Title
Evaluating Student Politics: A Social Media Approach to Assessing Ideological Skew on College Campuses

Author
Havey, Nicholas Francis

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Evaluating Student Politics: A Social Media Approach to Assessing Ideological Skew on College Campuses

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

Nicholas Havey

2022

ABSTRACT OF THE DISSERTATION

Evaluating Student Politics: A Social Media Approach to Assessing Ideological Skew on College Campuses

by

Nicholas Francis Havey

Doctor of Philosophy in Education

University of California, Los Angeles, 2022

Professor Mitchell J. Chang, Chair

From coeducation to integration, institutions of higher education in the United States of America have been the regular target of politically driven criticisms. Chief among these criticisms in recent years has been the idea that institutions of higher education are lacking in ideological diversity. These critiques have had significant impacts on institutions of higher education, as colleges and universities have faced persistent disinvestment and restructuring as a result of partisan policy making.

These criticisms are not, however, new and have been a hallmark of higher education since its inception in the United States of America. In response to these criticisms, a substantial amount of research has been conducted on student politics, student political behavior, and the politics of college campuses. This research, however, has been consistently driven by survey data and is dated, leaving room for a contemporary exploration of student politics. This study was designed with the limitations of previous research in mind, and with a focus on the modern information landscape and social media's inextricable connection to contemporary expressions of political ideology, and utilizes digital trace data to investigate student politics, ideological diversity, and ideological skew on college campuses.

Using digital trace data collected from Twitter and a latent attribute analysis of that data, I constructed a novel dataset of 8,554 students representing 43 states and 139 unique institutions of higher education. The average estimated political ideology of the students in the dataset was -0.337, which represented a left of center political position. With respect to students' information networks, the students in the dataset followed 43,958 unique information sources on Twitter, which had an average estimated political ideology of 0.4234, which represented a right of center political position.

The findings of the study indicate that, while the average college is moderate but leans left, there is no lack of ideological diversity on college campuses in the United States of America and previous survey-based research on the topic may over- and under-represent certain political populations. Institutional variables such as cost of attendance and institutional selectivity were not significantly predictive of student politics and campuses in general are exceedingly moderate, ideologically diverse, and not as politically extreme, specifically with respect to liberal skew, as they are accused of being. Similarly, the study finds that students consume a diverse swathe of information online, but that information is likely to be significantly more moderate than the political positions of most students. Similarly, students' information networks were ideologically diverse, but that diversity was less prevalent in the information networks of more conservative students. Given a lack of alignment between student ideologies and the available information online, a theory of constrained choice online was proposed and substantiated.

Finally, most students appear to prefer to associate with peers who share their political views and subsequently consume information that is aligned with those views, but extreme homophily, siloing, and selective exposure to ideologically consonant information is most prevalent among conservatives.

Implications from this study for research, practice, and policy largely centered on the utility of digital trace data as an alternative and novel data source and the reality that colleges and universities are not as politically extreme as they are perceived. In sum, student politics largely reflect the politics of the country as a whole and there is no significant liberal skew.

The dissertation of Nicholas Francis Havey is approved.

Ozan Jaquette

Safiya Umoja Noble

Zachary Steinert-Threlkeld

Mitchell J. Chang, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

A dissertation is not something you do alone, and I certainly could not have produced this work without the help, support, and guidance of a lot of people.

First, my committee. A lot of people in my life would probably be quick to call me a pusher. I like to push myself and I like to push other people and I am thankful to have had a committee that was willing to push me. From the start of my dissertation journey, Mitchell Chang encouraged me to think outside the box, to challenge myself, and to pursue personal (and methodological growth) within the doctoral program at UCLA and, even though I came in wanting to one dissertation, he pushed me and supported me to do another that has been significantly more fulfilling, challenging, and, frankly, interesting. Ozan Jaquette drove my methodological growth, always encouraging me to teach myself new things, ask questions, and turn to Google, GitHub, and StackOverflow when I had a vision of something I did not yet know how to execute. Zachary Steinert-Threlkeld helped me refine that vision and connected me with the people and resources, including a book he wrote, that I needed to know about to get the job done. And then he gave me the space and support to push myself even further and continue to learn and grow past the dissertation and for that I am eternally grateful. Finally, Safiya Noble challenged my thinking about my topic and encouraged me to conceptualize and frame my study, and my work in general, in new and exciting ways. And she wrote an incredible book, *Algorithms of Oppression*, that served as the launching pad for this new way of thought.

The professors I sought guidance from are not the only people I learned from during my time, and I could not have finished this study, and this program, without the friendship and support of my peers at UCLA and other institutions. To my cohort, Chantra Nhien and Michelle Sendowski, I love you both and sincerely appreciate how nice it has been to have people who are

supportive of anything and everything I do and I am so proud of both of you. To the Chang Gang– Jason Chan, Connie Chang, Jazz Kiang, Matt Griffith, Demeturie Toso-Lafaele Gogue, and Xiong Her– you are all stars and I am so thankful to have you as my academic family. Sharing work with you in RAC and having the opportunity to work with and alongside all of you has made me a better person and scholar and I am so grateful to know all of you. To the rest of my UCLA friends– Damani White-Lewis, Ben Kersten, Olivia Obeso, Patricia Martin, Crystal Chan, Elizabeth 'Zab' Fernandez, and anyone else I forgot but still cherish– thank you for making a new campus feel a little bit more like a home. To my pre-UCLA academic friends– Lucas Schalewski, Trace Camacho, Noel Hennessey, Cheree Meeks, Hannah Harp, Devon Degraffenreed, Jesse Stipek, and Brett Fera– your continued support, even after you stopped seeing me regularly in person, means everything.

This dissertation was financially supported by the UC National for Free Speech and Civic Engagement, led by the inimitable Michelle Deutchman and the unerringly helpful Brenda Pitcher, who I cannot thank enough. Having extra money during a graduate program is a huge boon, particularly if you've been working 20-40 hours a week outside of school so that you can pay rent and buy essentials like food, but the Center and the people who populate it have been more than just a source of extra income, they've been a family. I'm particularly thankful for Michelle Deutchman, Brenda Pitcher, Beth Niehaus, Jill Dunlap, Brian Soucek, Matt Griffith, and Xiong Her, all of whom have made my last two years so much better than they would have been without them.

To my friends outside of academia who have had to listen to me talk about research, my dissertation, and coursework for the last six years, thank you for not blocking my number and reading (or pretending to read) drafts of stuff that you did not really understand or care about

past caring about me. I love you Jessie Marman, Libby Willis, Andrew Bitzko, Ellie Bitzko, Hunter Kennedy, Dan Trotta, Stephen Bergauer, Sam Singh, Nicole Bradford, Hillary Litherland, George Ferakis, Allegra Amend, Eva Romero, Anna Cooper, Nate Palmer, Nick Mahon, Adan Garcia, Tanzida Zaman, Michelle Martin, Angelica McCarron, Francesca Maria, Spencer Lee, Kevin Snavley, Wyatt Gross, Katia Salgado, Syd Curtis, Jacqui Oesterblad, Ravi Patel, Erika Sena, Robert Jones, Khoi Nguyen, and Jaden Bankhead. You're all stars.

Finally, to my family, thank you for everything. I know most of my graduate school experience has gone over your heads and you're still recovering from trying to read my first thesis, but you're always willing to listen and that's maybe more important than understanding. I love you Mom and Dad, Ellie, Greg, Wally, and Sonny. To the rest of my walking-the-dog or driving-to-campus phone call contingent–Aunt Deb, Aunt Cindy, Aunt Michelle, and Uncle Vito– thanks for always taking my calls. And last but not least, shoutout to Bandit, who has been a steady companion for the hundreds of hours that doing this took and for never being mad that that time wasn't spent on walks.

VITA

Education

2019	M.A., Higher Education and Organizational Change
	University of California, Los Angeles
2018	M.A., Higher Education
	The University of Arizona
2016	B.S.H.S, Physiology
	The University of Arizona
2016	B.A., Spanish and Portuguese with Honors
	The University of Arizona

Experience

2022	Managing Editor, The Journal of Higher Education
	University of California, Los Angeles
2022	Grant Writer
	The Maria Group
2021	Instructor
	The University of Chicago, MSC in Threat & Response Management
2018-2021	Editorial Assistant, The Journal of Higher Education
	University of California, Los Angeles
2018-2021	Top-Rated Independent Contractor and Freelance Editor
	UpWork & Personal Clients
2016-2018	Graduate Assistant for Academic and Honors Initiatives
	The University of Arizona: The Honors College and Residence Life

Select Publications

Havey, N. (2021). "I can't be racist, I'm gay": Exploring queer white men's views on race and racism. JCSCORE: *Journal Committed to Social Change on Race and Ethnicity*, Vol. 7, No. 2. https://doi.org/10.15763/issn.2642-2387.2021.7.2.136-172

Havey, N. (2021). The New Right: Conservative student political repertoires and intragroup conflict. *Education Policy Analysis Archives*, 29(134). https://doi.org/10.14507/epaa.29.5669

Havey, N. (2021). Untapped Potential: Understanding How LGBQ Students Use Dating Applications to Explore, Develop, and Learn About Their Sexual Identities. *Journal of Woman and Gender in Higher Education*. https://doi.org/10.1080/26379112.2021.1988625

Havey, N. (2021). When the Campus is the Locker Room: A Queer Analysis of Student Athletics Bias Incidents. *Journal of Woman and Gender in Higher Education*. https://doi.org/10.1080/26379112.2021.1950740

Havey, N. (2020). Partisan Public Health: How Does Political Ideology Influence Support for COVID-19 Related Misinformation? *Journal of Computational Social Science*. https://doi.org/10.1007/s42001-020-00089-2

Havey N. (2020) "Radicalized on Campus? (Un)Coded Whiteness as Campus Social Movement", *Journal of Critical Thought and Praxis* 10(1). https://doi.org/10.31274/jctp.11566

Dorman, J., Havey, N., & Fagioli, L. (2019). Bridging the Gap: Summer Bridge Programs as an Effective Strategy for Improving Minority Student Academic Attainment in Community Colleges. *Journal of Applied Research in the Community College*, Vol. 27, No. 1.

