model 2 adds features tracking the number of Twitter users associated with the prediction school and other schools following the recruit in the month prior to commitment. Each coach, recruit, and current player from the prediction school following an athlete increases his odds of commitment by 69%, 47%, and 55%, respectively. As recruitment occurs in a competitive environment, connections with other schools are also analyzed. Each coach, recruit, and current college football player from another school

following the recruit in the previous month decreases the odds of selecting the prediction school by 6%, 8%, and 20%, respectively. These findings support the hypothesis that social media connections can be interpreted as signals of the strength of the relationship between an athlete and school, and ultimately predict his school choice decision.

Model 3 focuses on the effect of Twitter “interactions” (mentions, replies, quotes, and retweets posted by the recruit). Interacting with the prediction school at least once during the prior month is associated with a 94% increase in the odds of selecting that school. This result suggest that athletes may invest more effort into building relationships with their preferred schools, consistent with social capital theory (Lin, 2001), and that this information can be useful for modeling and predicting school choice.

Model 4 investigates the use of text data for predicting school choice, specifically the hashtags posted by the recruit in the month prior to commitment. Tweeting at least one hashtag relevant to the prediction school increases the odds of commitment by 257%, and referencing a competing school decreases the odds by 60%. These results suggest that athletes may signal their college preferences via content posted on social media.

Model 5 combines all of the features tested in Models 0-4. After applying lasso regression again to correct for potential collinearity issues, this process results in a final model with 19 features (11 offline recruiting features and eight social media features). Who a recruit chooses to follow in Twitter is a strong predictor of his commitment intentions. The odds of commitment increase 56% and 46% for each coach and recruit from the prediction school followed by athlete in the prior month, respectively. Each new friendship with a recruit committed to another school is associated with a 2% decrease in odds. Each current player from another school following the athlete is associated with a

19% decrease in odds of commitment to the prediction school. Interestingly, interactions were eliminated from the final model, suggesting that a coach seeking to estimate the likelihood that a recruit will commit may be better served by focusing on his social network and Twitter content. Posting a hashtag associated with the prediction school increases odds of commitment 201%, and the odds of selecting the prediction school decrease 58% when posting a hashtag relevant to another school. Overall, these results demonstrate the utility of considering information revealed by behavior and relationships in both offline and offline environments when predicting school choice decisions.

3.3.2 Predictive Performance

To evaluate the contribution of the different sets of features to school choice predictions, each model is trained on a random split of training and testing data over 100 trials. The baseline Model 0, which contains features related to cost/benefit objectives, alternatives, and decision-making heuristics, achieves an average AUC of 0.659. The features tracking the sequence of recruiting events yield a 3% improvement in AUC over the same model with these features omitted (AUC=0.640). This result supports the hypothesis that a quasi-rational approach may be appropriate when modeling and predicting commitment decisions in college football.

I also compare the performances of the baseline model containing only institutional and recruiting data with models incorporating features derived from online social media. Figure 4 displays predictive performance of each model, averaged over 100 trials.

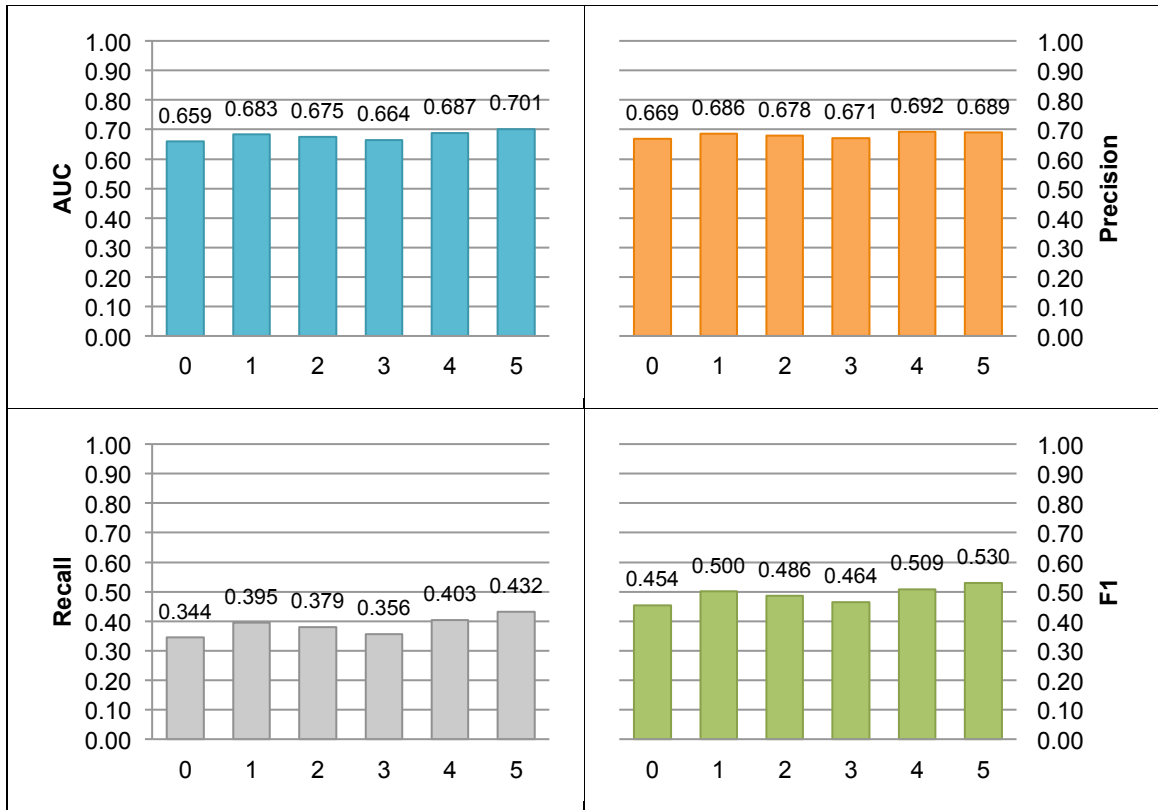


Figure 4 Performance of commitment predictions

These results demonstrate that models incorporating social media data (Models 1-5) dominate the baseline across all metrics. Among the individual sets of social media features, Model 4, which incorporates features related to the recruit’s Twitter content, demonstrates the largest performance increase relative to the baseline. Model 4 achieves a 4.4% increase in AUC ($p < 2.2 \times 10^{-16}$), 3.5% increase in precision ($p = 3.7 \times 10^{-7}$), 17.1% increase in recall ($p < 2.2 \times 10^{-16}$), and 12.1% increase in F1 score ($p < 2.2 \times 10^{-16}$) over the baseline. Model 1, tracking out-links to coaches, recruits, and current college football players, has the second-largest gains over the baseline, with a 3.7% improvement in AUC ($p < 2.2 \times 10^{-16}$), 2.6% improvement in precision ($p = 0.0004$), 14.7% improvement in recall ($p < 2.2 \times 10^{-16}$), and 10.2% improvement in F1 score ($p < 2.2 \times 10^{-16}$). Model 2, which incorporates features related to Twitter

followers, also makes significant gains across all metrics. Only Model 3, tracking the athlete's online interactions (mentions, replies, retweets, quotes), fails to improve on the baseline, and both features related to social media interactions are eliminated from the final model via lasso. These results suggest that features tracking the athlete's online social network structure and the content he posts may provide more insight for understanding and predicting his school choice intentions than his interactions with users affiliated with the schools that are recruiting him.

Ultimately, the final model (Model 5) is the top performer, with an average AUC of 0.701. In addition to achieving significantly higher performance than the baseline containing only recruiting and institutional data across all metrics ($p < 2.2 \times 10^{-16}$ for AUC, $p < 2.7 \times 10^{-4}$, $p < 2.2 \times 10^{-16}$ for recall, and $p < 2.2 \times 10^{-16}$ for F1 score), the combined model represents a significant improvement over any of the individual sets of social media features (Models 1-4) in terms of AUC, recall, and F1 scores, although it fails to make a significant improvement on the precision of Models 1 and 4.

Similar results are achieved when evaluating the school choice predictions as a ranking problem. Figure 5 displays the NDCG@2 scores for each model averaged over all recruits. Again, models using data from the recruit's social media consistently outperform the baseline, with Models 1-5 achieving gains of 2-9% over the Model 0 score. The results demonstrate that features related to the recruit's social network and Twitter content add more value than features tracking Twitter interactions, as Model 3 shows the smallest improvement. Again, Model 5 has the highest performance, with an average NDCG score of 0.755. Model 5 correctly ranks the commitment school first for 111 out of 182 recruits and second for an additional 42 recruits. These results indicating

that combining features related the costs and benefits of attendance, comparisons to alternative options, recruiting activities, decision-making heuristics, and social media can yield accurate school choice predictions. It is also possible to measure differences in NDCG across the same athlete, and Model 5 has the highest average gain over the baseline model.

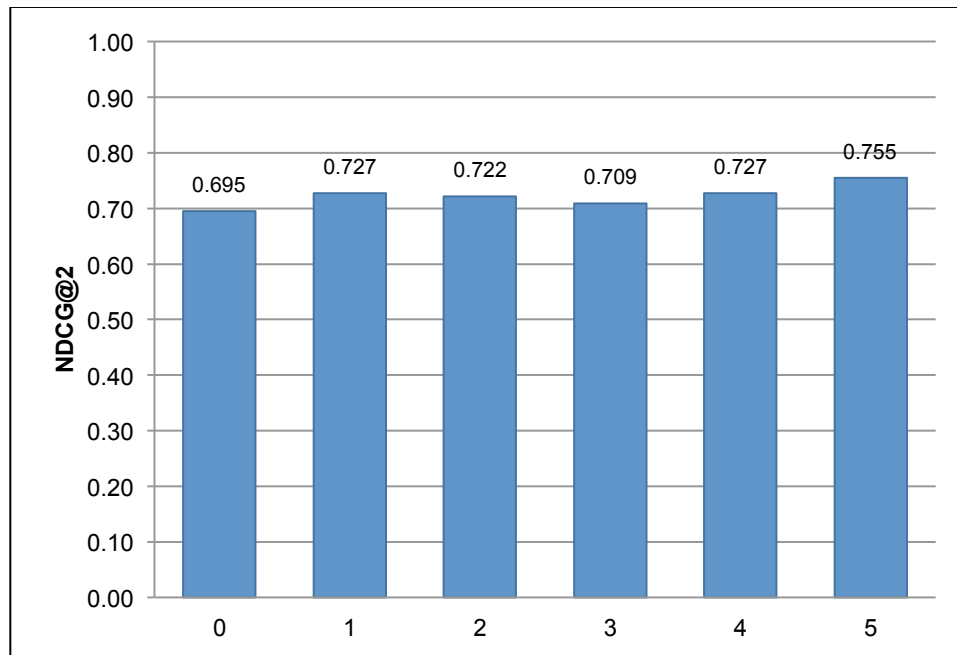


Figure 5 Performance of school choice rankings

3.3.3 Application to Recruiting Resource Allocation

In addition to better understanding the decision-making of college football recruits, this work is intended to provide practical insights for coaches and recruiters. Therefore, I produce a sample report that a college recruiting staff might consider when making decisions on how to identify recruits who are most likely to commit and effectively allocate their resources. Table 12 lists the predicted probability of ten high school athletes that received offers from the University of Iowa, but remained uncommitted as of January 2016.

Table 12 Top 10 Iowa prospects (January 2016)

Name	Star	Position	Hometown	P(Iowa)	Top Prediction
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%)
K.J. 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%)
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 (two coaches and five committed recruits followed him on Twitter during the prior month). Additionally, he showed his preference for Iowa by following two coaches, six committed recruits, and posting six Iowa hashtags during the same month. Ranked by predicted probability of attendance, Iowa's next prospect was Matt Farniok. He had made an official visit and was followed by two coaches and two committed recruits from Iowa. Unlike Jackson, he did not follow any Iowa accounts and did not post any Iowa hashtags. A coach interpreting these predictions might safely consider Jackson a stronger prospect. Given the fact that both recruits were offensive tackles, the coach could decide to focus his efforts on maintaining a relationship with Jackson during the final weeks of recruitment. Alternatively, if the team needed additional players at the same position, the coach might also decide to take an action intended to increase the probability of Farniok choosing Iowa. While the results of this study are not causal, they do suggest that having current players from Iowa follow a recruit on Twitter is associated with increased likelihood of commitment. Interestingly, Iowa football has a restrictive Twitter policy for current players, prohibiting tweeting but

allowing following and liking. As demonstrated in Chapter 2, Twitter content can have a significant impact on recruiting outcomes, suggesting that Iowa may be missing out on some of the benefits of peer recruiting. Ultimately, Jackson did commit to Iowa, while Farniok selected Nebraska.

This model of school choice in college football may also be useful for coaches seeking to identify which prospects are most likely to select a competitor. In addition to predicting a binary outcome, the logistic classifier estimates the probability of commitment for each college that has offered a recruit, allowing coaches to estimate their relative chances of securing a commitment. For example, K.J. Gray was predicted to have only 6% chance of choosing Iowa, compared to a 95% chance of committing to Rutgers. While Gray was followed by two Iowa recruits during the prior month and followed them back, he was followed by three Rutgers recruits and two current players and followed them back, in addition to being from New Jersey. 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, I believe that it shows the potential for the proposed model to be applied in real-world settings and inform the strategic decisions of college football coaches.

3.4 Discussion

This work represents a novel addition to the research literature examining the intersection of social media and individual decision-making. I extend previous studies predicting school choice (Dumond et al., 2008, Mirabile & 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 6% improvement in AUC, 3% improvement in precision, 25% increase in recall, and 17% improvement in F1 score over the baseline model. These results suggest that a model of school choice incorporating information gleaned from both offline and online behaviors would be most effective, and that combining features representing different aspects of the recruit's online social network may be more successful than any individual set of social-media-derived features.