Select Presentations

Havey, N. (2022). *Broadening What We Mean by Gender Scholarship: A Focus on Queer, Trans, Gender Non-Conforming, and Non-Binary Faculty, Staff, and Students.* Invited Panelist at the annual NASPA conference.

Havey, N. (2021). *Doing More Than Diversity: Moving Past Compliance Approaches to Serving Queer, Trans, and Nonbinary Students*. Invited Presentation at the annual meeting of the NASPA Special Issues Law Conference.

Havey, N. (2019). *Radicalized on Campus? Whiteness as (Un)Coded Campus Social Movement.* Research paper presentation at the annual meeting of the Association for the Study of Higher Education.

CHAPTER 1: INTRODUCTION

Background of Study

Historical Purpose of Higher Education

Higher education in the United States is associated with a multitude of positive outcomes including: credentials that facilitate the pursuit of upward mobility, improved critical thinking skills, civic and democratic engagement, and an informed citizenry that is the backbone of American democracy (Birnbaum, 1983; Thelin, 2011). While critics argue that credentialism, or the pursuit of credentials as a sort of badge rather than as a pursuit of the actual skills they represent, has led to a decreased focus on teaching and learning in higher education (Labaree, 1997), higher education remains a mainstay of society and culture in the United States.

Higher education has also historically facilitated much of what is necessary for the healthy functioning of democracy, namely the education of an informed citizenry that is interested in and can engage with the variety of interlocking systems that allow democratic societies to function (Thelin, 2011). Within the last several decades, the field of higher education and student affairs has more actively integrated civic engagement into the general curriculum (Miller & Gunnels, 2020) and overall campus experience, evidenced by the facilitation of voter registration and voter turnout programming (Benenson & Bergom, 2019; Bennion & Nickerson, 2021) and the use of civic engagement as a way to pursue "real world learning" (O'Connor & McEwen, 2021). Armed with specialized knowledge in a variety of areas and shaped through exposure to a diversity of ideas, this informed citizenry is better equipped to participate in civic and democratic processes than they would be without higher education. This focus on civic engagement is the core of American democracy, and the nation's students are regularly considered the nation's future culturally and politically. The increased attention to democratic education and civic engagement has not been without criticism, however, as higher education continues to be criticized for a perceived erosion of traditional, conservative values and an inherent liberalism and liberal skew that limits ideological diversity and serves to indoctrinate students into a liberal political identity they may not otherwise have. This criticism is not new. Given the persistence of this criticism, and the tangible impacts on institutions of higher education, a contemporary and nationally representative study of the ideological skew of American higher education is imperative.

Partisan Critiques of Higher Education

Higher education has long been a site for generational discourse around appropriateness and tradition, with students at the forefront of developing critical cultural changes and administrators and off-campus politicians questioning and criticizing the changes instituted by each new generation (Rhoads, 1998; Thelin, 2011). From challenging the concept of *in loco parentis*, the idea that faculty and administrators can and should act in the position of parents towards students (Thelin, 2011), to advocating for coeducation, desegregation, and divestment (Rhoads, 1998), higher education has always been political. It has thus become a topic and target of political ire across the political spectrum, though the idea that institutions of higher education are inherently liberal has been the most consistent political critique of higher education.

Contemporary conservative criticisms of higher education largely started with William F. Buckley, Jr. and are a product of his thinking. In 1951, Buckley, then a recent graduate of Yale and former editor at Yale's student newspaper, *the Yale Daily News*, published a controversial and polemic piece of writing condemning academic freedom as a "superstition" that he felt was designed to facilitate liberal indoctrination at America's colleges and universities and was designed against the interest of conservatives. This text, *God and Man at Yale: The Superstitions* *of 'Academic Freedom'* (Buckley, 1986, originally published in 1951), was perhaps the earliest and most explicit assault on academia as an institution, academic freedom as a core pillar of that institution, and clear criticism of academia as inherently and inappropriately liberal. Buckley condemned Yale for his perceived erosion of its Christian values and demanded administrators and trustees take direct action to curb what he believed was faculty misconduct. In his historical analysis of Buckley's writings, *The Academy on the Firing Line: William F. Buckley, Jr's God and Man at Yale and the modern conservative critique of higher education*, Laukaitis (2013) describes Buckley's position as an "early modern conservative critique of higher education" which "laid a foundation for later conservative critics and their charges of higher education's politically liberal bias and growing antipathy towards the ideals of Christianity and Western civilization" (p. 130). This is evident in the work of philosophical successors like Allan Bloom, whose book *The Closing of the American Mind* (2008) similarly criticized higher education by arguing that divestment from the "great books" of Western thought resulted in a cheapening of education; his central argument relies upon the antipathy Buckley brought attention to. This early conservative critique of higher education formed the basis for the ideological blueprint of what we now recognize as the contemporary conservative political position on higher education (Judis, 2001; Nemeth, 2020; Schneider, 1999).

In their ethnographic work, *Becoming Right: How campuses shape young conservatives*, which describes the development and behavior of contemporary campus conservatives and how campuses are shaping and being shaped by politically active students, sociologists Amy Binder and Kate Wood (2014) highlight Buckley's work as formative to conservative student identity development. Connecting contemporary issues like reproductive rights and affirmative action to traditionalist beliefs grounded in mid-20th century Christian religious doctrine, today's conservative students echo Buckley in their eager critiques of contemporary liberalism while relying on the same and now decades-old political theorizing of the university as liberal that he did. These critiques, predominantly focused on allegations of liberal indoctrination at the hands of the nation's faculty, have also served as the foundation for contemporary critiques of higher education and indicate a clear intellectual lineage descended from Buckley. As an example of some of these critiques, Dinesh d'Souza's *Illiberal education: The politics of race and sex on campus* (1991), condemned progressive policies like coeducation and affirmative action as eroding the allegedly meritocratic traditions of scholarship and individual achievement Buckley originally identified as under attack.

More recent and direct invectives like David Horowitz's (2009) *Indoctrination U.: The Left's War Against Academic Freedom*, which purports to "unveil the intellectual corruption of American universities by faculty activists who have turned America's classrooms into indoctrination centers for their political causes," and Ben Shapiro's (2010) *Brainwashed: How Universities Indoctrinate America's Youth*, which does more of the same, identify the American college and university as a recurring target of conservative political attention. This lineage is alive and well on college campuses, with conservative students identifying as increasingly under assault and unsafe on what they perceive as distinctly and dangerously liberal campuses (Havey, 2020a). Contemporary conservatives on and off college campuses have acted on their perception of ideological skew, weaponizing political polarization in the pursuit of supposed ideological diversity. Some states have even passed laws to assess and remedy the ideological diversity of their institutions of higher learning, if partisan policymakers deem it necessary.

This conservative disdain for and distrust of American higher education is also evident in general surveys of the American public. Surveys conducted every year since the 1970s revealed greater political polarization in this country in the last 10 years than ever before, with more respondents identifying as either staunch liberals or conservatives than previous generations (Twenge et al., 2016). Surveys specifically focused on higher education administered as recently as 2015 and 2019 reveal that Americans are starkly divided along partisan lines with respect to their opinions about colleges and universities, with Democrats more generally positive and Republicans exceedingly more negative on higher education in recent years. Surveyed Republicans have even stated that colleges and universities have an overall negative effect on the country (Trachtenberg, 2019). Further, according to a recent Pew survey, 19% of Republicans have no faith in professors to act in the public interest and 42% feel colleges are not open to a diversity of viewpoints and are unfairly slanted towards a liberal political agenda (Parker, 2019). Citing problems like skyrocketing tuition, conservative media outlets have amplified pre-existing dissent, deepening the already plummeting lack of confidence in higher education among conservatives (Trachtenberg, 2019). Conservative organizations like the American Enterprise Institute, Turning Point USA, and Professor WatchList have driven this dissent (Kissel, 2020).

This lack of confidence is reflected in policymaking and legislative chambers controlled by Republicans, with higher education law entering an era in which college and university leaders must respond to growing calls for change (Trachtenberg, 2019). Discontent, particularly with respect to the perceived political representativeness of colleges and universities (e.g., a consistent critique of institutions of higher education as inherently and unjustly liberal bastions that do not reflect the political makeup of the country), is functionally reshaping higher education. This is not, of course, the first time this has happened.

The Liberal Academy?

Students and student political behavior have historically driven social critiques of higher education resulting in public and political response. They have led, or at the very least supported, many of the civil rights movements that have been successful in this country, including movements for coeducation (Thelin, 2011), desegregation (Biondi, 2012; Wilder, 2013), racial equality and inclusion (Biondi, 2012; Rhoads, 1998), divestment and anti-war efforts (Rhoads, 1998), and more recently the nationwide responses to racist police violence (Morgan & Davis, 2019). Students have thus also been widely and consistently scrutinized for their political beliefs and student populations regularly derided by older generations as liberal pockets of the total population, destined to grow up and become less politically radical (Biondi, 2012; Rhoads, 1998). While this perception of students, and thus campuses, as disproportionately liberal is well-documented in popular culture, there is also a wealth of empirical research describing students' political beliefs, development, and behavior before, during, and after their college days.

Most students arrive at college with political identities that mirror their parents and their upbringing (Dunkel & Decker, 2012; Eagan et al., 2017). Not all students fit this mold, however, and some arrive with developed senses of political identity formed through pre-college experiences (Dunkel & Decker, 2012). Regardless of pre-college experience, college is a place for social and political experimentation, which may influence student political beliefs (Astin, 1977; Astin & lising-antonio, 2012).

Much of the literature on the civic and democratic influences of college focuses on students' future democratic participation (Terenzini, 1994), though some of it has also directly addressed the political demographics of student bodies and the potential politicizing influence of college. Early work identified a small but moderately liberalizing impact for students (Astin, 1977, 1993; Pascarella & Terenzini, 1991, 2005; Schiff, 1993), though some, albeit fewer, analyses contradict these findings and suggest that students may become more conservative or moderate during college as well (Astin, 1993; Dey, 1996, 1997; Gross & Simmons, 2013).