Additionally, I find that among the four types of social media features explored, those focusing on hashtag content (Model 4) contribute the most to predictive performance. Although this study focused on hashtags, it provides promising evidence that recruits' Twitter content is strongly predictive of school choice intentions. These results suggest that recruiters in athletics and other domains should consider the preferences communicated by candidates on social media when deciding how to direct recruiting resources. Structural features (Models 1 and 2) also add value to school choice predictions, indicating that recruits' and colleges' connective behaviors on social media function as clearer signals of interest than other relationship-building behaviors, such as mentions, replies, or retweets.

These findings can inform the recruiting strategies employed in both college football and HR, where evidence suggests that social media is not yet as widely adopted. While 96.2% of recruits in the class of 2016 with public Twitter accounts followed at least one college coach, only 35% of job-seekers report following a potential employer on Twitter (Westfall, 2017). Furthermore, 96.6% of recruits were followed by at least one college coach. Following an athlete or a job candidate online costs nothing and is

positively related to likelihood of commitment. Moreover, these relationships can have important effects post-commitment. Social ties in the workplace have been credited for superior socialization (McManus & Baratta, 1992) and increased willingness to share information (Chow & Chan, 2008). As stated in an analysis of Chinese firms' use of instant messaging and social media, "by strengthening social networks in the work-space, Web 2.0 applications have the potential to enhance coordination, collaboration, knowledge sharing, and problem solving" (Davison, Ou, Martinsons, Zhao, & Du, 2014, p. 2035). Specific to college sports, an analysis of Division I and III men's basketball found that trusting relationships with coaches were linked to measurable performance effects (Elsass, 2001).

The proposed model can also be compared to previous studies predicting college football commitments (Dumond et al., 2008; Mirabile & Witte, 2015). Both 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. Because only 17% of instances in the data correspond to commitments, I did not focus on overall accuracy (proportion of correctly predicted instances), which is not robust to class imbalance. However, when averaged over 100 trials, Model 5 achieves 89.6% accuracy and ranks the commitment school first in terms of predicted probability 61% of the time and or second 23% of the time. These findings suggest that it is worthwhile to include consideration of decision-making heuristics and social factors into a model of school choice. This study represents both a promising first step in analyzing and predicting school choice in college football using social media data and applying this approach other recruitment contexts.

CHAPTER 4 DECOMMITMENTS³

The decisions of individuals are often influenced by their social networks. In this study, I investigate the relationship between online social networks and organizational turnover. Specifically, I explore how a college football recruit's Twitter network may predict his level of commitment to a prospective college team.

This research makes a unique contribution to the body of knowledge on organizational networks by leveraging social media data. Social media facilitates the collection of large-scale network data in situations where observing offline social ties would be expensive or impossible, offering an unprecedented opportunity to study network effects on a large scale. While previous studies of social factors in employee turnover have generally focused on a single organization or a few companies in the same region (e.g., Feeley, Hwang, & Barnett, 2008), this work is unique in providing a holistic view of an entire labor market, tracking the offline recruiting activities and online social network connections 2,644 athletes in the college football recruiting class of 2016. Previous work in the field of information science has explored the power of online social network data to participation in online health communities (Wang, Zhao, & Street, 2017), box office success (Ding et al., 2016), and event attendance (Bogaert et al., 2015), among others. However, this study is among the first to assess how data from individuals' online social networks can be used to predict offline turnover.

Recruitment occurs in many contexts, and I draw comparisons between turnover in college football and human resources (HR). In HR, turnover refers to an employee's voluntary departure from an organization. In college football, decommitments occur

³ A version of this study has been submitted for publication to *Decision Support Systems*. This study was also presented at the 2017 INFORMS Workshop on Data Science, Houston, TX and the 2017 Winter Conference on Business Analytics, Snowbird, UT.

when a recruit reneges on a verbal commitment to attend a college. In comparison to other situations where an athlete separates from his college team—academic ineligibility, disciplinary suspensions, or transfers—I contend that decommitments most closely resemble HR turnover. First, decommitments are initiated voluntarily by the recruit. Furthermore, he may commit and decommit to any number of schools during the recruitment process without negative effects on their future college eligibility, analogous to an employee’s freedom to change jobs. I utilize data on athletes’ recruiting activities, college options, and Twitter networks to construct a predictive model of decommitment.

I expand the work presented in Chapter 3, which demonstrated that a recruit’s online social networks and Twitter content convey information about his commitment intentions, by predicting the occurrences of decommitments over time. This approach is unique, as the vast majority of previous turnover studies are either a) cross-sectional, collecting data on predictors and turnover intentions at a single point in time, or b) time-lagged, measuring predictors at one time, and outcomes at a later time. These methods do not allow researchers to capture key events that occur between the two times, nor to examine changes in predictors, and a recent meta-analysis of the literature has called for methods that “capture turnover processes in real time to establish temporal order” (Rubenstein et al., 2018, p. 52). I construct a dynamic model that considers an individual’s offline recruiting activities and online social media up to a specific month t , and predicts whether the athlete will decommit in the next month, $t+1$. Predicting whether an athlete will decommit during a specific month adds complexity, but also reflects the realistic use case, where coaches require up-to-date information to decide on recruiting strategies. I propose two main research questions:

Q1: What recruiting activities and institutional characteristics predict athletic decommitments?

While previous work in the HR domain has investigated individual, organizational, and environmental predictors of employee turnover (e.g., Rubenstein et al., 2018), this study is first to examine predictors of decommitment in college football recruiting, making a significant and timely contribution to the sports analytics literature. While a 2012 study in *Sports Illustrated* (Staples, 2012) estimated the rate of college football decommitment at approximately 12.4%, there are recent indications (Carson, 2016) that decommitments are occurring with increasing frequency. I draw upon school choice literature (e.g., Croft, 2008; Dumond et al., 2008; Sander, 2008) to identify context-specific factors related to the recruit's satisfaction with his current school and attractiveness of alternatives (other schools that have offered a scholarship). Because the decision-making process may vary by individual (e.g., Doyle & Gaeth, 1990), I also consider the relationship between the recruit's demographic characteristics, such as home state and star rating, and his chances of decommitting in a specific month.

Q2: What value does social media data add to decommitment predictions?

By first controlling for athletes' personal characteristics, school satisfaction, and perceived alternatives, I can measure the value added by social-media-derived network features. Previous research on network effects in employee turnover has primarily focused on offline, workplace networks. Using data from athletes' Twitter profiles, I expand my analysis to inter-organizational and personal social networks, constructing features relating to in-links, out-links, and behavioral diffusion. The effectiveness of

different types of features is assessed by comparing models focusing on various aspects of recruits' social networks.

The results indicate that considering information from athletes' Twitter networks improves decommitment predictions. This research can assist college coaches in identifying vulnerable commitments and informing their recruiting and retention strategies. The effects of the online social network features examined in this work are consistent with previous HR research on the relationship between offline social networks and employee turnover (e.g., Feeley et al., 2008), suggesting that this framework may be generalized to non-athletic contexts. Ultimately, this work represents both a promising first step in predicting decommitments in college football, and a novel addition to the field of information science, examining how the information that individuals reveal about their opinions and beliefs through online relationships can be used to understand and predict their offline decisions.

4.1 Related Research

Turnover presents a significant problem for organizations. In addition to losing the benefit of the departing employee's work product, knowledge, and network connections, companies must expend resources on recruiting a replacement, estimated to cost between \$4000 (Kridler et al., 2015) and \$6500 (Fernandez et al., 2000) per employee. High rates of turnover may be interpreted as a signal of poor organizational culture, working conditions, or leadership (Cable & Turban, 2003), and poor reputation has been linked to reduced pride in membership and employee tenure (Helm, 2013), suggesting a feedback loop of adverse outcomes. Thus, it is unsurprising that a large body of HR research has focused on identifying predictors of turnover. An employee's

decision to leave an organization is influenced by a multitude of factors, and one branch of turnover research centers on the effects of effects of individual, organizational, and environmental characteristics.

While the evidence supporting the effect of individual demographic variables (e.g., race, gender) is mixed, meta-analyses have consistently shown that age and tenure are negatively correlated to job turnover (Griffeth et al., 2000; Rubenstein et al., 2018). Regarding organizational factors in turnover, March and Simon (1958) proposed that an employee's turnover intention is strongly correlated to the "desirability of leaving" her current organization. Indeed, job satisfaction has been called "the single most reliable predictor of turnover," (Moynihan & Pandey, 2007, p. 208), and multiple meta-analyses (Griffeth et al., 2000; Rubenstein et al., 2018) have confirmed this relationship. Primary studies of turnover have differed in their measurement of satisfaction. Those using survey/questionnaire methods have often assessed satisfaction in a single, summary item (e.g., Feeley et al., 2008). The concept of satisfaction has also been operationalized into variables including pay and task complexity (Cotton & Tuttle, 1986). I follow the latter approach in this study, using publicly available data about the recruit's current school to estimate his level of satisfaction. Turnover has also been linked to job market conditions and the perceived ease of leaving the organization (March & Simon, 1958). Like satisfaction, perceived alternatives have been measured through both survey items and observable features like unemployment rate (Cotton & Tuttle, 1986). Meta-analyses have demonstrated that variables related to specific, concrete job alternatives and comparisons to an employee's current position are superior predictors of turnover (Griffeth et al., 2000; Rubenstein et al., 2018). In the current study, I consider features related to the

availability of alternative options (number of scholarship offers) as well as the attractiveness of these specific options in comparison to the recruit's current school.

The second branch of research on turnover considers social and relational factors. Mossholder, Settoon, and Henagan (2005) found that, after controlling for tenure, age, gender, and job satisfaction, network centrality and interpersonal citizenship behavior were significant predictors of turnover among healthcare workers. Feeley et al. (2008) also discovered that subjects who reported greater numbers of friendships in the workplace had a lower likelihood of turnover. While Moynihan and Pandey (2007) did not track actual social ties, they surveyed 326 non-profit employees about perceived coworker support and obligation toward coworkers, finding that both were negatively correlated to short-term turnover intentions. Introduced by Granovetter (1985), the theory of "embeddedness" suggests that turnover intention is strongly linked to the degree of attachment between members of an organization. Indeed, intra-organizational ties have been found to increase sense of person-organization fit, socialization, and the perceived difficulty of movement, ultimately decreasing the likelihood of employee turnover (Allen, 2006). In terms of social recruitment methods, referral hiring has been linked to higher job survival and a slight performance advantage (Zottoli & Wanous, 2000).

The current study provides a unique addition to the literature by investigating the relationship between inter-organizational ties and turnover as well. Previous research on social factors in employee turnover has been almost entirely dependent on small-scale, survey methods to record social ties within a single organization (e.g., Feeley et al., 2008). One exception is Moynihan and Pandey (2007), who tracked external networking (e.g., conferences, professional society memberships). The authors extended

Granovetter's (1973) "strength-of-weak ties" theory to turnover behavior, expecting that inter-organizational networks would influence turnover intention via the perceived ease of movement. While Moynihan and Pandey (2007) did not find strong support for their hypothesis, this result may be largely due to their focus on professional networking activities, rather than actual inter-organizational social ties. In contrast, I use Twitter data to observe in-links and out-links to coaches, players, and recruits from other schools that have offered a scholarship.

Yet previous research has suggested that the effects of social networks in the workplace are not always positive. In a two-year study of referral hiring at a telephone customer service center, Fernandez et al. (2000) found that employees hired via referrals were more likely to leave the organization after their referrers, suggesting a diffusion effect in turnover behavior. Informed by social comparison theory (Festinger, 1954), which holds that decision-making is informed by the context of "similar others," Felps, et al. (2009) proposed a theory of turnover contagion. The authors hypothesized that an individual's likelihood of turnover is influenced by coworkers' job embeddedness (a summary measure comprising person-job fit and network centrality) and turnover intentions. Across two studies—a survey of 8,663 employees of a nationwide hospitality and recreation organization and a survey of 234 bank employees—the authors concluded that coworker embeddedness was a significant, negative predictor of turnover whose effect was mediated by coworkers' job search activities.

Like turnover in HR, athletic decommitments have many negative impacts on college teams, and accurate predictions would be useful for preventing or mitigating these effects. Replacing a departing athlete is costly, and the process is complicated by

the competitive atmosphere and the constrained timeline of recruitment. Yet in contrast to the large body of literature on HR turnover, there is only one previous study of athletic decommitments. A 2012 article in *Sports Illustrated* measured the rate of college football decommitments from 2007-2011 (Staples, 2012). The author found that 12.4% of the Rivals.com top 100 recruits decommitted and signed with another school. My research makes a significant contribution to the athletic recruiting literature as the first to build a predictive model of decommitments.

In the absence of prior research predicting decommitments, I am informed by work on school choice in college football recruiting. Dumond et al. (2008) considered economic factors impacting school choice, with an individual recruit's expected utility of attendance determined by features including the college's amenities, geographic distance, academic ranking, and football team performance. I apply the authors' cost/benefit framework to create features that estimate the athlete's level of satisfaction with his current school. Additionally, I expand the scope of the analysis by considering recruits of all ability levels (rather than just the top 100 recruits).

Similar to employee turnover, decommitments are likely to be related to the recruit's social network. Mass media reports (Myerberg, 2015; Prunty, 2008; Rutter, 2015) and surveys of high school and college athletes (Croft, 2008; Lujan, 2010; Sander, 2008) have highlighted the influence of coaches, parents, and other athletes on school choice, though no previous work has examined the effect of social networks on decommitments. Mirabile and Witte (2015) considered the value of familial connections (father, brother, cousin, or uncle who played football at the same college) for predicting school choice. The authors analyzed ten years of recruiting data and achieved 65%

accuracy predicting the commitments of 19,815 recruits. In comparison, my work utilizes social media data to track the number of in-links and out-links to coaches, current players, and recruits affiliated with the athlete's current school and the other schools that are recruiting him.