In sum, while both the general population and American college students have become more politically polarized over the course of the last several decades (DiMaggio et al., 1996; Eagan et al., 2017; Eagan et al., 2014; Pryor et al., 2007), most college students arrive on campus with comparatively moderate beliefs. Students are also likely to leave with the ideologies they arrived with (Eagan et al., 2017; Hanson et al., 2012). Whether this moderation is the result of institutional and social self-selection (Elchardus & Spruyt, 2009), peer interaction (Dey, 1996), or simply chance, the literature regarding college student political orientation and development is mixed and suggests that college may simply function as a microcosmic slice of the rest of the population. I explore the history of research on college student political orientations, behaviors, and development, as well as the literature discussing campus political climate and the potentiality of a liberal ideological slant more thoroughly in Chapter 2.

The research discussed above suggests a tension between what is presented in the popular media and what is empirically known about the political climate on campuses and the subsequent potential reality of an inherent liberal skew. There are clear repercussions of that tension, such as increased disinvestment from higher education (Taylor et al., 2020), inconsistent federal guidance on critical laws and policies like Title IX, and partisan attacks on concepts grounded in diversity, equity, and inclusion. This supposed indoctrination and ideological slant, however, is almost entirely evidenced in self-report survey data with samples that are as far from nationally representative as they are dated. Or driven completely by the personally collected “anecdata” of those, such as William F. Buckley, Jr., who perceive an erosion of what they believe to be the foundations of higher education. Given the criticisms of higher education as a distinctly skewed bastion of liberalism, and the material consequences institutions of higher education might face, a contemporary investigation into the ideological skew of American higher education is necessary.

Summary and Problem Statement

Higher education has been consistently targeted by political partisans regarding the purported lack of ideological diversity present at colleges and universities in the United States of America. Criticized as intolerant of nonliberal views, colleges face increasing pressure at local, state, and federal levels to demonstrate ideological diversity to maintain societal relevance and, more importantly, funding. Increasing political polarization, driven by a human tendency for self-selection, homophily, and selective exposure, has only exacerbated this tension. Evaluating ideological skew and, subsequently, ideological diversity is thus particularly important in the college context for two reasons. First, students have become increasingly polarized (Binder & Wood, 2014; Eagan et al., 2017; Havey, 2020a) and are likely engaging in the same homophily, selective exposure, and partisan information seeking that is evident in studies of the general population (Barberá, 2015; Colleoni et al, 2014; Weeks et al., 2019). Second, the increasing polarization of the news, a driver of cultural and political isolation, has made ideological divisions clearer. With most of the population getting their news online, news sources cannot be separated from discussions of ideological diversity and political polarization, which I discuss further below and in Chapter 3. A contemporary assessment of student political polarization, and thus campus ideological skew, cannot be divorced from the realities of social media and online news. This study attends to those realities.

Purpose and Research Questions

This study utilizes student-level Twitter data, specifically an estimate of students' political ideology based on their online behavior and a secondary calculation based on the political and news accounts they interact with, to assess both students' individual ideological positions, and thus the ideological diversity of their institutions and the field of higher education writ large, and the ideological diversity of the information they are exposed to online, specifically on Twitter.

Twitter is a social media platform that is designed around the concept of a tweet, a short message ranging from a few characters to a paragraph or two. Started in 2006, Twitter has 300 million active monthly users, 70 million of whom are in the United States, making Twitter a logical choice from which to collect digital trace data. Secondary analyses incorporate data from the Integrated Postsecondary Education Data System (IPEDS), an institution-level dataset maintained by the Department of Education, to assess whether institutional variables, such as selectivity and racial composition, influence ideological diversity. This study is guided by the following research questions:

1) To what extent is the political ideology of students active on Twitter skewed towards liberalism?

2) To what extent do the sources students follow on Twitter overlap ideologically?

a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?

3) How ideologically diverse are students' information sources on Twitter?

4) What institution-level features predict the ideology of students on Twitter?

5) What institution-level features predict the ideological diversity of the information

students are exposed to on Twitter?

Why Digital Trace Data?

For many people, but particularly the so-called digital natives in the Millennial Generation and Generation Z, the internet and social media has become the primary source of content ranging from hard news and journalism to entertainment and gossip (Gottfried & Shearer, 2016; Shearer, 2018). In the online media ecosystem, new media outlets can find their audience without needing to overcome the costly barriers of traditional media, such as investing in a physical space such as a newsroom, hiring permanent staff, or building the reputation and distribution platform necessary to turn a profit (Munger, 2020). Contemporary online outlets rely on freelancers, increasingly affordable web development, and use social media platforms like Facebook, Instagram, and Twitter as content distributors and spaces to build their followings.

Twitter specifically is increasingly a major source of information and news for large swathes of the population and students are no exception (Pennycook & Rand, 2019). Twitter's interface allows its users to curate their feeds by following whoever they choose, and their algorithm is designed to direct users towards content and accounts that align well with their interests (Noble, 2018; Steinert-Threlkeld, 2018). This results in a media environment which supports the cultivation of ideologically fragmented bubbles, or filter bubbles (Pariser, 2011).

The outlets represented in these bubbles are also mostly disconnected from normative safeguards like journalistic ethics and fact-checking and instead rely on driving traffic to their sites to boost advertising revenue and build their reputations. As a result, contemporary media outlets have increasingly little incentive to print the truth (Licari, 2020; Munger, 2020; Noble, 2018) and can forego traditional reputation building centered on fact-based reporting, hard-hitting journalism, and reliable and desired news in favor of likes, retweets, shares, and virality.

This may have led to constrained choice online, with the options presented by social media platforms curated, reduced, and filtered to fit market demands and react to market incentives. As Noble (2018) has shown, tech companies and social media platforms such as Facebook and Google are driven by profit, not by some greater sense of morality, and offer their customers what is more profitable, not most popular. I propose a theory of this constrained choice online in Chapter 3 and test it using findings from this study in Chapters 5 and 6.

We now face a media ecosystem overpopulated by outlets that prioritize clicks over quality reporting and that are becoming more biased to drive readership (Metzger et al., 2020; Munger, 2020). This rise of clickbait media is particularly troubling given that most information consumers are less familiar with online outlets than with print outlets (i.e. knowing the reputation of your local or national newspapers, but being less familiar with new online-only outlets that crop up in your social media feeds; Jurkowitz, 2014). This is significant, as a more expansive media ecosystem driven by profit (Munger, 2020; Noble, 2018) has negative implications for informed democratic engagement (Mihailidis & Thevenin, 2013; Munger, 2020) and political polarization (Barberá, 2015; Colleoni et al, 2014; Weeks et al., 2019). Within the online media ecosystem, the burden of evaluating the quality of information, once the purview and responsibility of outlets and publishers, has shifted to the user (Flanagin & Metzger, 2007). This shift is concerning given that information consumers consistently seek out news consistent with their beliefs and rely on heuristics, such as the assent of other users in their social media feed, as proxies for credibility and relevance (Metzger et al., 2010; Metzger et al., 2020; Pearson & Knobloch-Westerwick, 2018).

With respect to Twitter, research indicates that political conservatives are far more likely to engage in homophily, clustering more tightly along ideological lines, than their more liberal peers (Colleoni et al., 2014). Further, Twitter users are also likely to follow political elites (such as politicians, pundits, and journalists) that align with their ideological positions (Weeks et al., 2019). Online, “birds of a feather tweet together” (Barberá, 2015; Himelboim et al., 2013). And studying online behavior through digital trace data can result in more nuanced findings than a simple survey.

The primary data source for this study is, thus, digital trace data. Digital trace data can refer to anything from emails, comments on forum-based sites like Reddit, and the cookies left behind by browsing behavior to more specific subtypes such as Tweets and social media activity. As I describe earlier in this chapter and detail more thoroughly in Chapter 2, most of the research describing student political identity and campus ideological skew is drawn from non-representative survey data or is anecdotal. Survey data, while useful in context, presents challenges such as self-selection, response bias, and interpretation bias (Sax et al., 2003). Subjective interpretation of a question centered on ideological identity, for instance, might result in someone who is objectively less liberal than their peers identifying as more liberal by sheer self-perception. Digital trace data, particularly data collected from Twitter, is thus increasingly useful for this study, as it provides participant-level data that is a function of that person’s choices and online behavior. By calculating individual-level ideological positions using participants’ digital trace data in comparison to standardized data (i.e., politicians with established voting records who can neatly be assigned an ideological position), this study responds to the issues introduced by survey data and other approaches to answering the same questions on ideological skew. Specifically, we know college students are using social media to get their news (Shearer, 2018; Wineburg & McGrew, 2019) and increasingly consuming information exclusively through digital mediums. We also know that social media sites like

Twitter are spaces for identity development, community building and maintenance, and that young people love them and engage with them almost as if responding to a social mandate (Boyd, 2008a, 2008b). The prevalence of college-aged students online is thus significantly higher than that of other populations (Kwak et al., 2010; Shearer, 2018; Steinert-Threlkeld, 2018) and their presence online can offer more information about their social and political behavior than a survey or interview, specifically by highlighting linkages to politicians, outlets, and other students, all of which contribute to their digital, and for the purposes of this study, political footprint.

Scope of the Study

To answer the previously stated research questions, this study's multisite quantitative design will utilize digital trace data, institution-matched data from the Integrated Postsecondary Education Data System, basic descriptive statistics, and linear models. The study's data is drawn from student digital trace data collected through the Twitter application programming interface (API) and data collected by IPEDs on the institutions included in my analyses. Application programming interfaces (APIs) provide users and researchers a pathway to engaging with the data produced, collected, and stored by an application such as Twitter. APIs generally afford users the ability to extract data that is accessible through the actual application, but is more readily digestible and useful when extracted from the API. In this study, the Twitter API allows access to students' and outlets' digital trace data. Digital trace data, such as the data collected for use in this study from Twitter, has been used to conduct research focused on social goods (Ediger et al., 2010; Steinert-Threlkeld, 2018), to assess and measure culture and political polarization (Bail, 2014), to identify the impact of growing information divides (Gil de Zúñiga & Chen, 2019), and to conduct demographic studies (McCormick et al., 2017).

As this research is a multisite quantitative analysis, I engage in purposive sampling to maximize variance and have sampled students across institutional types, selectivity levels, institutional control, and states to garner a nationally representative sample. Specifically, I used Department of Education data to identify the schools enrolling the most students in each state and sampled from those schools; additionally, I prioritized specific schools (such as religious or private institutions) that enrolled a smaller share of students but reflected a particular institutional type or potential student population. Students are selected because they identified themselves as active students at one of the site campuses, or were connected to other active students through their social networks and were tied to the institution in some way (listed employment on campus, involvement in a campus sorority or fraternity, etc.). Data collection persisted to ensure representativeness by state, institutional type, and institutional control. I describe the resulting dataset and the institutions it represents, specifically detailing their compositions and types, in greater detail in Chapters 5 and 6.

I use Twitter to collect students' digital trace data and calculate student-level variables. The digital trace data I collected consists of students and the information sources they interact with on Twitter and will be used to calculate variables for the individual actors (estimated political ideology calculated using Barberá (2015)'s Tweetscores R package for both students and information sources, an average of the estimated ideologies of news accounts students follow, the standard deviation of those ideologies, and the difference between the student's calculated ideology and their news average). Secondary analyses explore whether any institutional features (institutional selectivity, racial demographics, etc.) predict student-level ideology and the ideology of news students are exposed to online. The data sources and analyses are thoroughly detailed in Chapter 4.

Contribution of the Study

This study seeks to contribute to the growing literature on student political ideology and provides updated, empirical evidence that can help inform policy making. Further, the data collected and created for this study will provide a contrast to the existing research on student political ideology and ideological skew that is driven by limited, often biased, self-report survey data that results in narrow samples being generalized to larger populations; these data can also function in comparison to existing voter registration data by county and state, as well as aggregate and institution-level data on students maintained by organizations like the Higher Education Research Institute at the University of California, Los Angeles.

Additionally, this study makes a contribution to the greater body of higher education research by demonstrating the further utility of social media data and social media space as a potential and compelling venue to study students and interactions. It is this study's intent to also encourage additional inquiry into the social media space for educators, as the internet and social media provide a nearly limitless and untapped space for educational research which considers relationships, interactions, and online communities. There is a comparative lack of work utilizing digital trace data or employing computational methods in higher education and I hope to highlight the potential of both approaches using this study's findings. Additionally, this study provides the basis for further work examining student political homophily, selective exposure, and how these behaviors influence social and political acts such as civic engagement. Similarly, this study will provide data that can be used in further analyses of the ideological skew and diversity of institutions of higher education which consider faculty and administrator politics. By first generating student-level data, future work will be able to leverage comparative analyses to assess whether ideological diversity is represented in different campus groups and how that representation differs, if at all, across groups.

Finally, this study offers a potential theoretical contribution in the presentation and subsequent validation of a constrained theory of choice online. As much of the discourse surrounding political polarization and ideological skew centers on news and media (i.e. "fake news," the "mainstream media"), investigating whether the information ecosystem and social media platforms such as Twitter are narrowing the field is imperative. Using student-level data, I will explore this constrained theory of choice online and interrogate whether highly polarized platform users, on either end of the political spectrum, are limited in their news and information consumption online. This exploration into choice will lay the groundwork for future studies of political and ideological skew and may provide empirical evidence that can problematize simplistic narratives regarding the skew of news media and how it might be impacting political identification, polarization, and radicalization.

Significance of the Study

This study is significant for three main reasons. The first is that it provides context and empirical evidence regarding the ideological skew of institutions of higher education in the United States of America, what is potentially driving that skew, and data that can help to inform policy making and institutional responses to state and federal policy. The second is that, outside of pure political reaction, ideological diversity is generally good for students and democracy, and exposure to diverse viewpoints has been shown to increase creativity and support problem solving (Luo, 2021). Assessing ideological skew in the interest of maximizing ideological diversity is paramount to supporting students to participate in a diverse democracy. Finally, there is a clear impact of partisan policymaking on higher education, and understanding the institutions, counties, and states that are vulnerable to partisan policymaking is crucial for mitigating the damage caused by politically driven changes and disinvestment.

Recent partisan policy and decision making lays bare the significance of this study. The recent bill passed in Florida (HB 233), which "prohibits State Board of Education and Board of Governors from shielding students, staff, and faculty from certain speech; requires the State Board of Education to conduct an annual assessment on intellectual freedom & viewpoint diversity; creates a cause of action for recording or publication of certain video or audio recordings; revises provisions related to protected expressive activity, university student governments, & codes of conduct," is one example. The bill, designed to assess the ideological diversity of institutions of higher education, and framed as being in support of freedom of speech and the protection of a diversity of ideas, comes on the heels of state and, prior to the change of administration, federal bans on important educational topics like critical race theory. While these are some of the most recent partisan attacks on higher education, they are not the first.

Some of the earliest work on public opinion of higher education, partisan identification, and public policy indicates that higher education used to be more of a platform issue for political candidates, and voters, than it is today (Doyle, 2007). Higher education has traditionally been thought of as a public good that is valuable and reciprocal (Birnbaum, 1983, Doyle, 2007; Thelin, 2011). Public investment in higher education at the local, state, and federal levels reflected that. As time passed and other policy issues became more important politically, higher education policy increasingly became a product of rational choice theory on both ends of the political spectrum. Liberal policymakers favored an opportunity approach that widened access and ensured the nation's youth had a path to social mobility. Conservative policymakers favored an efficiency and accountability approach to higher education. Analysis of voting records on higher education policy between 1965 and 2004 bears this out and indicates a clear partisan continuum of support for higher education policy measures, with Republicans favoring funding cuts and austerity measures and Democrats opting to support investment in the interest of improved access, retention, and representation in higher education (Doyle, 2010). These records indicate that, at least along the red/blue, Donkey/Elephant, Democratic/Republican American political binary, there is a strong difference of opinion on the public utility, and purpose, of higher education (Doyle, 2007, 2010). Further work has revealed how this plays out in both higher education policy and appropriations.

While higher education has become less of a platform issue in recent years, increasing political polarization and the subsequent influence of single-issue voters concerned with policy items like financial aid (who can access it?) and affirmative action (almost unilaterally con) have forced politicians to acknowledge and campaign around concerns that may not reflect the majority of their constituents (Dar, 2012). These highly polarized voters place pressure upon politicians to play to the center of the electoral field to garner their votes and, theoretically, future electoral success. A recent analysis of this phenomenon, which looked at lobbying and political advocacy work produced by political Think Tanks focused on higher education policy topics, indicates a clear partisan rift when it comes to higher education (Gándara & Ness, 2019). The authors' analysis of Think Tank documents reflects much of what Doyle (2007, 2010) found in his analysis of elected representatives' voting records: conservatives rely on a culpability and accountability framework (i.e. deciding what is and what is not a waste of *their* tuition and tax dollars) whereas liberals are more likely to discuss state (dis)investment, pointing to slashed funding as an equity and access issue.

Given the redistributive focus of Democratic higher education policy and the more austere approach to Republican higher education policy, playing to the center, or courting the perceived median voter, can result in higher education that leans right and prioritizes disinvestment. This was true across a fixed regression effects analysis of state budgets over 20 years, in which Dar and Lee (2014) identified that higher education is consistently the largest discretionary item in state budgets and is exceedingly contingent on partisan control, particularly in times of economic instability. For instance, following the Great Recession (2007-2009), Republican-controlled state legislatures dramatically cut higher education funding in an austerity move that many state systems have not recovered from (Dar, 2012; Dar & Lee, 2014). The analyses indicate that state disinvestment in higher education functions as a sort of balance wheel, with higher education funding cut to maintain other state budget priorities on the assumption that lost revenue can be recouped through tuition hikes and other budgetary measures (Dar & Lee, 2014), though the balance wheel does not always spin back in higher education's favor. As the disinvestment is far more common among Republican legislatures (Dar & Lee, 2014), and Republicans report increasing distrust in higher education (Parker, 2019), it is clear that policy, and funding for higher education, is increasingly sensitive to political polarization.

There is not merely a specter of austerity, however, as more recent empirical work demonstrates the influence of partisan policy making on higher education. In their analysis investigating how party control of state government and the racial demographics of institutions of higher education jointly explain state appropriations to those institutions, Taylor and his colleagues (2020) identified an even clearer pattern of partisan decision making than evidenced in Dar and Lee's (2014) study. Following the same balance wheel understanding of higher education funding described above, Taylor et al. (2020) leveraged racial threat theory-- the idea that resources will be diverted to avoid supporting members of the racial outgroup (here, nonwhite students)-- to explain higher education appropriations. Specifically, the authors describe the white racial homogeneity of the Republican party and white racial resentment as a driving factor of state-level disinvestment and show that Republican legislatures are more likely to cut funding when the racial compositions of colleges and universities become less white and Republicans maintain control of the legislature (Taylor et al., 2020). Conversely, the authors found increased Republican spending for white students. This sort of partisan decision making, and clear divestment from higher education, framed as an inefficiency reform or austerity measure by conservatives and rightly criticized as limiting access and compromising student success by liberals (Cantwell & Taylor, 2020), should be of huge concern to the field of higher education, as well as higher education policymakers, policy scholars, and policy advocates. In assessing ideological skew in two ways, as well as predictors of it at an institutional level, this study can aid campus, system, and state administrators in responding to partisan policy decisions like those described above and may even provide counterevidence regarding claims of a lack of ideological diversity already levied at institutions of higher education.

Organization of the Study

The accompanying chapters include a review of the relevant literature and theoretical bases that guide this study, the attendant methodology, the findings, a discussion of those findings, implications drawn from those findings, and a conclusion. Chapter 2 discusses the guiding literature for the study. I start with an overview of the research relevant to this study, focusing on the last several decades of research on student political behaviors, the politics of campus, what might be influencing students' political identities, and close with a critique of the data, and thus findings, of that body of research. In Chapter 3 I discuss the underlying theory for this work, specifically focusing on concepts of homophily, selective exposure, algorithmic amplification, and content moderation and how each individual concept contributes to contemporary internet users' behaviors and what I view as a constrained theory of choice online. This chapter further substantiates the utility of digital trace data for the study at hand and examines how individual decision making online can reflect a person's political position. I close the chapter with a diagram that explains an operationalization of this theory with respect to the data I collect and its subsequent impact on the study's analyses. Chapter 4 presents the methodological approach that was used to conduct the study, with a primary focus on collecting and analyzing digital trace data and secondary attention to predictive linear models. Chapter 5 presents the findings of the initial descriptive statistics for all students, an assessment of aggregate political ideology by institution, and data interrogating the overlap between students' online sources and the subsequent ideological diversity of these sources. Chapter 6 presents the findings of the predictive linear models and specifically focuses on how institutional features, integrated from IPEDs, may influence student political identity and the political position of the sources students follow on Twitter. Chapter 7 synthesizes the major findings and discusses how institutions may respond to questions of ideological skew and calls for greater ideological diversity, and offers potential interventions and considerations for research, practice, and policy. Chapter 8 summarizes the study for the reader and provides the reader a detailed account of how the study was conducted, as well as how the data was collected and subsequently analyzed.