Despite evidence that coaches and athletes increasingly use social media to communicate and build connections (e.g., Crabtree, 2016), there is very little research examining its role in the recruiting process. Illustrating the relationship between social media and decommitments is the case of Tate Martell and Texas A&M. When the third-rated quarterback in the class of 2017 decommitted in May 2016, the damage caused by losing a marquee recruit was compounded by the social media comments of an assistant coach condemning a lack of "loyalty" (as quoted in Khan, 2016). Several other committed recruits quickly decommitted, indicating the negativity expressed by the coach and Texas A&M fans as the cause. In this study, I analyze how a recruit's online connections reveal information about his decommitment intentions as well as how decommitment behavior can diffuse in online social networks.

4.2 Methods

This study explores the benefit of considering online social networks for predicting American college football decommitments over time. In other words, I seek to predict whether a committed athlete will decommit in October, in November, and so on each month until National Signing Day. By constructing several models from different sets of features, I measure the value added by social media data and compare the performance of models focusing on various aspects of athletes' social networks.

4.2.1 Data

As detailed in Chapter 2, I gathered data on the recruiting activities, college choices, and Twitter profiles of the athletes in the class of 2016. For the purposes of this project, I define a decommitment as occurring when a recruit with a current verbal commitment to a college revokes this pledge and either commits to another institution or returns to an uncommitted status. I identified decommitments from 247Sports.com recruiting data in two ways: (1) from decommitment announcements tracked in an athlete's timeline, and (2) from instances when an athlete with a previous verbal commitment subsequently commits to another institution. Of the 2,644 recruits in the full dataset, 536 unique individuals (20.3%) decommitted at some point during recruitment—34 recruits decommitted twice and one decommitted three times, resulting in 572 total decommitment events. This rate is greater than the 12.4% observed in previous empirical work (Staples, 2012), and supports claims by sports media that decommitments are becoming more frequent (Carson, 2016).

Because this analysis focuses on the relationship between online social media and offline decommitments, I excluded 315 recruits without public Twitter accounts. Furthermore, I considered only recruits that were verbally committed at any point between October 2015 and February 2016. A recruit cannot decommit if he is not currently committed or has already signed an NLI, and this time range was selected so that at least one month of retrospective Twitter data is available for each recruit.

These steps reduced the data to 1,785 athletes and 370 decommitments. To address the problem of predicting the occurrence and timing of decommitments, I constructed a dataset of “athlete-month” instances. For each recruit, I created an instance

for each month that he was verbally committed between October 2015 and February 2016. This process resulted in 7,128 instances, each associated with a binary outcome tracking whether the athlete decommitted during the specified month, and features related to institutional characteristics, recruiting events, and Twitter connections, as recorded up to the end of the prior month. Each recruit may be associated with multiple instances. For example, instance 80622_October asks whether Jacob Huesma, a two-star quarterback recruit (ID 80622), will decommit in October, given his offline recruiting activities and online social media measured through the end of September, while 80622_November predicts whether he will decommit in November based on predictors through the end of October. While 20.8% of all 2016 recruits decommitted at some point between October and February, only 5.2% of all athlete-month instances represent a situation where the specified athlete decommitted during the specified month.

4.2.2 Modeling Decommitments

College football is often categorized as a two-stage model. First, coaches narrow the pool of 2,000 or 3,000 recruits to a few hundred prospects and decide who to extend scholarship offers to, a process that is discussed in Chapter 2. Then recruits select a college out of those that have offered him a scholarship, described in Chapter 3. However, the recruitment process does not truly end with a commitment. Recruits can make a verbal commitment to a college offering a scholarship at any time (NCAA, 2016). But because these “verbals” are non-binding, recruits can decommit and commit to another school until signing an NLI. Although an early signing period was adopted for the class of 2018 (USA Today High School Sports, 2017), during the period of data collection for this study most recruits could not sign NLIs until National Signing Day

(February 3, 2016). In this study, I focus on the recruit's decision to decommit from his current school and predict whether he will decommit in a specified month between October 2015 and February 2016. I assume that this decision will be based on his satisfaction with his current school, the availability of alternatives, and his social network. The proposed model of decommitment includes different sets of features corresponding to these choice factors.

I first construct a set of baseline features from the 247Sports.com recruiting data and institutional data. As decision-making varies by individual, I include features related to the recruit's personal characteristics, e.g., star rating, the length of his current commitment, and whether he has previously decommitted from another school (Table 13). I extend the concept of job satisfaction to the college football recruiting process, expecting that features that increase the benefits and decrease the costs associated with attendance at the current college (termed "original school" in Table 13) will decrease the likelihood of decommitment. For example, recruits may be less likely to decommit from an in-state school. Conversely, I expect that features that decrease the benefits associated with attendance, such as a head coaching change or poor record, will increase the likelihood of decommitment. I also consider the affinity between the recruit and his current college demonstrated by unofficial, official, and coach visits. As with HR turnover, athletic decommitments are likely to be impacted by the perceived ease of movement, and I create features that measure the availability of concrete alternatives, i.e., the other colleges that have offered him a scholarship (termed "other" in Table 14). I also expect that the relative attractiveness of alternatives, such as the number of schools that have superior records, will impact the chances of decommitment.

Table 13 describes the 44 baseline decommitment features. I note that these features are time-consistent. For example, to predict if an athlete will decommit in January, only official visits to other schools that occurred before January 1 are counted.

Table 13 Baseline decommitment features

Type	Feature	Description
<i>Demographic</i>	star	Recruit star rating (0, 2, 3, 4, 5)
	height	Height (inches)
	weight	Weight (pounds)
	BMI	Body mass index ($weight/height^2$) \times 703
	position	Recruit position (ATH, DB, DL, LB, OB, QB, RB, ST, WR)
	region	U.S. Census region of recruit (INT, MW, NE, S, W)
	hotbed	1 if hometown located in recruiting hotbed states of CA, FL, or TX, 0 otherwise
<i>Satisfaction</i>	days_committed	Number of days since verbal commitment
	prior_decommit	1 if recruit has previously decommitted from another school, 0 otherwise
	distance	Distance between recruit hometown and original school (miles)
	in_state	1 if recruit hometown in same state as original school, 0 otherwise
	type	Institutional type of original school (military, private, public)
	us_news	Original school academic ranking in U.S. News
	division	NCAA division of original school (FBS, FCS, II, III, JUCO)
	power	1 if original school is a member of Power 5 conference, 0 otherwise
	AP	Original school ranking in AP poll
	postseason	1 if original school played in bowl game during 2014 season, 0 otherwise
	historic_win	Winning percentage of original school over past five seasons (2010-2014)
	current_win	Winning percentage of original school during 2015 season
	commits	Number of other recruits verbally committed to original school
	coach_change	1 if head coach change at original school during 2015 season, 0 otherwise
sanctions	1 if football program under active NCAA sanction or probation, 0 otherwise	
first_offer	1 if original school was first to offer recruit, 0 otherwise	
unofficial	Number of unofficial visits to original school	
coach_visit	Number of coach visits from original school	
official	1 if recruit has taken official visit to original school, 0 otherwise	
<i>Alternatives</i>	other_offer	Number of offers from school other than original school
	other_closer	Number of offering schools closer than original school
	other_us_news	Number of offering schools with higher U.S. News ranking
	other_power	Number of offering schools from Power 5 conference
	other_AP	Number of offering schools with higher AP poll ranking
	other_postseason	Number of offering schools that played in bowl game during 2014 season
	other_historic_win	Number of offering schools with greater winning percentage over last five seasons (2010-2014)
	other_current_win	Number of offering schools with greater winning percentage during 2015 season

Table 13 (cont.)

other_coach_change	Number of offering schools with head coach change during 2015 season
other_sanctions	Number of offering schools with football program under NCAA sanction or probation
offers_post	Number of offers from other schools since commitment
other_unofficial	Number of unofficial visits to other schools
unofficial_post	Number of unofficial visits to other schools since commitment
other_coach_visit	Number of coach visits from other schools
coach_visit_post	Number of coach visits from other schools since commitment
other_official	Number of official visits to other schools
official_post	Number of official visits to other schools since commitment
days_to_NSD	Number of days remaining until National Signing Day

Following previous research examining network effects in personnel turnover (Feeley et al., 2008; Mossholder et al., 2005; Moynihan & Pandey, 2007), I expect that the recruit's decision to decommit will be related to his social network. I extend this concept to the online domain, and the second set of features centers on Twitter "friends," i.e., other Twitter users followed by the recruit. I hypothesize that recruits' online connections reveal information about their decommitment intentions, and that recruits with strong commitments will follow individuals from their new school. I am informed by the theory of network realignment, which holds that overlap in the individual social networks of the members of a dyad will increase with the intensity of the relationship (Jowett & Timson-Katchis, 2005) as well as previous surveys of Division I athletes (Sander, 2008). Because of the dynamic nature of the prediction problem (predicting whether an athlete will decommit in a specified month), I focus on tracking new friends added or dropped in the previous month rather than the cumulative number of friends. I construct six features tracking the number, affiliation, and type of the recruit's Twitter out-links (Table 14). Out-links to coaches, recruits, and current players at the athlete's

original commitment school are denoted as “original,” and connections to individuals from other schools that have offered a scholarship to the athlete are termed “other.”

Table 14 Social media decommitment features

Type	Feature	Description
<i>Out-links</i>	coach_friends	Number of coaches from original school followed by user over prior month
	recruit_friends	Number of 2016 recruits from original school followed by user over prior month
	college_friends	Number of current players from original school followed by user over prior month
	coach_friends_other	Number of coaches from other schools that have offered a scholarship followed by user over prior month
	recruit_friends_other	Number of 2016 recruits from other schools followed by user over prior month
	college_friends_other	Number of current players from other schools followed by user over prior month
<i>In-links</i>	coach_followers	Number coaches from original school following user over prior month
	recruit_followers	Number of 2016 recruits from original school following user over prior month
	college_followers	Number of current players from original school following user over prior month
	coach_followers_other	Number of coaches from other schools that have offered a scholarship following user over prior month
	recruit_followers_other	Number of 2016 recruits form other schools following user over prior month
	college_followers_other	Number of current players from other schools following user over prior month
<i>Diffusion</i>	committed	Total number of reciprocated connections to currently committed 2016 recruits
	decommitted	Total number of reciprocated connections to 2016 recruits that have previously decommitted

The third group of features focuses on the in-links from other Twitter users, i.e., “followers.” These features are intended to capture the intensity of social recruiting and peer recruiting (Myerberg 2015; Rutter, 2015). Previous work on sports recruiting has highlighted the importance of recruiting networks, with a hockey coach stating, “It’s a sport of connections; your athletes are your biggest recruiters, athletes that you coached, athletes that you’ve played with” (as cited in Elliot & Maguire, 2008, p. 169). Thus, I expect that the likelihood of decommitment will decrease as the number of followers

from the original school increases. I construct six features recording the type, number, and affiliation of new Twitter followers during the previous month (Table 14).

The fourth group of features examines the actions of athletes' social network neighbors. A large body of research in psychology and sociology has explored how social networks impact individuals' decisions (e.g., Festinger, 1954), and the behavior of one's coworkers has specifically been linked to personnel turnover (Fernandez et al., 2000; Felps et al., 2009). I expect that the commitment status of the recruit's peers will impact his likelihood of decommitment. Consistent with a threshold model of diffusion, I count the total number of reciprocated connections to committed and decommitted recruits prior to the prediction month (Table 14).

4.2.3 Predicting Decommitments

As different classification approaches are suited to different types of problems, model selection may greatly impact prediction quality. I initially test five classification algorithms from the scikit-learn package for Python (Pedregosa et al., 2011), logistic regression, decision tree, support vector machine, artificial neural network, and random forest. More details about each classification method can be found in *Introduction to Data Mining* (Tan et al., 2006). Beginning with the same set of 44 baseline features (Table 14), I performed feature selection specific to each method and selected appropriate parameters via grid search, described below:

- Logistic regression: I conducted feature selection using a lasso regression with L1 penalty ($C=0.1$) to build a baseline model. Manually removing predictors whose coefficients reduced to zero, the size of the baseline model decreased to 20

features. After removing extraneous features, I compare the performance of logistic regression implemented without regularization to the other classifiers.

- Decision tree: Entropy was selected as the split criterion for the decision nodes. The maximum number of features considered when searching for the best split was set to $\sqrt{n_features}$. The maximum tree depth was set to five levels.
- Support vector machine (SVM): Based on the grid search, I implemented a radial basis function kernel with a penalty weight of 1. To avoid overfitting and accelerate training speed, I applied univariate feature selection (based on an ANOVA F-test of each feature). This process reduced the size of the baseline model to 20 features.
- Artificial Neural Network (ANN): The hyperbolic tangent function was used as the transfer function between the layers, with the Adam algorithm for weight optimization.
- Random Forest: Based on a grid search, the ensemble size was set to 35 individual trees. Splits were based upon entropy, with consideration of $\sqrt{n_features}$.

Stratified Monte Carlo cross validation was used to assess classifier performance.

For 100 independent trials, the full data is randomly split into two equal subsets, each with the same proportion of positive and negative instances. I ensure that athlete-month instances corresponding to the same individual were kept together in order to avoid training and testing on the same athlete. Given that the distribution of the outcome is highly unbalanced (5% decommitments), oversampling is applied to decrease bias in the training data and improve classifier performance. I create an oversampled training set

with 50% positive instances—examples where the specified athlete decommitted in the specified month—by randomly copying and inserting positive instances. I evaluate predictive performance on the hold-out set using standard metrics: AUC, precision, recall, and F-1 score.

Table 15 displays the mean performance of each classifier over 100 trials.

Standard deviations are shown in italics next to each column.