CHAPTER 2: GUIDING LITERATURE

This chapter explores the relevant literature and concepts used to frame, analyze, and understand student political behavior and campus ideological skew using digital trace data. The chapter is divided into three parts. I begin with an overview of the link between higher education, civic engagement, and democratic education and explain why the political ideology and behavior of college students is often scrutinized and discussed on and off college campuses. Next, I detail this research from its origins in the earliest analyses of the American student to more contemporary work that explicitly looks at how students participate in modern politics and how institutions of higher education are influencing them. I close with a discussion of the limits of past research, specifically work conducted via survey, and explain why accurately assessing the ideological skew, or lack thereof, of contemporary higher education is necessary for civic, democratic, and educational outcomes.

PART I: Higher Education, Civic Engagement, and Democratic Education

One of the many outcomes associated with higher education in the United States is civic and democratic engagement (Birnbaum, 1983; Erhlich, 2000; Thelin, 2011). Higher education institutions have stood sentry against the threat of democratic tyranny by functioning as sanctuaries of nonrepression, dedicated to scholarly autonomy and freedom of academic association (Gutmann, 1987). By facilitating space for open discourse and educating generations of students toward the collectivist goal of resisting tyranny and an undemocratic imposition of elite or dictatorial will on the people, institutions of higher education have also historically supported much of what is necessary for the healthy functioning of democracy (Bowen, 1977). The informed citizenry that is theoretically the result of higher education can engage with and participate in the myriad interlocking systems (voting, legislation, the judicial system) that allow democratic societies to function (Boyte & Hollander, 1999; Erlich, 2000; Thelin, 2011). Civic engagement and democratic education lie at the heart of forming this citizenry. While the idea of democratic education has been central to the goals of higher education since its inception and into the contemporary era (Gutmann & Ben-Porath, 2014; Gutmann & Thompson, 1998), democratic education has shifted from a theoretical and necessary byproduct of higher education to a core export (Riddle & Apple, Eds., 2019).

Within the last several decades, institutions of higher education, driven by the fields of higher education and student affairs, have increasingly integrated civic engagement and democratic education into the general curriculum, opting to explicitly include promoting civic and democratic engagement within the collegiate experience. Directly promoting civic engagement has thus become one of the key features of contemporary higher education, as campuses regularly serve as hubs for voter registration and subsequent electoral participation (Benenson & Bergom, 2019; Bennion & Nickerson, 2021), as well as other initiatives designed to promote civic engagement and democratic education. From courses designed around organizing to legislative partnerships (Howe & Fosnacht, 2017), civic engagement has been integrated into both curricular and cocurricular experiences within institutions of higher education (Miller & Gunnels, 2020) and campuses have become incubators for the politically engaged and informed citizenry necessary for democracy. Schools have also emphasized the importance of civic engagement initiatives, such as service learning, problem-based pedagogies, and campus organizing as examples of “real world learning” (O'Connor & McEwen, 2021).

While these efforts to drive civic engagement and democratic education are necessary, the highly partisan political ecosystem in the United States has resulted in continuous accusations of liberal skew within institutions of higher education and questions and criticisms of whether the democratic education at the core of higher education is, in itself, inherently liberalizing. Accurately evaluating political ideology and student politics has thus become a consistent project within the research literature on higher education in the United States of America. The next section describes the origins of this work and tracks the history of research on student political identity and behavior on American college campuses.

PART II: Student Political Identity and Behavior on American College Campuses

The impact of college on students has been a central line of inquiry in higher education research in the United States of America (Feldman, 1972; Feldman & Newcomb, 1969; Gurin, 1971; Hyman & Wright, 1978; Levine, 1966; Pascarella & Terenzini, 1991). Driven by a variety of factors including a persistent desire to quantify the benefits of education, the college experience-- and its impact on students-- is one of the most consistent topics of study in higher education research and has been examined regularly over the course of the last several decades. Within that broader category, investigations of students' political identities, attitudes, and behaviors have been a similarly consistent topic of study. While much of this work has been focused explicitly on identifying the explicit distribution of students politically (i.e., what percentage of students are liberal and what percentage are conservative), descriptive work has largely been integrated into inferential work describing how college might influence a students' political identification. I detail the descriptive and inferential literature on student politics below.

The Origins of Educational Research Focused on Student Politics

Perhaps the earliest consistent empirical work conducted explicitly on student political identity and behavior on American college campuses was conducted by Alexander Astin of the University of California, Los Angeles (UCLA). As the founding director of the Cooperative Institutional Research Project (CIRP), which oversees an ongoing study of more than fifteen million college students, 300,000 faculty and staff, and 1,800 institutions of higher education, and the founding director of the CIRP subsidiary Higher Education Research Institute (HERI) at UCLA, which administers and analyzes nationwide surveys on behalf of CIRP, Astin's work on students and the impact of college influences most of the higher education research literature. Based on longitudinal survey research conducted by HERI over the course of more than a decade, Astin's *Four Critical Years* (1977) presents some of the earliest empirical evidence on the effect of college on students' political development and identifies college as an environment in which students have the space to explore contradictory beliefs and gain exposure to new experiences. Astin identifies that, politically, the majority of students are moderate, liberals outnumber conservatives, and college has a marginal liberalizing impact on students (1977). These analyses, and the idea that college is a place to try new identities and grow, is reiterated in his other work, including in the findings of a follow up study twenty years later (*What Matters in College? Four Critical Years Revisited,* Astin, 1993) and in the assessment-focused book *Assessment for Excellence*, co-authored with anthony lising-antonio in 2012. Specifically, this work acknowledges that students will likely be exposed to a greater diversity of peers, and experiences, than they may be previously exposed to, particularly if their high school or childhood neighborhood was not diverse, or if their pre-college experiences were homophilic, and that this exposure may result in a slight liberalizing effect on student politics.

Astin's work has been continued by a number of scholars, including students of his at the Higher Education Research Institute at UCLA. Outside of the consistent descriptive reports put out by HERI (e.g., *The American Freshman,* Eagan et al., 2014, 2017), which examine demographic, behavioral, and experiential trends among students, staff, and faculty, many studies utilizing the research institute's data have explicitly examined how college influences students, their political identities, and their political attitudes. Schiff's (1993) doctoral dissertation, advised by Astin and other UCLA faculty members, explicitly describes how college experiences influence students' political identification and attitudes. Using panel data from HERI representing nearly 20,000 students, Schiff found that peer interactions had a significant impact on students' political development and identification. Schiff's findings largely reiterate Astin's earlier work and describe again that most college students are moderate, liberals outnumber conservatives, and that college has a moderately liberalizing influence on students.

Continuing this work, Dey (1996, 1997) utilized HERI data and Weidman's (1989) model of undergraduate socialization to explore how peer, faculty, and social influences impact undergraduate political attitudes. Dey's 1996 study identified effects consistent with the general population, that is student political orientations change within a given political context, as a result of shifting personal identities, and as a result of normative peer and social contexts (i.e., spending time with a particular political group might suggest belonging to that group or a shift towards belonging to that group). In a departure from Astin and Schiff's work, but using the same data, he also found a positive influence of equal magnitude on students of all political stripes, specifically noting that liberal students entering liberal institutions are more likely to stay liberal or become slightly more liberal while conservative students entering conservative institutions are similarly likely to maintain their political identity or become increasingly conservative. This finding is consistent with the general process of socialization described by Weidman (1989), with students moving in the direction of whatever institutional norms they are exposed to, though Dey unequivocally rejects the notion that campuses are inherently politicizing in one way or another in stating that "popular concerns about faculty creating an environment that produces political clones is incorrect" (Dey, 1996, p. 551). He further asserts:

“It is true that students do seem to move toward political orientations consistent with those held by their faculty, but these also happen to be similar to the political orientations held by their peers and are also consistent with general social trends. Attributing such changes exclusively to the influences of liberal faculty is at best imprecise and at worst misleading” (Dey, 1996, p. 552). Dey’s 1997 follow-up work, *Undergraduate Political Attitudes: Peer Influence on Changing Social Contexts*, largely reiterated his 1996 findings. Again using panel data from HERI and Weidman’s (1989) framework for understanding student socialization, Dey (1997) similarly found that peer interactions and institutional norms influenced students’ political attitudes and identified changing social contexts as important for contextualizing and grounding any analyses of shifting political attitudes. Specifically, Dey noted that, while his findings were consistent with past work, including his own (Astin, 1993; Dey, 1988, 1996), changing political attitudes may simply be a result of a “long-standing social trend toward political liberalization” (Dey, 1997, p. 410). This social trend toward liberalization (Alwin & Krosnick, 1991; Davis, 1992; Weil, 1985), and the subsequent findings of his study, Dey (1997) argues, may simply be a result of the historical boundedness of his data. He concludes by recommending that future work take Alwin and colleague’s (1991) advice to consider the effect of period on social or political results, as findings and effects may simply be a result of context (Dey, 1997). Given the historical nature of much of the work on this topic, this dissertation study intends to follow that advice.

Reviews of The Research

More dated research focused on the impact of college on students has also been regularly reviewed and synthesized by researchers such as Ernest Pascarella and Patrick Terenzini, who have explained how college affects students across the decades (Pascarella & Terenzini, 1991, 2005). In their landmark text *How College Affects Students*, Pascarella and Terenzini (1991,

2005) explored how college students and college outcomes have come to be defined as an area of study, detailed the major theoretical models used to study college students, and review and synthesize the empirical findings of decades of research on college students. With specific respect to college student political behaviors, Pascarella and Terenzini (1991, 2005) report that student political development and subsequent identity and behavior is often a product of students' home lives prior to college, their peer interactions within college, and diversity experiences, which is consistent with what has been reported in the research literature (Astin, 1977, 1993; Dey, 1996, 1997; Schiff, 1993). While some of the research reviewed indicated a multidirectional effect of college on students' political attitudes (Dey, 1996, 1997; Schiff, 1993), it largely emphasized a modest but persistent liberalizing influence (Astin, 1977, 1993; Pascarella & Terenzini, 1991, 2005). Given the impact of period on this research identified by Dey (1996, 1997), it is not surprising that research on students' political attitudes and behaviors persisted into the early 2000s and is, as evidenced by this study, important today.

Contemporary Approaches to Understanding Student Politics

As I discussed in the introduction, most students arrive at college with a political identity informed by their parents' more established senses of political self (Binder & Wood, 2014; Davies, 1965; Gross, 2013; Jost et al., 2008) or the students' own experiences prior to college (Dunkel & Decker, 2012; Eagan et al., 2017). Experiences like volunteering with a political group, protesting, or identifying as LGBTQ may have been formative for some students prior to college. This sense of political self may influence their subsequent peer interactions at college, their behaviors on their campus, and the experiences they seek out. As Astin (1977, 1993) described, college is a place for exploration and experiential growth. Students with less stable political ideologies, subsequently, may experience cognitive dissonance as a result of the exploration possible once they arrive at campus and when they are confronted by peers with beliefs that contradict their own (Dunkel & Decker, 2012). The sort of peer interaction that can influence students' development, particularly with respect to their political behaviors and attitudes, has been at the core of historical approaches to understanding students (Astin, 1977, 1993; Dey, 1988, 1996, 1997; Pascarella & Terenzini, 1991, 2005) and is central to contemporary inquiry into student politics.

More recent work on student political behaviors and attitudes has examined topics like whether institutional type influences students' political attitudes (Hanson et al., 2012), how discipline influences students' political attitudes (Elchardus & Spruyt, 2009), how diversity experiences influence political views (Pascarella et al., 2012), and whether college students are actually liberal (Bailey & Williams, 2016). Extending previous work (Astin, 1977, 1993; Dey, 1988, 1996, 1997), there is also new research exploring students' political behaviors and attitudes that leverages HERI panel data across the last decade (Havey, 2021; Havey & Schalewski, 2021).

To explore the relationship between attending a liberal arts college and students' political views, Hanson and colleagues (2012) leveraged the Wabash National Study of Liberal Arts Education (WNS). According to the authors, "The WNS is a large, longitudinal investigation of the personal and cognitive effects of liberal arts colleges and liberal arts experiences" (Hanson et al., 2012, p. 358) and specifically oversamples liberal arts colleges. In their sample of 2,159 full-time undergraduate students, the authors found that liberal arts students generally enter college with slightly more liberal views than their peers at other 4-year institutions and become more liberal over the course of four years than peers at other institutions (Hanson et al., 2012). Consistent with other research, the authors report that the vast majority of students exit college with the same political identities they enter with (Eagan et al., 2017; Havey, 2023; Havey & Schalewski, 2022). The authors conclude that, while they were able to identify specific mechanisms underlying the institutional influence they identified within their study, student self-selection into institutions that mirror their political orientations is one possibility for understanding how institutions influence student politics (Hanson et al., 2012).

Elchardus & Spruyt (2009) similarly assessed the influence of a particular factor on student political attitudes: students' choice of academic discipline and subsequent socialization in their field. Using both cross-sectional and panel data, the authors identified self-selection into certain fields as more influential than peer interaction and socialization in influencing students' political attitudes (Elchardus & Spruyt, 2009). They do, however, note that their findings are consistent with previous research that has highlighted the influence of peer interactions on students' political attitudes (Astin 1977, 1993; Dey, 1988, 1996, 1997; Pascarella & Terenzini, 1991, 2005), though they acknowledge these findings are modest.

Pascarella and colleagues (2012) explored how diversity experiences, a topic central to contemporary political discourse and criticisms of higher education, influence first-year students' political attitudes. Using survey data of just under 3,000 students, the authors (2012) identified that greater exposure to diversity predicted more liberal political views among college students but that these interactions did not necessarily influence the development of specific social or political views. This is consistent with extant qualitative research (Binder & Wood, 2014; Havey, 2020a). Similarly, the authors found that men were more likely to be influenced by diversity experiences on campus than women, which is consistent with Sax's (2008) findings, and that more general peer interactions influenced student political development (Dey, 1988, 1996, 1997). Their findings largely reiterate that exposure to proximate or like political ideologies supports further identification with those ideologies, confirming previous work suggesting students develop through peer socialization (Dey, 1996, 1997).

Finally, Bailey and Williams (2016) directly took on the question of whether college students are actually liberal in their exploration of student political ideologies. Using a nationally representative but comparatively small survey sample of students (n < 200, whereas other studies range between 2,000 and 60,000), the authors interrogated whether student self-identification as a particular political identity accurately reflected their support for specific political policies. In their sample, they identified strong inconsistencies between students' actual and expressed political ideologies (i.e., expressed support for a decidedly progressive policy) and perceived political ideologies (i.e., self-identifying as a conservative). Their findings suggest a serious discrepancy and tension between students' perceptions of their political attitudes and what they actually believe (Bailey & Williams, 2016) and warrant a greater investigation into students' expressed political behaviors and attitudes. This future direction for research is supported by the work of Woessner and Kelly-Woessner (2020) who, in their effort to prove that college students do indeed drift left, identified that party identification is more resistant to change than policy positions and that operational versus symbolic ideology is important to parse. Their findings suggest that self-report data (i.e. identifying as liberal) and expressed political practice (i.e., support for a particular policy) are often at odds within student populations.

More recent empirical work supports previous research in identifying both the distribution of student political ideologies in contemporary student populations as moderate and the major political shifts for students across their first year of college (Havey & Schalewski, 2022) and across four college years (Havey, 2023) as marginal, with most students retaining their political orientations across their first year and four college years or, if they shift at all, shifting by one degree (from very liberal to liberal, for instance). Using panel data collected by the Higher Education Research Institute (HERI) at the University of California, Los Angeles, Havey and Schalewski (2022) examined changes in student political orientation over the course of students' first college year and what background characteristics, plans, and experiences predicted these changes. Across their sample of 65,123 first year students, the authors found that students identifying as middle-of-the-road were the majority across both time points (41.84% to 40.06 over the course of the first college year) and that most students, regardless of their beginning political orientation, retained that orientation across their first college year (70%; Havey & Schalewski, 2021). Of the students who did change their political orientation, women moving left (from middle-of-the-road to liberal) were the highest change population (n = 2,475), though the next largest change population was women moving right (from liberal to middle-of-the-road, n = 1,715). When disaggregated by race, white students were the two largest change populations, with 1,343 white students shifting right (from middle-of-the-road to conservative) and 1,336 white students shifting left (from conservative to middle-of-the-road). The authors also ran parallel logistic regression models to ascertain whether there were any significant demographic or experiential predictors of one particular political shift (becoming more conservative or becoming more liberal, for example), which largely reiterate previous findings that indicate most students do not, in fact, change politically over the course of their first college year (Dey, 1996, 1997; Eagen et al., 2017; Havey & Schalewski, 2022; Hanson et al., 2012). While the authors indicate that the majority of students do not change much politically over the course of their first year, their significant regression findings do demonstrate that certain social factors associated with conservatism, such as wealth and religiosity, predict a shift to the right whereas social factors associated with liberalism, such as a desire to change the political structure, to promote racial equity, and to interact with a diversity of people predict a shift towards the left (Havey & Schalewski, 2022). This is supported by previous research (Pascarella et al., 2012) and recent work that identified that conservative white students in particular as less likely to consider and value the perspectives of others when it comes to social and political issues (Johnson et al., 2017). Consistent with previous research findings (Astin, 1977, 1993; Dey, 1996, 1997), the newer data and findings presented in this study suggests that claims of indoctrination on college campuses are likely overblown based on the small percentage of students who shift their political orientations (Havey, 2023; Havey & Schalewski, 2022) and reiterate the slight shifts towards liberalism evident as a result of social changes and education in the general population (Mariani & Hewitt, 2008). Findings are similar for students across four college years.

Using similar panel data collected by HERI, Havey (2023) examined changes to student political ideology across four college years and identified significant predictors of these changes. His analysis of 56,169 students across 8 student cohorts spanning 10 years supports previous findings, specifically: the majority of students (60%) do not change their political orientations while slightly more than double (28%) shifted left while 13% shifted right during the course of their four college years. Consistent with Havey and Schalewski (2022), the largest change populations across all groups were students who did not change their political orientation at all, followed by students who moved one degree (from far left to liberal, for instance), suggesting an overall moderating effect of college consistent with previous work (Astin, 1977, 1993; Dey 1996, 1997; Pascarella & Terenzini, 1991, 2005). While Americans and American college students have become more politically polarized over the last several decades (DiMaggio et al., 1996; Eagan et al., 2017; Eagan et al., 2014; Pryor et al., 2007), this polarization seems to only be moderately, if at all, influenced by college and moderates remain the largest population on and off college campuses (Eagan et al., 2017; Gross & Simmons, Eds., 2014; Havey, 2023; Havey & Schalewski, 2022). The author also ran a linear regression to assess what college experiences, student demographics, or anticipated experiences might predict political development and change and found that demographic variables are not very predictive of change in political orientation and that peer interactions are key drivers of political change (Havey, 2023) consistent with prior research (Dey, 1996, 1997). Similarly, students who had positive diversity experiences on their campuses were more likely to shift to the left over the course of their college experience (Havey, 2023), which is again consistent with previous research (Pascarella et al., 2012).

PART III: An Assessment of Limitations on Student Politics and Future Research

The quantitative research presented in this chapter suggests, if there is any effect of college on students' political orientations, that that effect is small, modest, and multidirectional (Astin 1977, 1993; Bailey & Williams, 2016; Dey, 1988, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Hanson et al., 2012; Pascarella & Terenzini, 1991, 2005; Sax, 2008; Schiff, 1993). While some of these studies present data with sizable and significant sample populations (Astin, 1977, 1993; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Pascarella & Terenzini, 1991, 2005; Sax, 2008; Schiff, 1993) while others offer much smaller sample analyses or analyses conducted at a single site (Bailey & Williams, 2016; Elchardus & Spruyt, 2009; Hanson et al., 2012; Johnson et al., 2017), there are significant limitations to both the data and the findings. Specifically, these studies are limited by their focus on a particular institution or institutional type (such as a liberal arts college), are not nationally representative, and rely upon self-report survey data, which several of the studies critique themselves.