Table 15 Decommitment classifier testing results

Method	AUC		Precision		Recall		F1	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Logistic	0.653	<i>0.016</i>	0.171	<i>0.042</i>	0.435	<i>0.080</i>	0.237	<i>0.029</i>
Decision Tree	0.639	<i>0.025</i>	0.148	<i>0.034</i>	0.432	<i>0.104</i>	0.213	<i>0.032</i>
SVM	0.616	<i>0.018</i>	0.177	<i>0.043</i>	0.319	<i>0.039</i>	0.224	<i>0.033</i>
ANN	0.582	<i>0.021</i>	0.156	<i>0.026</i>	0.235	<i>0.057</i>	0.184	<i>0.026</i>
Random Forest	0.509	<i>0.032</i>	0.241	<i>0.057</i>	0.271	<i>0.094</i>	0.238	<i>0.032</i>

These results indicate that, given the same data and initial set of features, logistic regression performs as well or better than the majority of other classifiers. Logistic regression achieves an AUC that exceeds the other methods by 7% on average. T-tests demonstrate that this is a significant improvement over the methods tested ($p = 0.008$ for decision tree, $p = 7.49 \times 10^{-9}$ for SVM, $p = 1.37 \times 10^{-13}$ for neural network, and $p < 2.2 \times 10^{-16}$ for random forest). While the random forest classifier achieves the highest precision score, logistic regression achieves significantly higher precision than the decision tree ($p = 2.23 \times 10^{-5}$) and neural network ($p = 0.003$) methods. The logistic classifier achieves the highest recall, outperforming the other methods by 46% on average. This is a significant improvement over the SVM, neural network, and random forest classifiers ($p < 2.2 \times 10^{-16}$ each). Logistic also achieves the highest F1 score, with significantly higher scores than the decision tree ($p = 8.81 \times 10^{-8}$), SVM ($p = 0.003$), and neural network ($p < 2.2 \times 10^{-16}$) methods. Across the metrics evaluated, logistic

regression also has low standard deviation, suggesting that its predictive power is fairly stable when trained and tested on different splits of the data.

In addition to achieving the best overall performance in this preliminary test, logistic regression has the advantage of being easily interpreted. Interpretability is paramount in real-world sports recruiting applications according to Jake Loos, director of basketball analytics for the NBA's Phoenix Suns:

You can create any statistic you want — but if you're not able to clearly explain what it is and how it can be utilized to improve the team, then the statistic is inconsequential... Researchers need to take a step back and think about how their work could be communicated to people who think Random Forest is a national park in California. (Eastwood, 2018)

Interpretability is also useful when comparing findings with previous research on HR decommitments and examining the effects of specific features on the estimated likelihood of decommitment. Based on this preliminary test, I present results based on logistic regression for the remainder of the chapter.

In order to evaluate the effect of social networks on decommitments, I construct five logistic models with different combinations of feature groups. Model 0 (the baseline model) uses only data on the athlete, his college options, and his recruiting activities. Because some of these baseline features may be correlated to each other, I apply lasso regression ($C=0.1$) to perform feature selection. Model 1 adds to the baseline the features related to Twitter "friends," i.e., the coaches, recruits, and current college players followed by the athlete in the month prior to prediction. Model 2 adds the features tracking "followers" in an athlete's online social network to the baseline, and Model 3

includes the features related to network diffusion. To create Model 4, I combine all features from Models 0-3 and then perform feature selection again with lasso regression to combat collinearity and spurious effects. In order to compare the coefficients and performance across different models, Models 0 and 4 are refit without regularization.

As with the preliminary test of different classification algorithms, stratified Monte Carlo cross validation is used to assess the predictive performance of each model. Over 100 trials, the full data is randomly split into two equal subsets, each with the same proportion of positive and negative instances and ensuring that instances corresponding to the same individual were kept together. After training on oversampled data, each model is applied to the hold-out testing set. I evaluate predictive performance on the using standard metrics: AUC, precision, recall, and F-1 score.

4.3 Results

This study seeks to predict turnover in college football recruiting using institutional data, recruiting activities, and social media. The following sub-sections report the results of the different logistic regression models, compare their predictive performance, and present an example decision support application for the proposed predictive model.

4.3.1 Factors Related to Decommittment

Each model is fit using the full set of 7,148 instances to examine the marginal effect of each feature. By applying the exponential function to the coefficients, we can estimate the effect of each variable on the odds of decommitment. Table 16 lists the coefficients and significance for each feature across the five models as well as the pseudo R-squared.

Table 16 Fitted decommitment regressions

Feature	Model 0		Model 1		Model 2		Model 3		Model 4	
intercept	-1.5812	***	-1.6692	***	-1.6146	***	-1.5089	***	-1.6395	***
prior_decommit	0.8555	***	0.8006	**	0.7672	**	0.8627	***	0.7711	**
in_state	-0.3664	**	-0.4014	**	-0.4100	**	-0.3975	**	-0.4117	**
us_news	-0.0002		-0.0000		0.0000		-0.0004		0.0000	
AP	0.0034		0.0040		0.0039		0.0028		0.0040	
current_win	-0.0040		-0.0042		-0.0036		-0.0043		-0.0035	
commits	-0.0362	**	-0.0255	°	-0.0273	*	-0.0219		-0.0261	°
coach_change	0.9091	***	0.9312	***	0.8349	***	0.7990	***	0.8430	***
sanctions	0.0985		0.0778		0.0635		-0.0197		0.0571	
first_offer	-1.1874	***	-1.1203	***	-1.1361	***	-1.1578	***	-1.1188	***
unofficial	-0.1889	*	-0.2122	*	-0.2142	*	-0.1876	*	-0.208	*
official	-0.3619	*	-0.3053	°	-0.2517		-0.3779	*	-0.2701	
other_offer	-0.0117		-0.0136		-0.0070		-0.0145		-0.0051	
other_closer	-0.0007		0.0011		0.0006		-0.0044		0.0003	
other_us_news	0.0373		0.0433		0.0464		0.0513		0.0458	
other_sanctions	-0.0786		-0.0637		-0.0529		-0.1118		-0.0534	
offers_post	0.1556	**	0.1497	**	0.1294	*	0.1427	**	0.1263	*
unofficial_post	0.4179	***	0.4017	***	0.4040	***	0.4278	***	0.3985	***
coach_visit_post	0.3142	*	0.2797	°	0.2022		0.3387	*	0.2113	
official_post	0.7519	***	0.6595	***	0.4936	***	0.7291	***	0.5023	***
days_to_NSD	-0.0138	***	-0.0150	***	-0.0167	***	-0.0127	***	-0.017	***
coach_friends			-0.0431							
recruit_friends			-0.2427	***					-0.1739	*
college_friends			-0.1288						-0.0707	
coach_friends_other			0.0180							
recruit_friends_other			0.0694	***					0.0068	
college_friends_other			0.1590	*					0.1553	°
coach_followers					-0.0992				-0.0807	
recruit_followers					-0.1970	**			-0.0566	
college_followers					-0.0988					
coach_followers_other					0.3122	***			0.3073	***
recruit_followers_other					0.0902	***			0.0844	**
college_followers_other					0.2368	°			0.1235	
committed							0.0784	***		
decommitted							-0.0371	***		
Pseudo R-squared	0.1326		0.1516		0.1687		0.1400		0.1715	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ° $p < 0.1$

Model 0 contains features that relate to the athlete’s individual characteristics, cost/benefit factors impacting satisfaction with his current school, the availability and relative attractiveness of alternatives. After applying lasso regression, Model 0 is reduced to 20 features. While most of the personal predictors are removed during this feature selection process, a previous decommitment is associated with a 135% increase in the odds of decommitment. As hypothesized, features that decrease costs and increase

benefits of attendance at the recruit's current college are associated with a decreased likelihood of decommitment. The recruit's odds of decommitment decrease by approximately 31% when he is committed to a college in his home state. Interestingly, the number of committed recruits is negatively correlated to decommitment. While one might expect that a full recruiting class would encourage an athlete to decommit due to increased competition for playing time, each additional commit is associated with a 4% decrease in odds of decommitment. One possible explanation for this result is that a solid recruiting class may be linked to the quality and stability of a football program. Features that represent decreased athletic capital benefits are associated with an increased likelihood of decommitment. Commitment to a team that has experienced a recent head coaching change has the largest impact on likelihood of decommitment, increasing the odds by 148%. Recruiting activities are also strong predictors of decommitment intentions. Each official and unofficial visit to the original commitment school is associated with a 17% and 30% decrease in the odds of decommitment, respectively. As shown in Chapter 3, the sequence of recruiting events may be related to the athlete's satisfaction with his current school. Specifically, the evidence suggests that recruits have an emotional connection to the first school that offered a scholarship. Indeed, the odds of leaving the original school decrease 69% when it was the first offer.

The regression results for Model 0 also indicate that the availability and attractiveness of alternatives is associated with increased likelihood of decommitment. Especially notable are the features tracking recruiting events after the athlete has announced a commitment. Each scholarship offer received from another school after committing is associated with a 17% increase in the odds of decommitment. The odds

increase by 52% for each unofficial visit to a competing school after committing and 87% for each official visit. Furthermore, each coach visit from another school after announcing a verbal commitment is associated with a 37% increase in the odds of decommitment. In an unexpected result, the recruit's odds of decommitment increase 1% for each day closer to National Signing Day. While one might expect that the limited time to find a new team would discourage decommitments, this result could be related to coaches' last-minute efforts to recruit athletes committed to other colleges in order to fill out their rosters. Referring to controversial recruiting practice of "poaching" recruits, Tommy Bowden, then-head coach at Clemson, stated "Especially in this part of the country...no means go," (as cited in Staples, 2012). Finally, several features of this model were not significant, although their coefficients show the expected signs.

Model 1 adds features related to out-links to Model 0, focusing on the number of Twitter friends added or dropped by the recruit during the previous month. Consistent with the theory of network realignment, following accounts associated with the original school decreases the odds of decommitment. Each new friendship with a recruit committed to the same school is associated with a 22% decrease in the odds. Conversely, following accounts associated with other schools is associated with an increased likelihood of decommitment. For each new out-link to a committed recruit or current football player at another college, the athlete's odds of decommitting increase 7% and 17%, respectively. These results indicate that recruits' online connections reveal information that can be used to predict decommitments. The features tracking friendships with coaches are not significant, although their coefficients have the expected signs.

Model 2 adds features measuring social recruiting efforts to the baseline, specifically the number of new Twitter followers associated with the commitment school and other schools. As hypothesized, an increase in social media followers from the commitment school is associated with a decreased likelihood of decommitment. Each new recruit following the athlete in the prior month decreases the odds of decommitment by 18%. Additional followers from other schools are associated with increased odds—37% for each coach, 9% for each recruit, and 27% for each current player. These findings indicate that, in addition to the recruit’s own actions online, the actions of individuals associated with his current school or other schools may be related to his decommitment intentions. While the other features tested in this model are not significant, their coefficients carry the expected signs.

Model 3 investigates diffusion, adding features tracking the behavior of other recruits in the athlete’s social network to Model 0. Each reciprocated friend who has decommitted increases the recruit’s odds of decommitting by 8%. Each reciprocated friend who is currently committed decreases the odds of decommitment by 4%. These results suggests that recruits may be influenced by the behavior of their peers and that the decision to decommit can be viral in a social network.

Model 4 adds the social network features tested in Models 1, 2, and 3 to the baseline. Applying lasso regression again to account for potential collinearity issues and spurious effects reduces model to the 29 features shown in Table 16 (20 baseline features based on recruiting and institutional data and nine social-media-derived network features). The coefficients of the final model suggest that, after considering personal, organizational, and environmental factors, social network features do have a significant

impact on likelihood of decommitment. For example, each new friendship with a fellow recruit from the original commitment school is associated with a 16% decrease in the odds of decommitment. The odds increase 17% for each new friendship with a current college player at another school. Additionally, the odds of decommitment increase 36% for each coach and 9% for each recruit from another school following the athlete in the prior month. Model 4 also excludes the both the diffusion features of Model 3. This result suggests that the structure of the recruit's social network (the number and affiliation of his social network neighbors) may reveal more information about his preferences and be more useful for predicting decommitments than the behavior of his peers.

4.3.2 Predictive Performance

Stratified Monte Carlo cross validation is used to assess the predictive performance of each model. I compare the performance of the baseline model containing only features derived from recruiting and institutional data (Model 0) with models incorporating social media data (Models 1-4).

Figure 6 displays the predictive performance of the five different models, averaged over 100 trials.