Secondary data inherently comes with limitations based on the instrumentation and the sampling. For instance, surveys like the Freshman Survey and the College Senior Survey, both administered by the Higher Education Research Institute, often feature overrepresentations of certain populations, which may skew the results towards these populations. For instance, work directly examining HERI surveys indicates that women are far more likely to respond than men (Sax et al., 2003). Similarly, analyses conducted using survey data, such as the ones described above, are limited to finite instrumentation, such as Likert-scale responses, that restrict how respondents might interpret a question. Given that the majority of work on student political attitudes and behaviors in higher education has been conducted using HERI data (Astin, 1977, 1993; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Pascarella & Terenzini, 1991, 2005; Sax, 2008; Schiff, 1993), this is a serious concern.

There are also major concerns regarding self-report survey questions within the literature. Specifically, educational researchers have questioned, and assessed, whether college students can accurately evaluate, and report, their own identities, learning, and development (Bowman & Seifert, 2011; Herzog & Bowman, Eds., 2011; Pascarella, 2001; Porter, 2011). Concern around this topic often centers on four interlocking factors: social desirability (wanting to be seen as something you are not), formatting of items (and subsequent respondent confusion around them), halo effects (response inflation consistent with social acceptability or peer expectations; Pike, 1999), and clarity of measures (Dugan, 2015). Put simply, college student surveys are inherently limited by the nature of asking students to report on themselves, their identities, behaviors, and experiences with any degree of objectivity and truth. As the majority of the data described in this chapter relies on self-report student survey data (Astin, 1977, 1993; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Pascarella & Terenzini, 1991, 2005; Sax, 2008; Schiff, 1993), the serious concerns with the validity of that data cannot be overlooked.

With respect to political orientation, assuming respondents, particularly students, have an accurate interpretation of what constitutes a certain political position and how their own ideologies align with that interpretation may result in at best imprecise and at worst inaccurate self-report data. As Havey (2020a) found, students often report or identify with a political orientation that is inconsistent with the beliefs and policy positions they subsequently describe. Similarly, Bailey and Williams (2016) identified that students do not have a good understanding of policy with respect to political positions and self-identification. Their findings that students often perceive an identity that they subsequently fail to express were further substantiated by Woessner and Kelly-Woessner (2020). Though Bailey and Williams' (2016) nationally representative survey was small (n < 200) and Woessner and Kelly-Woessner's (2020) dataset was similarly small but nonrepresentative nationally, they both found a significant lack of ideological inconsistency (i.e., identifying as a liberal but espousing support for markedly conservative policy, consistent with Havey's (2020a) findings) among students. Relying on self-reported, static political orientations is an inherent limitation of the studies presented throughout this chapter.

Finally, as noted by Dey (1996, 1997), the changing social contexts that influence college campuses and the general population cannot be overlooked. Fluctuations in political climate, inciting events like wars or national protests, and changing demographics warrant an attention to the effect of period on findings associated with any study but particularly studies concerned with political attitudes. Given the constantly shifting sociopolitical context in the United States of America, the unprecedented occurrence of a now-politicized pandemic, and recent and ongoing national protests and discourse surrounding police violence and racism in this country, it is important now more than ever to interrogate students' political behaviors and attitudes.

Summary

This chapter described the connections between higher education, civic engagement, and democratic education and detailed the literature focused on evaluating students' political behaviors and attitudes within higher education. The literature presented in this chapter was driven exclusively by self-report survey data, the limitations of which were discussed at the end of the chapter. Given both the effect of period and social change noted by Dey (1996, 1997) and the importance of accurately evaluating students' political ideologies based on their expressed rather than perceived stances (Bailey & Williams, 2016; Woessner & Kelly-Woessner, 2020), this study intends to evaluate the political ideologies of contemporary students using a novel social media method that more realistically depicts their expressed rather than self-perceived politics. The research covered in this chapter animates two of the driving research questions for this study. Specifically:

1) To what extent is the political ideology of students active on Twitter skewed towards liberalism?
2) What institution-level features predict the ideology of students on Twitter?

I discuss the remaining research questions, animated by my guiding theory, in the following chapter.

CHAPTER 3: GUIDING THEORY

This chapter explores the guiding theory used to frame and understand student political behavior and campus ideological skew using digital trace data and how the information sources students follow on Twitter can serve as secondary assessments of their ideological positions. The chapter is divided into four parts. I begin by describing Twitter as a platform and product of the contemporary information ecosystem and highlight both the functional utility of the platform alongside its cultural relevance and staying power. In this first section, I describe how contemporary market incentives have led to management of media in the form of content moderation (Roberts, 2016; 2017; 2018) and algorithmic amplification (Donovan & boyd, 2021; Noble, 2018; Roberts, 2021) and reiterate Noble's (2018) findings that social media platforms like Twitter present consumers with what is most profitable, not necessarily the most popular. Then, I follow with an overview of approaches to processing information and describe several possible pathways an information consumer might take when deciding what information to curate into their feed and subsequently consume. This is critical, as this curation can also be understood as a performance of, in this study, political identity. Next, I detail social, cognitive, and behavioral influences of these processes, describing how concepts of homophily and selective exposure can limit an individual's choices and influence their decision making. These social behaviors are readily evident in my analyses, as students' calculated political positions are products of their social interactions (i.e., students who are networked entirely with liberal politicians or pundits are likely liberal). Finally, I present a theory of constrained choice online, describing how the contemporary information ecosystem, approaches to processing information, and social and cognitive influences like homophily and selective exposure, result in a limited information field online. In this section I present a model demonstrating this theory and describe how the analyses in this study will serve to explore and, potentially, validate the theory. This exploration motivates research questions 2, 3, and 5. At the close of this chapter I reiterate the linkages between the sections and describe how these linkages inform the methodological approach in this study, explained in the following chapter.

PART I: Social Media and the Information Ecosystem

The advent of the internet and the massive explosion of available information has necessitated increased attention to digital, information, and media literacies. This explosion of information has driven students, and most information consumers, online, and the majority of students today get their news from the internet, social media, and news aggregators (Gottfried & Shearer, 2016). Twitter is one platform in which students interact, organize, and get information (Gottfried & Shearer, 2016; Graham & Smith, 2016). As I described in the introduction of this study, students are exceedingly present on Twitter.

Twitter: Social and Information Network

Some of the earliest research on Twitter suggests that Twitter is both a social network and a source of information for most users, as the majority of users follow less than 10,000 other users and because the site boasts lower reciprocity than other, earlier social networking sites (Facebook, Yahoo; Kwak et al., 2010). This more diffuse, but curated, organization of the platform is reflected in the average path length for users (4.12, which is relatively short considering Twitter’s size) and suggests that the platform is mostly used for information seeking after users’ accounts have matured and stabilized following their initial network curation (Kwak et al., 2010). The subsequent overall density of Twitter allows for the broad, and comparatively swift, dissemination power of retweets (one user sharing another user’s content), with retweets serving as communication channels that require less than a day to achieve maximum spread

(Kwak et al., 2010). Subsequent work further substantiated Twitter's utility as a space for news dissemination and consumption and highlighted how digital behavior often reflects social behavior (Ediger et al., 2010). Online, users behave much like they do offline, interacting with peers and like-minded strangers for social and informational purposes (Ediger et al., 2010). In fact, social media sites like Twitter have been characterized as spaces for identity development and the creation and maintenance of community and young people have expressed a clear interest in them (boyd, 2008a, 2008b).

Later work utilizing Twitter data found that, several years into Twitter's existence, most users utilized the platform primarily for information consumption, limiting reciprocity (ties / follows) to their social networks while maintaining low reciprocity for the accounts they used for information seeking (Myers et al., 2014). Those accounts, such as news networks, showed similar behavior, interacting and connecting with peers (other outlets) while limiting reciprocal ties with information consumers (everyday users; Myers et al., 2014). This research largely confirmed Kwak and colleagues' (2010) initial hypotheses that Twitter users start curating their social feeds based on information seeking but, as time passes, network growth is largely social (Myers et al., 2014). This high degree of assortativity suggests that small, tightly-linked social networks function very effectively for information distribution, as subsets of social groups likely to be interested in information shared by their peers are most likely already closely linked (one or two user lengths, or hops, away; Myers, 2014).

The dual utility of Twitter as a space for information seeking and social interaction has been well-documented within the literature (boyd & Ellison, 2007; Ellison & boyd, 2013; Kwak et al., 2010). In addition to work identifying the utility of Twitter for demographic and social science research like this study (McCormick et al., 2017), Twitter has been identified as a counterpublic space within the more general “networked public sphere” of the internet for marginalized groups such as Black Twitter (Graham & Smith, 2016, p. 434), an effective space for information dissemination in and outside of formal education (Anthony & Jewell, 2017; Arceneaux & Dinu, 2018; Watson, 2020), and a place for students to explore topics of interest and engage in related discourse (Dennen et al., 2020; Gleason, 2018). Given the social and informational aspects central to Twitter, as well as the nature of online socialization and presentation in which minute interactions such as a like or retweet can serve as a public marker of community (Marwick & boyd, 2011), Twitter is a logical place to examine students’ political identities and behavior.

Content Moderation, Algorithmic Amplification, and the Changing Media Market

While Twitter is a powerful social media platform that boasts millions of unique users, is widely used by individuals and institutions, and is moderated by automated features and staff, it is also a business. And businesses operate with a profit incentive. Twitter is no exception (Munger, 2020). Twitter, for instance, operates as a distribution platform and reputation builder for news outlets, who leverage social media metrics like likes, retweets, and comments to bolster advertising revenue and profit (Berry & Sobieraj, 2013). Platforms like Twitter have carefully positioned themselves as just that: platforms, making strategic claims regarding their functions and purview while simultaneously acting as de facto curators of public discourse (Gillespie, 2010). Platforms cannot, thus, be assumed to be apolitical.

Businesses, such as Twitter, thus have a profit incentive to engage in moderation of content on their platforms and amplification of content that is profitable (Noble, 2018). Given the reality that media entities on Twitter are also businesses (e.g., *The New York Times*) and curate their content in the interest of profit (Munger, 2020), understanding how content moderation and algorithmic amplification have altered the contemporary media ecosystem is necessary.