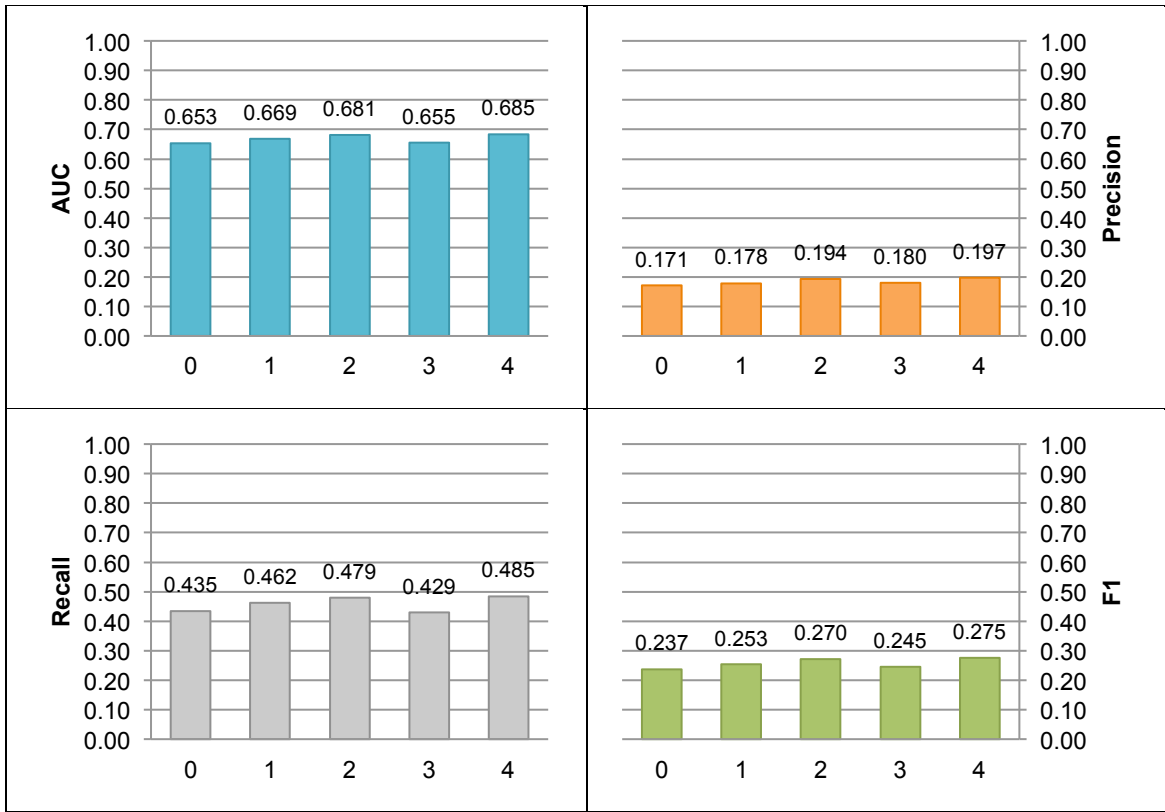


Figure 6 Performance of decommitment predictions

Models 1, 2, 3, and 4 dominate the baseline in terms of AUC, precision, and F1 score, demonstrating that social network features improve the performance of the baseline model. Among the individual sets of social network features (Models 1, 2, and 3), the model tracking in-links from coaches, recruits, and current college athletes demonstrates the largest performance increase relative to the baseline. Model 2 achieves a 4.3% increase in AUC ($p < 2.2 \times 10^{-16}$), 13.4% increase in precision ($p = 9.47 \times 10^{-5}$), 10.1% increase in recall ($p = 1.77 \times 10^{-5}$), and 14.1% increase in F1 score ($p = 1.17 \times 10^{-13}$) over the baseline. Model 1, tracking out-links, makes significant gains on the baseline in terms of AUC ($p = 2.66 \times 10^{-12}$), recall ($p = 0.009$), and F1 score ($p = 2.71 \times 10^{-5}$). Model 3, with features tracking behavioral diffusion, shows no significant improvement over the baseline. These results further suggest that features

focusing on the recruit's social network structure add more value to decommitment predictions than those related to diffusion and social influence.

Ultimately, Model 4 is the top performer. It achieves a 4.8% improvement in AUC ($p < 2.2 \times 10^{-16}$), 15.3% improvement in precision ($p = 8.11 \times 10^{-6}$), 11.4% improvement in recall ($p = 2.05 \times 10^{-6}$), and 15.9% improvement in F1 score ($p < 2.2 \times 10^{-16}$) over the baseline with only recruiting and institutional data.

Furthermore, the combined model achieves significantly higher AUC scores than Models 1-3, as well as significantly higher precision, recall, and F1 scores than Models 1 and 3.

The performance of Model 4 suggests that a combination of features measuring different aspects of online social network structure may be more useful for predicting decommitments than any individual set of network features.

Although oversampling improves the AUC and recall scores for the models, it also leads to decreased precision. The models generate between 685 and 828 false positives when trained on oversampled data. However, because these models predict both the occurrence and timing of an athlete's decommitment, it is possible that some of these errors are situations where they correctly predict that an athlete will decommit, but during the wrong month. Indeed, 17-20% of the false positives across the different models are the result of predicting a decommitment too early. For example, Model 4 predicted that Kevin Harmon, a wide receiver committed to the South Carolina, would decommit in November when in fact he decommitted in December. In the direct use case of athletic recruiting, early warnings are likely to be a welcome result, giving coaches more time to salvage a vulnerable commitment or recruit a replacement athlete. Similarly, early warning of an employee's turnover intention can give a manager valuable lead time to

compose a counter offer or adjust workflow. Thus, the performance measures may actually underestimate the true utility of the proposed decommitment model.

4.3.3 Application to Recruiting Decision Support

In addition to demonstrating the utility of social media-derived network features for predicting turnover, this work is intended to provide practical decision support for recruiters. Table 17 displays a sample report for the University of Utah, tracking each committed recruit’s predicted probability of decommitment as of November 1, 2015. Note that because of the large effect of head coach turnover on predicted likelihood of decommitment, I selected a program that did not experience a head coaching change during the 2015-2016 season.

Table 17 Decommitment predictions for Utah recruits (November 2015)

Name	Star	Position	Commit Date	P(Decommit)	Decommit?
Lorenzo Neal	3	Defensive Tackle	8/10/15	0.725	Y*
Jay Griffin	3	Wide Receiver	6/29/15	0.701	Y*
Devontae Henry-Cole	3	Running Back	7/11/15	0.635	Y
Micah Croom	3	Safety	12/21/14	0.501	Y**
Demari Simpkins	3	Wide Receiver	8/2/15	0.485	N
Ohaji Hawkins	3	Safety	8/30/15	0.390	N**
Daevon Vigilant	3	Running Back	6/26/15	0.197	N**
Alema Pilimai	3	Athlete	10/1/2015	0.193	N**
Kahi Neves	3	Quarterback	10/6/14	0.186	N**
Tucker Scott	3	Offensive Tackle	3/30/15	0.124	N
Cole Fotheringham	3	Tight End	7/6/15	0.120	N
RJ Hubert	3	Wide Receiver	10/1/2015	0.107	N

* *Decommitted in November*, ** *Decommitted later*

According to Model 4 (which combines both recruiting data and social network features), Lorenzo Neal, Jay Griffin, Devontae Henry-Cole, and Micah Croom were predicted to decommit during the month of November. For Griffin, his likelihood of decommitment was increased by the fact that he was an out-of-state commit who had not taken any visits to Utah. Additionally, he had been followed by five recruits from other schools during the month of October, and had not followed any additional Utah coaches,

current players, or recruits. Similarly, Neal was an out-of-state recruit who unfollowed one Utah coach and six recruits, followed five new coaches from other schools, and was followed by two recruits from other schools in October. Armed with the information that both recruits were more than 70% likely to decommit in the next month, it would not be unreasonable for a coach to conclude that Griffin and Neal were “lost causes,” and turn the attention of his recruiting staff toward securing replacement players. Both Griffin and Neal did in fact decommit in November, eventually signing with New Mexico and Purdue, respectively.

Because the logistic classifier provides both predicted outcomes and probabilities, it offers coaches more nuanced information to shape recruiting and retention strategies. Henry-Cole and Croom’s predicted probabilities of decommitting in November were lower than Griffin and Neal’s. In this situation, a coach might decide to hedge his bets, simultaneously taking actions to improve the relationship and discreetly pursuing other recruits. While Croom did not decommit in November as predicted, he did decommit in December, signing with Dartmouth. Henry-Cole remained committed to Utah. Thus, our model correctly predicted 10 of the 12 recruits’ chances of decommitting in November, with one false positive simply being an early warning.

A coach could also use a report like this to note recruits who are not expected to decommit in the next month, but whose predicted probability of decommitment is still relatively high. For example, Model 4 predicts that Ohaji Hawkins had a 39% chance of decommitting in November. Hawkins did eventually decommit in February, signing with Eastern Michigan. A coach could use information about which recruits may be on the cusp of decommitting to take actions intended to salvage the commitment, such as

increasing communication and/or visiting the recruit. A risk-averse coach might even be concerned by Daevon Vigilant, Alema Pilimai, and Kahi Neves, each of whom was predicted to have ~20% probability of decommitment. Indeed, these three athletes did end up decommitting in the following months. While this study uses the default 50% threshold to classify decommitments (instances where the predicted probability is greater than 50% are classified as decommitments) coaches could adjust this parameter as desired. Ultimately, the threshold at which a coach determines a commitment to be vulnerable may depend on his individual preferences or on other factors, such as the value of the recruit, playing position, remaining time until National Signing Day, or personal relationship, to name a few. This report is intended to illustrate how the current model, incorporating both recruiting and social media to predict decommitments over time, may assist coaches in shaping recruiting strategies.

The proposed model could also be used by coaches to monitor changes in athletes' predicted probability of decommitment over time. For each of the 10 Utah recruits who were committed as of October 2015, Figure 7 displays their predicted probabilities during the final months of the recruiting season. The timelines for athletes who decommitted are visualized as red lines, and athletes who remained committed as gray lines.

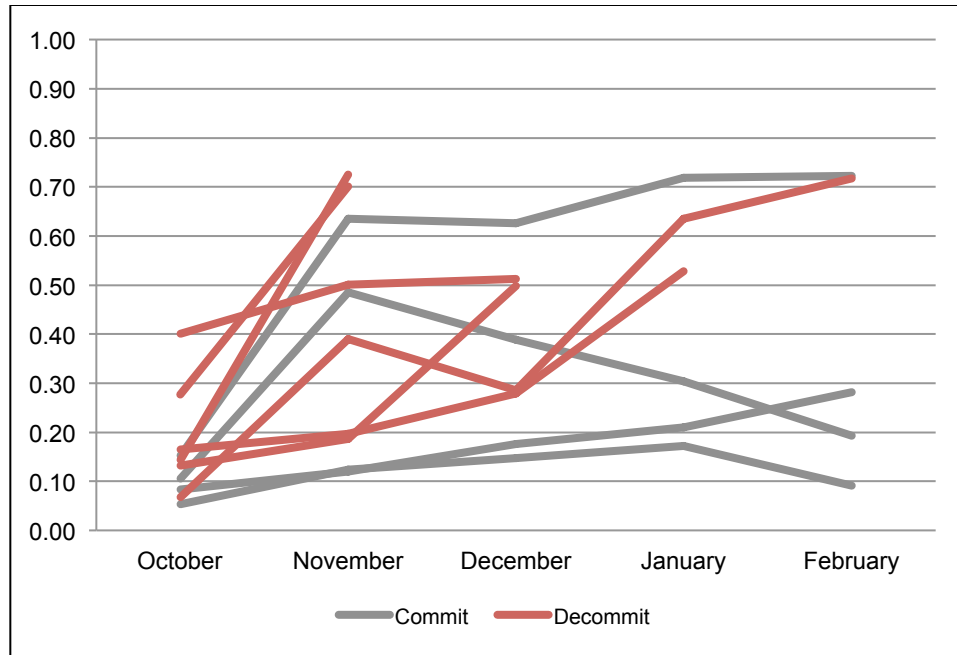


Figure 7 Trends in predicted probability (October 2015-February 2016)

Recruits who remained committed tended to have decreasing, or only slightly increasing, trends in probability of decommitment, with most predictions under 50% likelihood of decommitment. An exception was Devontae Henry-Cole, who had a large jump in predicted likelihood of decommitment between October and November, but then plateaued. Recruits who eventually decommitted tended to show strongly increasing trends, with more predictions exceeding 50% likelihood of decommitment. Thus, a coach could use the decommitment model not only to produce a static prediction for a given month, but also monitor changes in recruits' predicted probability over time, paying special attention to large increases over the course of one or more months. In a realistic application, data could be gathered to produce new predictions on a bi-weekly or weekly basis, giving recruiters up-to-date information.

4.4 Discussion

This research makes a unique contribution to literature on social networks and organizational turnover. I investigate the extensibility of established predictors of personnel turnover to the athletic domain, and my results support the importance of satisfaction and perceived alternatives. Features representing the costs and benefits of enrollment at the original prediction school and comparisons to other schools that have offered scholarships to the athlete significantly predicted the likelihood of decommitment. However, I find mixed evidence for the effectiveness of athletes' personal characteristics, as only the feature tracking whether the athlete had decommitted in the past was a significant predictor. This may be due to taking a dynamic approach to predicting decommitments; static features like star rating or height may do very little to predict the changing odds over time. In general, results were consistent with the satisfaction/alternatives framework, and such theoretical basis suggests that the findings of this work may be generalizable to other organizational settings.

This study explores the value of social network features for predicting turnover in the context of American college football recruiting. I take a unique approach to exploring the intersection between social networks and turnover by utilizing social media data. Although one's online connections are an imperfect proxy for offline social ties, I find that—after considering personal, organizational, and environmental factors—online social network features have a consistent and significant impact on the likelihood of decommitment. This result aligns with previous research on workplace social networks (e.g., Feeley et al., 2008).

Performance testing indicates that social media-derived network features consistently add value to predictive models. Model 4, which includes a combination of features measuring different aspects of social network structure, is the highest performer, achieving a 5% improvement in AUC, 15% improvement in precision, 11% improvement in recall, and 16% improvement in F1 score over the baseline model using only recruiting and institutional data. This finding is especially significant in light of the prohibitive cost of tracking offline social networks. Social network data may be retrieved from social media websites easily and in large quantities, offering more opportunities to perform holistic network analyses.

Additionally, I find that among the three groups of network features explored, those focused on structural aspects of the online social network (Models 1 and 2) contribute more to predictive performance than those based on network diffusion (Model 3). Thus, the effect of turnover contagion (Felps et al., 2009) was not supported in the context of athletic decommitments. However, a more detailed analysis may yield different results. Specifically, focusing on the behavior of recruits from the same high school or committed to the same college provides an opportunity to further explore the question of decommitment diffusion in future work.

As expected, making new connections to individuals from the recruit's original school is associated with decreased odds of decommitment, while new ties to other schools are associated with increased odds. In addition to considering the affiliation of the recruit's online social connections, I also consider the type (coach, recruit, college football player). My results suggest that suggests that forming new out-links to recruits and current players may be more predictive of commitment strength than connections

with coaches, as the lasso procedure removes both features tracking out-links to college coaches. This effect could also be due to the time range of the data; athletes may be more likely to follow coaches earlier in the recruiting process. Yet I find a different effect for followers, where both features tracking in-links from current college football players are removed in the final model. Only the features related to in-links from coaches and recruits at other schools were significant. Again, this result may be influenced by the time range of the analysis; at the tail end of the recruiting process, it is possible that all of the individuals from the athlete's original school have already followed him.

This work represents both a promising first step in predicting decommitments in college football and using online social network data to explain and predict turnover in other organizational settings. I find that recruits' online social networks are strong predictors of decommitments, and contend that recruiters in other domains should also consider the information conveyed by candidates' online social networks. Moreover, this study contributes to more broadly to the field of information science by investigating how the information created by individuals on social media can provide insight into their offline decisions and the decisions of others.

CHAPTER 5 CONCLUSIONS AND FUTURE WORK

This dissertation explores how the information people reveal about themselves online can be analyzed to understand and predict their decisions in the offline world as well as how this information can diffuse and influence the decisions of others. Specifically, I examine the complex interactions between online social media and offline recruiting outcomes, using data from American college football recruiting. In addition to contributing to the growing body of work on sports analytics, my work represents a unique addition to the fields of personnel psychology and management by demonstrating the value of social media data for modeling and predicting recruiting decisions. This research is the first to document the use of social media by athletes and coaches during the recruitment process, making a novel contribution to the information science literature, which has previously described the social media usage of different age groups (e.g., Xie et al., 2012) and professions (e.g., Nentwich & Konig, 2013).

The first study focuses on the relationship between a recruit's Twitter content and the coach's decision to extend a scholarship offer. Informed by impression management theory (Goffman, 1959), I construct classifiers to identify instances of self-promotion and ingratiation in recruits' tweets. I find that athletes tended to engage in more self-promotion (19.2% of tweets) than ingratiation (13.9%). These results are consistent with previous research on impression management during in-person employment interviews (e.g., Stevens & Kristof, 1995), suggesting the opportunity for fruitful borrowing between the HR and athletic domains.

The labeled tweets are then used to construct explanatory models of the number of offers received in the next month. Compared to a baseline model using only

demographic and recruiting data, both models tracking online impression management are a significantly better fit, but Model 1, which focuses on self-promotion, is a marginally better fit than the ingratiation model (Model 2). Furthermore, a logistic classifier that considers the recruit's personal characteristics and recruiting activities to-date as well as his social media activities during the prior month achieves 74% accuracy predicting whether he will receive an offer in the next month. These results are noteworthy, suggesting that information created on social media by college football recruits may impact the decisions of coaches during recruitment. Previous work on selection processes in athletics and HR has indicated that recruiters in both domains are increasingly turning to social media as a cost-effective and convenient source of information on candidates, and this work fills a significant gap in our understanding of the offline impacts of online behaviors (Roth et al., 2016).

Yet even more important than the content posted by athletes was simply possessing a Twitter account. Indeed, athletes with public or private social media profiles had a significant advantage over their peers without an online presence in terms of attracting scholarship offers. These results provide evidence that social media is not changing the recruiting game, but rather, has already changed it.

The second study examines the athlete's school choice decision. I extend prior research that focused on a cost/benefit approach to commitment decisions (e.g., Dumond et al., 2008), by constructing a logistic classifier that includes features related to the availability heuristic. Averaged over 100 trials, this baseline model achieves an AUC of 0.659, a 3% improvement over the purely same model without the heuristic features. This work is the first to explore heuristics in school in athletic school choice, and my results

suggest that both football coaches and HR professionals may benefit from considering cognitive biases and heuristics for better understanding job choice decisions.

This study also investigates social factors in school choice and specifically assesses the value of social media data for predicting commitments. Comparing the baseline with only institutional and recruiting data to models incorporating different sets of social media features, the results demonstrate that features tracking out-links, in-links, and hashtag content contribute more to predictive performance than features recording interactions (replies, retweets, mentions) between recruits and colleges. Model 4, which combines offline recruiting data with online social media data, achieves the highest predictive performance (AUC=0.701). This research represents a novel contribution to the athletic recruiting literature, and my findings highlight the value of online social media data for predicting offline recruiting outcomes in other contexts.

The third study examines the problem of athletic decommitments. Informed by extant theories of personnel turnover (e.g., March & Simon, 1958), I first consider the role of satisfaction and perceived alternatives and build a logistic classifier predicting the occurrence of decommitments over time using institutional data and recruiting data from 247Sports. The baseline model achieves modest success (AUC=0.653), but provides preliminary evidence that the HR turnover framework can be extended to the athletic domain. This work is the first to build a predictive model of decommitments and represents a significant addition to the sports analytics literature.

I also assess the value of social media data for predicting decommitments by comparing the baseline to models tracking different aspects of the athlete's online social network (out-links, in-links, behavioral diffusion). In comparison to previous work on

personnel turnover, which has primarily focused on networks within a single company (e.g., Feeley et al., 2008), I consider both intra- and inter-organizational connections. The results indicate that forming new connections to the current college is associated with decreased likelihood of decommitment, while stronger external networks are linked to increased likelihood. I also find that structural features add more value to decommitment predictions than features tracking the commitment status of the recruit's peers. The combined model with institutional data, recruiting data, and social media data achieves the highest performance (AUC=0.685), suggesting that a combination of features measuring different aspects of social network structure is more successful than any individual set of social network features.

I note some potential limitations of this research. First, the accuracy of the recruiting data utilized in this dissertation is dependent on its source. 247Sports incorporates both user-supplied data (athletes, coaches, and other users may request a profile page or report updates) and expert data (recruiting professionals compile rankings and curate the site), which I contend makes it both a comprehensive and up-to-date source of recruiting information. Second, the project uses data on a single recruiting class, and results may differ for other recruiting years. Third, because of the scope of my data collection, I was only able to gather detailed network data for the last six months of recruitment. This resulted in being unable to consider the relationship between Twitter connections and offers in Chapter 2, and the analyses of Chapters 3 and 4 were focused on the final months of recruitment. It is possible that predicting earlier commitments and decommitments may yield different results. Fourth, my commitment and decommitment predictions were limited to athletes with public Twitter profiles, and may not be

generalizable to athletes without a presence on social media. Finally, as these were not experimental studies, no causality can be inferred from the results. My findings indicate that, in addition to data readily available from recruiting websites like 247Sports, athletes' online relationships and content may be useful for predicting offline recruiting outcomes. While these findings are consistent with existing social network and communications theories, I cannot state that these variables cause offers, commitments, or decommitments.

There are several interesting directions for future work building on the data already collected. Chapter 2 focuses specifically on the relationship between self-promotion, ingratiation, and offers, and further investigation into other types of online self-presentation may prove fruitful, especially the impact of negative behavior (e.g., cursing, sexual content) on scholarship offers. Currently, the offer analysis is school-neutral, aimed at modeling and predicting new offers from any college team. However, both college football recruits and job candidates may be more interested in the question of whether they will receive an offer from a specific organization. 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. Similarly, impression management is often tied to the concept of person-organization and person-job fit (Kristof-Brown, Barrick, & Franke, 2002). Deeper analysis of tweet content, such as estimating the similarity between coaches' and recruits' tweet topics, may provide insight into the function of Twitter as an information source for athlete-school fit. Additional analysis of recruits' and coaches' demographic characteristics and online social networks may also prove useful for investigating issues

of fit. Furthermore, it is likely that recruits of different skill levels may utilize significantly different social media strategies during recruitment, i.e., low-rated recruits may expend more effort on self-promotion in order to garner coaches' attention. Examining differences in Twitter activity and impression management by star rating presents a potentially fruitful extension of the current study.

In Chapters 3 and 4, my analyses primarily utilize information about athletes and the schools that are recruiting them, but the influence of recruiter characteristics on applicant job choice intentions has been studied in the HR domain (e.g., Alderfer & McCord, 1970; Wyse, 1972). 247Sports also tracks data on coaches—including their alma mater, job history, position, salary, and age—that could be used to investigate the effect of coach characteristics and athlete-coach fit on recruiting outcomes. My commitment and decommitment predictions also focus on connections to college coaches, current college football players, and other recruits, but further analysis of centrality, connectivity, and community structure in the Twitter network may provide insight into the role of social networks during recruitment. 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 factors than lower-ranked recruits, and Popp et al. (2011) compared the college selection process for international and domestic student-athletes as well as differences by gender, finding that female athletes rated academic factors significantly higher than males. Looking at differences in school choice decisions and decommitment decisions between early commits and late commits as well as between athletes at different positions (quarterback, kicker, etc.) or star ratings would be a

promising extension of this research. In addition, expanding the analysis of decommitments to include instances of involuntary turnover, i.e., having a scholarship offer rescinded, would be an interesting future direction. This work could also form the starting point for predicting transfers or on-field performance outcomes. The decision support aspect could also be further developed with an intention toward generating prescriptive analytics.

While this dissertation focuses on using online social media data to model and predict offline outcomes in college football recruitment, my approach could be easily be applied to other sports and, ultimately, to personnel recruiting. The proliferation of business networking platforms such as LinkedIn has made it easier to gather large amounts of data on individuals' educational backgrounds, job histories, and inter-firm mobility. This presents a promising opportunity for future analysis of real hiring data.

Overall, this research takes a small step toward answering larger questions in the field of information science about the offline consequences of online social media use. Though the power of social media data to predict both significant (e.g., Franch, 2012), and trivial (e.g., Zhang et al., 2015) outcomes has been demonstrated in the information science literature, my work is one of the first to focus on recruitment, an application area with potentially life-changing consequences. Across three case studies, I find preliminary indications that social media has become an indispensable tool for professional branding and networking in college athletics, and that a recruit's actions on social media may significantly influence others' perceptions of him. Additionally, my results suggest that a recruit's connections and content on social media predict his likelihood of joining (or leaving) an organization. While this dissertation demonstrates only one approach to

analyzing the information that individuals reveal about themselves on social media and its relationship to their offline decisions and the decisions of others, my results underscore the growing importance of social media on recruitment processes in athletics and beyond.

REFERENCES

- Alderfer, C., & McCord, C. (1970). Personal and situational factors in the recruitment interview. *Journal of Applied Psychology* 54, 377-385.
- Allen, D.G. (2006). Do organizational socialization tactics influence newcomer embeddedness and turnover? *Journal of Management*, 32(2), 237-256.
<http://doi.org/10.1177/0149206305280103>
- Asur, S., & Huberman, B.A. (2010). Predicting the future with social media. In X.J. Huang, I. King, V. Raghavan, & S. Rueger (Eds.) *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01* (pp. 492-499). Washington, DC: IEEE Computer Society. <http://doi.org/10.1109/WI-IAT.2010.63>
- Auger, P., Devinney, T.M., Dowling, G.R., Eckert, C., & Lin, N. (2013). How much does a company's reputation matter in recruiting? *MIT Sloan Management Review* 54(3), 79-88.
- Baade, R.A., & Sundberg, J.O. (1996). Fourth down and gold to go? Assessing the link between athletics and alumni giving. *Social Science Quarterly* 77(4), 789-803.
- Barden, J.Q., Bluhm, D.J., Mitchell, T.R., & Lee, T.W. (2013). Hometown proximity, coaching change, and the success of college basketball recruits. *Journal of Sport Management* 27, 230-246.
- Behera, B., & Pathy, M. (2013). Employee referrals—The best leveraged talent acquisition strategy amid recession. *IOSR Journal of Business and Management* 14(1), 1-10.

- Bella, A., Colladon, A.F., Battistoni, E., Castellan, S., & Francucci, M. (2018). Assessing perceived organizational leadership styles through Twitter text mining. *Journal of the Association for Information Science and Technology* 69(1), 21-31.
<http://doi.org/10.1002/asi.23918>
- Bergman, S.A., & Logan, T.D. (2014). The effect of recruit quality on college football team performance. *Journal of Sports Economics* 17(6), 578-600.
<http://doi.org/10.1177/1527002514538266>
- Berkowitz, S., Schnaars, C., Hannan, M., Alewine, M., Brugh, J., Harris, R., Hunt, J., Perez, C.M., Rodriguez, L.U., Wu, Z., & Thomas, K. (2016). NCAA Finances. *USA Today*. Retrieved from <http://sports.usatoday.com/ncaa/finances>.
- Bhattacharya, S., Yang, C., Srinivasan, P., & Boynton, B. (2015). Perceptions of presidential candidates' personalities in Twitter. *Journal of the Association for Information Science and Technology* 67(2), 249-267.
<http://doi.org/10.1002/asi.23377>
- Bian, Y. (1997). Bringing strong ties back in: Indirect ties, network bridges, and job searches in China. *American Sociological Review* 62(3), 366-385.
- Bogaert, M., Ballings, M., & Van den Poel, D. (2015). The added value of Facebook friends data in event prediction. *Decision Support Systems* 82, 26-34.
<http://doi.org/10.1016/j.dss.2015.11.003>
- Bohnert, D., & Ross, W.H. (2010). The influence of social networking Web sites on the evaluation of job candidates. *Cyberpsychology, Behaviour, and Social Networking* 13(3), 341-347.

- Brady, E., Kelly, J., & Berkowitz, S. (2015, February 3). Schools in power conferences spending more on recruiting. *USA Today*. Retrieved from <https://www.usatoday.com/story/sports/ncaaf/recruiting/2015/02/03/college-football-recruiting-signing-day-sec-power-conferences/22813887/>
- Bricker, J., & Hanson, A. (2013). The impact of early commitment on games played: Evidence from college football recruiting. *Southern Economic Journal* 79(4), 971-983. <http://doi.org/10.4284/0038-4038-2010.119>
- Brown, R. (1993). An estimate of the rent generated by a premium college football player. *Economic Inquiry* 31, 671-84.
- Browning, B., & Sanderson, J. (2012). The positives and negatives of Twitter: exploring how student-athletes use Twitter and respond to critical tweets. *International Journal of Sport Communication* 5, 503-521.
- Broughton, A., Foley, B., Ledermaier, S., & Cox, A. (2013). The use of social media in the recruitment process. *ACAS*. Retrieved from <http://m.acas.org.uk/media/pdf/0/b/The-use-of-social-media-in-the-recruitment-process.pdf>
- Butts, C. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology* 11, 13-41. <http://doi.org/10.1111/j.1467-839X.2007.00241.x>
- Caers, R., & Castelyns, V. (2010). LinkedIn and Facebook in Belgium: The influences and biases of social network sites in recruitment and selection procedures. *Social Science Computer Review* 29(4), 437-448. <http://doi.org/10.1177/0894439310386567>

- Cable, D. M., & Turban, D. B. (2003). The value of organizational reputation in the recruitment context: a brand equity perspective. *Journal of Applied Social Psychology* 33, 2244–2266.
- Caldwell, D. (2015, October 1). Missteps on social media can ruin recruitment. *Athletic Business*. Retrieved from <http://www.athleticbusiness.com/web-social/missteps-on-social-media-can-ruin-recruitment.html>
- Caro, C.A. (2012). College football success: The relationship between recruiting and winning. *International Journal of Sports Science & Coaching* 7(1), 139-152.
- Carson, C. (2016, March 15) Introducing the 247Sports Decommitment Tracker. *247Sports*. Retrieved from <http://247sports.com/Bolt/Introducing-the-247Sports-Decommitment-Tracker-44280624>
- CBS Sports (n.d.). FB Recruiting. *247Sports*. Retrieved from <http://247sports.com>
- Chapman, D.S., Uggerslev, K.L., Carroll, S.A., Piasentin, K.A., and Jones, D.A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology* 90(5), 928-944. <http://doi.org/10.1037/0021-9010.90.5.928>
- Chow, W.S., & Chan, L.S. (2008). Social network, social trust, and shared goals in organizational knowledge sharing. *Information & Management* 45, 458-465. <http://doi.org/10.1016/j.im.2008.06.007>
- Cotton, J.L., & Tuttle, J.M. (1986). Employee turnover: A meta-analysis and review with implications for research. *Academy of Management Journal*, 11(1), 55-70.

- Crabtree, J. (2016, January 26). The 'social' science of recruiting. *ESPN*. Retrieved from http://espn.go.com/college-football/recruiting/story/_/id/14646545/social-media-becomes-powerful-aide-dangerous-connection-recruiting
- Croft, C. (2008). *Factors influencing Big 12 Conference college basketball student-athletes' selection of a university* (Doctoral Dissertation). Retrieved from Digital Commons@University of Texas at El Paso. (AAI3313419)
- Crossley, C.D., & Highhouse, S. (2004). Relation of job search and choice process with subsequent satisfaction. *Journal of Economic Psychology* 26, 255-268. <http://doi.org/10.1016/j.joep.2004.04.001>
- Davenport, R. (2015, December 29). Under Armour All-American survey. *ArkansasOnline*. Retrieved from <http://www.arkansasonline.com/news/2015/dec/29/under-armour-all-american-survey/>
- Davison, R.M., Ou, C.X.J., Martinsons, M.G., Zhao, A.Y., & Du, R. (2014). The communicative ecology of Web 2.0 at work: Social networking in the workspace. *Journal of the Association for Information Science and Technology* 65(10), 2035-2047. <http://doi.org/10.1002/asi.23112>
- Ding, C., Cheng, H.K., Duan, Y., & Jin, Y. (2016). The power of the 'like' button: The impact of social media on box office sales. *Decision Support Systems* 94, 77-84. <http://doi.org/10.1016/j.dss.2016.11.002>
- Downs, C. W. (1969). Perceptions of the selection interview. *Personnel Administration* 32, 8-23.

- Doyle, C.A., & Gaeth, G.J. (1990). Assessing the institutional choice process of student-athletes. *Research Quarterly for Exercise and Sport* 61(1), 85-92.
<http://doi.org/10.1080/02701367.1990.10607482>
- Dronyk-Trosper, T., & Stitzel, B. (2015). Lock-in and team effects: Recruiting and success in college football athletics. *Journal of Sports Economics* 18(4), 376-387.
<http://doi.org/10.1177/1527002515577885>
- Dumond, J.M., Lynch, A.K., & Platania, J. (2008). An economic model of the college football recruiting process. *Journal of Sports Economics* 9(1), 67-87.
<http://doi.org/10.1177/1527002506298125>
- Eastwood, B. (2017, December 28). *Watch for these sports analytics developments in 2018* [Press Release]. Retrieved from
<http://mitsloan.mit.edu/newsroom/articles/watch-for-these-sports-analytics-developments-in-2018/>
- Elliot, R., & Maguire, J. (2008). Getting caught in the net: Examining the recruitment of Canadian players in British professional ice hockey. *Journal of Sport & Social Issues* 32(2), 158-176. <http://doi.org/10.1177/0193723507313927>
- Elliott, B. (2015, August 4). 8 pieces of advice from parents of star college football recruits. *SB Nation*. Retrieved from <http://www.sbnation.com/college-football-recruiting/2015/8/4/9060917/ncaa-football-recruit-parents>
- Elsass, P. (2001). Trust and team performance in NCAA basketball. *The Academy of Management Executive* 15, 137–138. <http://doi.org/15.5465/AME.2001.5229673>

- Feeley, T.H., Hwang, J., & Barnett, G.A. (2008). Predicting employee turnover from friendship networks. *Journal of Applied Communication Research* 36(1), 56-73.
<http://doi.org/10.1080/00909880701799790>
- Feldman, B. (2007). *Meat market: Inside the smash-mouth world of college football recruiting*. New York, NY: ESPN Books.
- Felps, W., Mitchell, T.R., Hekman, D.R., Lee, T.W., Holtom, B.C., & Harman, W.S. (2009). Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal*, 52, 545-561,
<https://doi.org/10.5465/AMJ.2009.41331075>
- Fernandez, R.M., Castilla, E.J., & Moore, P. (2000). Social capital at work: networks and employment at a phone center. *American Journal of Sociology* 105(5), 1288-356.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations* 7, 117-140.
- Forsythe, S.A. (2015). *The impact of a head coaching change on intercollegiate student-athletes* (Doctoral Dissertation). Retrieved from Digital Commons@Western Kentucky University. (88)
- Franch, F. (2012). Wisdom of the crowds²: 2010 UK election prediction with social media. *Journal of Information Technology & Politics* 10(1), 57-71.
<http://doi.org/10.1080/19331681.2012.705080>
- Gabert, T.E., Hale, J.L., & Montalvo, G.P. (1999). Differences in college choice factors among freshmen student-athletes. *Journal of College Admission* 164, 20-29.
- Goff, B. (2000). Effects of university athletics on the university: A review and extension of empirical assessment. *Journal of Sport Management* 14, 85-104.

- Goffman, E. (1959). *The presentation of self in everyday life*. New York, NY: Anchor Books.
- Goidel, R.K. & Hamilton, J.M. (2006). Strengthening higher education through gridiron success? Public perceptions of the impact of National Football Championships on academic quality. *Social Science Quarterly* 87(4), 851-862.
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology* 78, 1360-80.
- Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3), 481-510. Retrieved from <http://www.jstor.org/stable/2780199>
- Greene, W.H. (1994). *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models* (NYU working paper no. EC-94-10). New York, NY: New York University.
- Griffeth, R.W., Hom, P.W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management* 26, 463-488. <http://doi.org/10.1177/014920630002600305>
- Han, J. (2009). Network-based recruiting and applicant attraction in China: Insights from both organizational and individual perspectives. *International Journal of Human Resource Management* 20(11), 2228-2249. <http://doi.org/10.1080/09585190903239633>

- Helm, S. (2013). A matter of reputation and pride: associations between perceived external reputation, pride in membership, job satisfaction and turnover intentions. *British Journal of Management* 24, 542–556. <http://doi.org/10.1111/j.1467-8551.2012.00827.x>
- Highhouse, S. (1997). Understanding and improving job-finalist choice: The relevance of behavioral decision research. *Human Resource Management Review* 7(4), 449-470.
- Hilgert, R., & Eason, L. (1968). How students weigh recruiters. *Journal of College Placement* 28, 99-102.
- Horn, T. (2015, February 4). The no-star recruit. *The Players' Tribune*. Retrieved from <http://www.theplayerstribune.com/the-no-star-recruit/>
- HRreview. (2013, November 6). 60% of job seekers do not like recruiters searching social media profiles. *HRreview*. Retrieved from <http://www.hrreview.co.uk/hr-news/recruitment/60-of-job-seekers-do-not-like-recruiters-searching-social-media-profiles/49324>
- Huffman, L.T., & Cooper, C.G. (2012). I'm taking my talents to...: An examination of hometown socio-economic status on the college choice factors of football student-athletes at a southeastern university. *Journal of Issues in Intercollegiate Athletics* 5, 225-246.
- Humphreys, B.R., & Mondello, M. (2007). Intercollegiate athletic success and donations at NCAA Division I institutions. *Journal of Sport Management* 21, 265-280.

- Hunsberger, P.K. & Gitter, S.R. (2015). What is a blue chip recruit worth? Estimating the marginal revenue produce of college football quarterbacks. *Journal of Sports Economics* 16(6), 664-690. <http://doi.org/10.1177/1527002515580938>
- Huynh, C.H., & Simon, J. (2016). Using means objectives to present risk information. *Decision Analysis* 13(2), 117-127. <http://doi.org/10.1287/deca.2015.0328>
- Jones, E.E., & Pittman, T.S. (1982). Toward a general theory of self-presentation. In J. Suls (Ed.), *Psychological Perspectives on the Self, Vol. 1* (pp. 231-262). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jones, W.A. (2013). Exploring the relationship between intercollegiate athletic expenditures and team on-field success among NCAA Division I Institutions. *Journal of Sports Economics* 14, 6: 584-605.
<http://doi.org/10.1177/1527002511433469>
- Jowett, S., & Timson-Katchis, M. (2005). Social networks in sport: Parental influence on the coach-athlete relationship. *The Sport Psychologist* 19, 267-287.
- Kacmar, K.M., Delery, J.E., & Ferris, G.R. (1992). Differential effectiveness of applicant impression management tactics on employment interview decisions. *Journal of Applied Social Psychology* 22(16), 1250-1272.
- Khan, S. (2016, May 5). A&M assistant's tweets add fuel to fire of No. 3 QB's decommitment. *ESPN*. Retrieved from http://espn.go.com/college-football/recruiting/story/_/id/15465600/texas-aggies-lose-second-recruit-night-coach-tweets

- Kian, E., & Sanderson, J. (2014). High school football recruiting and Twitter: An unstable mix for self-branding efforts? *The Sport Journal*. Retrieved from <http://thesportjournal.org/article/high-school-football-recruiting-and-twitter-followers-an-unstable-mix-for-self-branding-efforts/>
- Kilduff, M. (1990). The interpersonal structure of decision making: A social comparison approach to organizational choice. *Organizational Behavior and Human Decision Processes* 47, 270-288.
- Kirnan, J.P., Farley, J.A., & Geisinger, K.F. (1989). The relationship between recruiting source, applicant quality, and hire performance: an analysis by sex, ethnicity, and age. *Personnel Psychology* 42, 293-308.
- Klenosky, D., Templin, T., & Troutman, J. (2001). Recruiting student athletes: A means-end investigation of school-choice decision making. *Journal of Sport Management* 15, 95–106.
- Koo, G.Y., & Dittmore, S.W. (2014). Effects of intercollegiate athletics on private giving in higher education. *Journal of Issues in Intercollegiate Athletics* 7, 1-16.
- Krider, J., O'Leonard, K., & Erickson, R. (2015). WhatWorks brief: Talent acquisition factbook. [Fact sheet]. Retrieved from <https://legacy.bersin.com/uploadedfiles/042315-ta-factbook-wwb-final.pdf>
- Kristensen, J.B. Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M., & Bornakke, T. (2017). Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PLOSOne* 12(9), e0184562. <http://doi.org/10.1371/journal.pone.0184562>

- Kristof-Brown, A., Barrick, M.A., & Franke, M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management* 28(1), 27-46.
[http://doi.org/10.1016/S0149-2063\(01\)00131-3](http://doi.org/10.1016/S0149-2063(01)00131-3)
- Lampe, C., & Ellison, N.B. (2010). Student athletes on Facebook. In K. Inkpen, G. Gutwin, & J. Tang (Eds.) *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)* (pp. 193-196). New York, NY: Association for Computing Machinery. <http://doi.org/10.1145/1718918.1718954>
- Langelett, G. (2003). The relationship between recruiting and team performance in division 1A college football. *Journal of Sports Economics* 4(3), 240-245.
- Lebel, K. (2013). *Professional athlete self-presentation on Twitter* (Doctoral Dissertation). Retrieved from Scholarship@Western University. (1303)
- Lewis, M. (2003). *Moneyball: The art of winning an unfair game*. New York, NY: W.W. Norton & Company.
- Lin, N. (2001). *Social capital: A theory of social structure and action*. Cambridge, UK: Cambridge University Press.
- Loper, E., & Bird, S. (2002). NLTK: The natural language toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1 (ETMTNLP '02)* (pp. 63-70). Stroudsburg, PA: Association for Computational Linguistics.
<http://doi.org/10.3115/1118108.1118117>

- Lovejoy, K., & Saxton, G.D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication* 17(3), 337-353. <http://doi.org/10.1111/j.1083-6101.2012.01576.x>
- Lujan, C. (2010). *How well informed are high school student-athletes about post-secondary options?* (Masters Thesis). Retrieved from University of Arizona Campus Repository. (193407)
- Malm, T. (1954). Recruiting patterns and functioning of labor markets. *Industrial and Labor Relations Review* 7(4), 507-525. Retrieved from <http://www.jstor.org/stable/2519101>
- Manning, C.D., Raghavan, P. & Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press. Retrieved from <https://nlp.stanford.edu/IR-book/html/htmledition/irbook.html>
- March, J.G., & Simon, H.A. (1958). *Organizations*. New York, NY: Wiley.
- Marsden, P.V. (1994). The hiring process: Recruitment methods. *American Behavioral Scientist* 37(7), 979-991. <http://doi.org/10.1177/0002764294037007009>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.) *Frontiers in Econometrics* (pp. 105-142). Berkeley, CA: Academic Press.
- McManus, M.A., & Baratta, J.E. (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.
- Meer, J., & Rosen, H. (2009). The impact of athletic performance on alumni giving. *Economics of Education Review* 28(3), 287-294.

- Mirabile, M.P., & Witte, M.D. (2015). A discrete-choice model of college football recruit's program selection decision. *Journal of Sports Economics* 18(3), 211-238. <http://doi.org/10.1177/1527002514566278>
- Mixon, F.G. & Hsing, Yu. (1994). The determinants of out-of-state enrollments in higher education: A tobit analysis. *Economics of Education Review* 13(4), 329-335.
- Montgomery, J.D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review* 81(5), 1408-1418. Retrieved from <http://www.jstor.org/stable/2006929>
- Mossholder, K.W., Settoon, R.P., & Henagan, S.C. (2005). A relational perspective on turnover: Examining structural, attitudinal, and behavioral predictors. *Academy of Management Journal* 48(4), 607-618.
- Moynihan, D.P., & Pandey, S.K. (201). The ties that bind: Social networks, person-organization value fit, and turnover intention. *Journal of Public Administration Research and Theory* 18, 205-227. <http://doi.org/10.1093/jopart/mum013>
- Muthukadan, B. (2011). Selenium with Python. Retrieved from <http://selenium-python.readthedocs.io/>
- Myerberg, P. (2015, July 10). The best quarterback recruits become the best recruiters. *USA Today*. Retrieved from <http://www.usatoday.com/story/sports/ncaaf/2015/07/09/college-football-recruiting-quarterbacks-dwayne-haskins-coaches/29920671/>
- National Collegiate Athletic Association. (n.d.). *Major infractions database*. Retrieved from <https://web1.ncaa.org/LSDBi/>

- National Collegiate Athletic Association. (2015). Estimated probability of competing in college football. Retrieved from <http://www.ncaa.org/about/resources/research/football>
- National Collegiate Athletic Association. (2016). *2015-2016 NCAA Division I Manual*. Indianapolis, IN: National Collegiate Athletic Association.
- National Collegiate Athletic Association. (2018). 2017 football attendance. Retrieved from <http://www.ncaa.org/championships/statistics/ncaa-football-attendance>
- Nentwich, M., & Konig, R. (2013). Academia goes Facebook? The potential of social network sites in the scholarly realm. In Bartling, S., Friesike, S. (Eds.) *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing* (pp. 107-124). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_7
- Parnes, H. (1954). Research on labor mobility: An appraisal of research findings in the United States. *Social Science Research Council Bulletin* 65, 144-190.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825-2830. Retrieved from <http://www.jmlr.org/papers/v12/pedregosa11a.html>
- Pegoraro, A. (2010). Look who's talking—athletes on Twitter: A case study. *International Journal of Sport Communication* 3: 501-514. <http://doi.org/10.1123/ijsc.3.4.501>

- Pegoraro, A., & Jinnah, N. (2012). Tweet 'em and reap 'em: The impact of professional athletes' use of Twitter on current and potential sponsorship opportunities. *Journal of Brand Strategy*, 1(1), 85-97.
- Peng, K., Cooke, J., Crockett, A., Shin, D., Foster, A., Rue, J., ... Rhodes, M. (2018). Predictive analytics for University of Virginia football recruiting. In *2018 Systems and Information Engineering Design Symposium* (pp. 243-248), <https://doi.org/10.1109/SIEDS.2018.8374745>
- Penttila, C. (2009, July 01). Build a social media hiring strategy. *Entrepreneur*. Retrieved from <http://www.entrepreneur.com/article/202466>
- Pew Research Center. (2014). *Social media update 2014*. Retrieved from http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf
- Pitts, J.D., & Rezek, J.P. (2012). Athletic scholarships in intercollegiate football. *Journal of Sports Economics* 13(5), 515-535. <http://doi.org/10.1177/1527002511409239>
- Pope, D.G., & Pope, J.C. (2009). The impact of college sports success on the quantity and quality of student applications. *Southern Economic Journal* 75(3), 750-780.
- Popp, N., Pierce, D., & Hums, M.A. (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://doi.org/10.1016/j.smr.2010.08.003>
- Prunty, B. (2014, December 28). Colleges often entice top prospects by recruiting their mentors, too. *The New York Times*. Retrieved from <http://www.nytimes.com/2014/12/29/sports/ncaabasketball/colleges-often-entice-top-prospects-by-recruiting-their-mentors-too.html>

- Quaglieri, P.L. (1982). A note on variations in recruiting information obtained through different sources. *Journal of Occupational Psychology* 55, 53-55.
- R Development Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>
- Rees, A. (1966). Information networks in labor markets. *American Economic Review* 56, 559–66.
- Roth, P.L., Bobko, P., Van Iddekinge, C.H., & Thatcher, J.B. (2016). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of Management* 42(1), 269-298.
<http://doi.org/10.1177/0149206313503018>
- Rubenstein, A.L., Eberly, M.B., Lee, T.W., & Mitchell, T.R. (2018). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover. *Personnel Psychology*, 71(1), 23-65. <http://doi.org/10.1111/peps.12226>
- Rutter, E. (2015, October 8). 3-Star LB Andre Polk talks Minnesota commit and peer recruiting. *Today's U Sports*. Retrieved from <http://www.todaysu.com/recruiting/cfb-recruiting/3-star-lb-andre-polk-chooses-minnesota/>.
- Rynes, S.L., Heneman, H.G., & Schwab, D.P. (1980). Individual reaction to organizational recruiting: A review. *Personnel Psychology* 33, 529-542.

- Sander, L. (2008, December 19). For college athletes, recruiting is a fair (but flawed) game. *Chronicle of Higher Education* 55(17), A1. Retrieved from <https://www.chronicle.com/article/For-College-Athletes/8066>
- Safian, N. (2016). *Cornerstone Reputation college athletics survey report*. Somerville, MA: Cornerstone Reputation.
- Sanderson, J. (2011). To tweet or not to tweet: Exploring Division I athletic departments' social-media policies. *International Journal of Sport Communication* 4: 492-513.
- Sanderson, J. (2013). Stepping into the (social media) scene: Building athlete identity via Twitter. In L. Luppacini (Ed.), *Handbook on Research on Technoself: Identity in a Technological Society, vol 2* (pp. 419-438). Hershey, PA: IGI Global.
- SB Nation College News. (2014, December 7). The regular season's last football rankings: playoff, polls, and computers. *SB Nation*. Retrieved from <http://www.sbnation.com/college-football/2014/12/7/7347525/college-football-rankings-top-25-playoff-ap-coaches>
- Schawbel, D. (2012, July 09). How recruiters use social networks to make hiring decisions now. *TIME*. Retrieved from <http://business.time.com/2012/07/09/how-recruiters-use-social-networks-to-make-hiring-decisions-now/>
- Schmitt, N., & Coyle, B. (1976). Applicant decisions in the employment interview. *Journal of Applied Psychology* 61, 184-192.
- Scrapy. (2008). Retrieved from <https://docs.scrapy.org/en/>.
- Silver, N. (2012). *The signal and the noise: Why so many predictions fail—but some don't*. New York, NY: Penguin Press.

- Singh, P., & Finn, D. (2003). The effects of information technology on recruitment. *Journal of Labor Research* 24(3), 395-408.
- Smith, R. (2009). College football and student quality: An advertising effect or culture and tradition? *The American Journal of Economics and Sociology* 68(2), 553-579.
- S-O'Brien, L., Read, P., Woolcott, J., & Shah, C. (2011). Understanding privacy behaviors of millennials within social networking sites. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-10.
- Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics* 87, 355-374.
- Staples, A. (2012, January 20). The commitment project: Studying recruits, college football edition. *Sports Illustrated*. Retrieved from <http://www.si.com/more-sports/2012/01/20/commitment-project>
- Steinberg, L., & Cauffman, E. (1996). Maturity of judgment in adolescence: Psychosocial factors in adolescent decision making. *Law and Human Behavior* 20(3), 249-272. <http://doi.org/10.1007/BF01499023>
- Stevens, C.K., & Kristof, A.L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology* 80(5), 587-606.
- Stinson, J.L., & Howard, D.R. (2007). Athletic success and private giving to athletic and academic programs at NCAA institutions. *Journal of Sport Management* 21, 235-264.
- Tan, P.N., Steinbach, M. & Kumar, V. (2006). *Introduction to Data Mining*. Boston, MA: Pearson Education, Inc.

- Tenbrunsel, A.E., & Diekmann, K.A. (2002). Job-decision inconsistencies involving social comparison information: The role of dominating alternatives. *Journal of Applied Psychology* 87(6), 1149-1158. <http://doi.org/10.1037//0021-9010.87.6.1149>
- Toma, J.D., & Cross, M.E. (1998). Intercollegiate athletics and student college choice: Exploring the impact of championship seasons on undergraduate applications. *Research in Higher Education* 39(6), 633-661.
- Towers Watson. (2014). *The targeted employee value proposition: Drive higher performance through key talent and differentiated rewards*. Retrieved from <https://www.towerswatson.com/en/Insights/IC-Types/Survey-Research-Results/2013/12/2013-2014-talent-management-and-rewards-study-north-america>
- Treadway, D.C., Adams, G., Hanes, T.J., Perrewe, P.L., Magnusen, M.J., & Ferris, G.R. (2014). The roles of recruiter political skill and performance resource leveraging in NCAA football recruitment effectiveness. *Journal of Management* 40(6), 1607-1626. <http://doi.org/10.1177/0149206312441836>
- Tsai, W, Chen, C., & Chiu, S. (2005). Exploring boundaries of the effects of applicant impression management tactics in job interviews. *Journal of Management* 31(1): 108-125. <http://doi.org/10.1177/0149206304271384>
- Tversky, A., & Kahnemann, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science* 185(4157), 1124–1131.
- Twitter (2016). *REST APIs*. Retrieved from <https://dev.twitter.com/rest/public>
- U.S. News and World Report (2014). *Best Colleges*. Washington, DC: U.S. News and World Report.

- USA Football (2001). Academics: What do coaches look for in players? Retrieved from <http://www.usafootball.com/articles/displayArticle/5349/4500>
- USA Today High School Sports (2016, April 8). NCAA opens door for unlimited texting between recruits, football coaches. *USA Today*. Retrieved from <http://usatodayhss.com/2016/ncaa-opens-door-for-unlimited-texting-between-recruits-football-coaches>
- USA Today High School Sports (2017, May 8). Early signing period in football has arrived. *USA Today*. Retrieved from <http://usatodayhss.com/2017/december-early-signing-period-football-recruiting>
- Vecchio, R.P. (1995). The impact of referral sources on employee attitudes: Evidence from a national sample. *Journal of Management* 21, 953-965.
- Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57(2), 307-333. <http://doi.org/10.2307/1912557>
- Xie, B., Huang, M., & Watkins, I. (2012). Technology and retirement life: A systematic review of the literature on older adults and social media. In M. Wang (Ed.) *The Oxford Handbook of Retirement* (pp. 493-509). New York, NY: Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780199746521.013.0150>
- Wang, X., Zhao, K., & Street, N. (2017). Analyzing and predicting user participations in online health communities: A social support perspective. *Journal of Medical Internet Research*, 19(4), e130. <http://doi.org/10.2196/jmir.6834>
- Wayne, S.J., & Liden, R.C. (1995). Effects of impression management on performance ratings: A longitudinal study. *Academy of Management Journal* 38(1), 232-260.

- Westfall, B. (2017, August 11). How Fortune 500 companies engage talent on Twitter. *Software Advice*. Retrieved from <https://www.softwareadvice.com/resources/how-fortune-500-companies-engage-talent-on-twitter-update/>
- Wong, K.F.E., & Kwong, J.Y.Y. (2005). Between-individual comparisons in performance evaluation: A perspective from prospect theory. *Journal of Applied Psychology* 90(2), 284-294. <http://doi.org/10.1037/0021-9010.90.2.284>
- Wong, R. (2014, November 20). The biggest Twitter beefs in professional sports. *Sports Illustrated*. Retrieved from <https://www.si.com/extra-mustard/2014/11/20/biggest-twitter-beefs-professional-sports>
- Wyse, R. (1972). *Attitudes of selected black and white college business administration seniors toward recruiters and the recruiting process* (Doctoral Dissertation). Retrieved from Dissertation Abstracts International (osu1486736094430511)
- Zhang, X., Zhao, J., & Cao, G. (2015). Who will attend? Predicting event attendance in event-based social network. In *Proceedings of the 16th IEEE International Conference on Mobile Data Management* (pp. 74-83). <http://doi.org/10.1109/mdm.2015.231>
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software* 27(8), 1-25. <http://doi.org/10.18637/jss.v027.i08>
- Zottoli, M.A., & Wanous, J.P. (2000). Recruitment source research: Current status and future directions. *Human Resource Management Review* 10(4), 353-382.