Content Moderation

Commercial content moderation is, at its core, a business practice that is usually very opaque and underexplained by the companies engaging in it (Roberts, 2016; 2017). Content moderation is also regularly outsourced, exacerbating the opacity of practices that dictate what content gets promoted and what content gets removed from a site. While some view the internet as a free place for expression or a contemporary public square or sphere (Gil de Zúñiga & Chen, 2019), the internet, but more specifically the spaces curated and maintained by private companies, is rife with moderation (Roberts, 2016), though this moderation may go unnoticed by the average consumer. The terms and policies describing moderation, if available at all, may also be buried in legalese and in complicated and lengthy user agreements, which is itself a tactic designed to preserve the propriety and opacity of companies' practices (Roberts, 2016).

Content moderation did not, of course, start as a commercial practice. The internet has historically been resistant to censorship, though moderation naturally arose as a need in early internet communities (Roberts, 2017). In spaces like the fictional world of Azeroth, the land at the center of the video game World of Warcraft, game moderators were coded into the DNA of the game, with human staff assigned to monitor in-game behavior and respond to player complaints. On the world's largest forum-based site, Reddit, each subreddit (for example r/Dissertation, a subforum for people, like me, seeking support while working through their dissertations) has a moderator or team of moderators that ensures that the space remains usable and, in the best cases, friendly. In most cases, content moderation has, thus, been a choice made in the interests of improving user experience and the viability of the space in which content is being moderated (Roberts, 2017). With the rise of the internet age and the wild outgrowth of content in general, content moderation moved rapidly into the commercial sphere and functions most regularly as a form of risk management which requires human discretion (Roberts, 2017).

As a commercial practice, content moderation functions to maintain both popularity and, thus, profit. The mark of a good commercial content moderator is, then, that it is not noticeable or perceptible to the average consumer that they did something at all (Roberts, 2016; 2018). While banning a relatively unknown account or removing a post that has only gained traction with one or two users would be comparatively invisible as an action of commercial content moderation (Roberts, 2016), commercial content moderators have to balance their companies' and their own principles against a profit motive. This often means that commercial content moderators' work places them squarely in some of the most toxic places in the world, sorting through inflammatory and hate-filled speech and deciding what survives. Many have offered the advice to never read the comments (Smith, 2020), but commercial content moderators have to. The inevitable removal of former president Donald Trump from Twitter is an example of this balancing act— while he consistently broke the terms of service he agreed to when he created a Twitter account, his popularity and platform, which Twitter itself helped to cultivate and grow, was a financial asset to the firm.

This careful balance, however, is neither objective nor viewed consistently by those impacted by content moderation decisions. Content moderation has, as a result of the discretionary and imbalanced perception many hold of content moderation (i.e., only people I like are being banned and none of my enemies ever seem to be held accountable), become a hotly contested policy, advocacy, and public concern (Gillespie et al., 2020). Woven into the complex tapestry of discussion surrounding content moderation is another form of content moderation which, in lieu of removal and modification, serves to amplify content deemed desirable.

Algorithmic Amplification

Algorithmic amplification describes how human intelligence and choice, in the form of a carefully designed algorithm, results in the cultivation and increased spread of specific and selected content. A product of the rise of the “Googlearchy” (Hindman, 2008, p. 15), the hypervisibility and utility of search engines, and the growing cessation of information seeking authority to and subsequent trust in sites like Google and Twitter as good-faith arbiters of information (Hargittai et al., 2010), algorithmic amplification is the other side of the content moderation coin.

While the word algorithm has mutated in contemporary parlance to mean something akin to artificial intelligence driven magic, the most basic meaning of the term describes a set of rules to be followed in sequence or calculation in the aim of solving a problem. In common practice, an algorithm describes a set of decisions and choices that a human has made that that same human would like a computer to execute. This could range from something as common as identifying a particular string of letters as a word, locating that string or word in all incoming emails, and rerouting emails containing that string to a junk folder in an email inbox to something as complicated and potentially biased as searching for a collection of color identifiers assigned to a pixel, assessing how those pixels are arranged in sequence, evaluating that sequence as an image, and categorizing that image as good or bad. In sum, while algorithms may seem and feel like they are making decisions based on objective, carefully selected logics, they are a function of humans and any automation, algorithmic or not, is a human intervention (Roberts, 2021). The assumption that algorithms, and the automation and artificial intelligence that can stem from them, is somehow better or more trustworthy than human discretion is baked into the foundation of modern technology and how companies operate in an increasingly tech-heavy age (Noble, 2018; Roberts, 2021).

The routine denial of the human component behind artificial intelligence is both a risk and a calculated choice in commercial spheres. While the classic adage garbage in garbage out applies in droves to natural language processing as a computational moderation tool (i.e., having a computer search for strings and words in sequence to decide if they are actionably bad versus having a human do this at the most basic level), a "logic of opacity" and intentional obfuscation of moderation, and amplification, can be exceedingly useful for tech companies like Twitter (Roberts, 2021, p. 54). This "computational overconfidence" (Roberts, 2021, p. 56) is a business decision. And it is not a decision without consequence.

With respect to media companies, such as Twitter, the choices that are made with respect to coverage and amplification drastically impact both the spread of misinformation and the growth of polarizing rhetoric online (Donovan & boyd, 2021). Safiya Noble's (2018) investigation of search terms, content platforms like Google and Facebook, and how a human-designed and enforced algorithm results in biased and racist results in the name of profit is perhaps the best example of this.

Her 2018 book, *Algorithms of Oppression*, documents in great detail how programming, done by and for humans, is inherently political and how massive tech conglomerates– functional monopolies–artificially and selectively dictate what is available online. Noble's (2018) work focuses on investigating how certain narratives are pushed by companies and framed as merely the result of 'the algorithm.' Her earliest examples in the book describe what happens when she searched for Black girls on Google: the results featured a lot of porn. She describes in detail how

Google's page rank system was designed to appear as if the top search results were simply the most popular but explains that they are, in fact, just what was the most profitable (Noble, 2018). Business interests are primary, she argues, and tech companies have worked hard to normalize the notion that the 'best information' is what they present, even if it is racist and sexist.

The organizational climates that business incentives drive, Noble (2018) argues, have resulted in a lowest common denominator 'majority reader' (p. 95) that reflects contemporary social and racial climates and marginalizes those who are already on the margins of the reality that is driving profit. In this reality, some communities, such as those seeking porn or racist information, are supported while some are further marginalized (Noble, 2018). In practice, this has resulted in the creation and framing of social media as a race-less space, with algorithms carefully designed to promote what is profitable even when what is profitable is white nationalist sentiment (Daniels, 2018). As both Noble (2018) and Daniels (2018) argue, bias, and racism specifically, is a feature of platforms, not a bug. This intentionally created divide is the result of algorithmic amplification and the collective efforts made to frame human intervention as neutral and objective further reinforces existing biases at the human level. Algorithmic amplification can, thus, result in both increased attention to biased, problematic, and toxic information, such as the continued spread of profitable misinformation around the COVID-19 pandemic (Havey, 2020b), white nationalist propaganda (Daniels, 2018), and anti-Black content (Noble, 2018).

Why Does This Matter?

Increased trust in the internet in general and in platforms like Google and Twitter to promote the best and most accurate information has resulted in an epistemology of ignorance online (Bhatt & MacKenzie, 2019). Aggressive content moderation and algorithmic amplification of content driven by profit has also narrowed the field of information available for consumption, displacing responsibility and ensuring a consumer class who simultaneously feel informed while being spoon-fed information that, regardless of veracity, is going to make a company money (Bhatt & MacKenzie, 2019; Munger, 2020; Roberts, 2017). Algorithms, and trust in them, have created asymmetries of power along information lines (Bhatt & MacKenzie, 2019; Noble, 2018) that necessitate increased attention to how users are curating the information they consume (Bhatt & MacKenzie, 2018; Hargittai et al., 2010; Metzger et al., 2020; Noble, 2018; Wineburg & McGrew, 2019).

Efforts to improve the curation of online information networks have largely relied upon digital literacy approaches that center checklists and other small heuristics like "considering the source" and checking the domain name of a website to assess its credibility (.edu versus .com, for instance) prior to assuming its veracity and utility (Meola, 2004; Ostenson, 2014). Unfortunately, these checklists are simply not being used, and contemporary information consumers cede decision making with respect to information evaluation to search engines like Google and social networks like Twitter (Hargittai et al., 2010; Metzger et al., 2020; Noble, 2018; Wineburg & McGrew, 2019). Between network curation and overreliance on heuristics, which can create a false sense of security given the ease of digital presentation and the market incentive for information producers to appear credible (Munger, 2020; Wineburg & McGrew, 2019), additional attention to the information that is curated and consumed, and how that curation results in a narrowed information ecosystem, is needed.

One strategy that has proved effective in improving information curation and generally improving the quality of information consumed online is teaching information consumers to behave like fact-checkers, who tend to read laterally, leaving the site they are on to corroborate the information it presents with other sources (Addy, 2020; McGrew et al., 2019; Wineburg &

McGrew, 2019). Traditional approaches to information curation and evaluation, like the checklist, emphasize reading vertically, scouring the page for a byline and other heuristics or indicators that might suggest that the information therein is credible (Meola, 2004; Ostenson, 2014). This has not proven effective, particularly because it can be difficult to accurately identify some of the items included on checklists and people tend not to work very hard to do so (Brem et al., 2001; Julien & Barker, 2009; O'Brien & Symons, 2007; Ostenson, 2014). In experimental settings, a small intervention designed to teach students to read laterally modestly improved evaluation of content, though students tended to rely on heuristics such as site appearance and attractiveness prior to the intervention (McGrew et al., 2019). This intervention was based on the researcher's other work, which relied upon the expertise of a group of 45 internet-savvy people across three groups: 10 historians with doctorates, 10 professional fact checkers, and 25 Stanford University undergraduate students (Wineburg & McGrew, 2019). The key findings of this study indicate that two of these subgroups, historians and undergraduate students, overly relied upon easily manipulated website features such as site design and domain names and often read vertically, staying too long on the site without learning much additional information that could be used to accurately evaluate the content (Wineburg & McGrew, 2019). Professional fact checkers, conversely, read laterally, leaving the sites their investigation originated from to corroborate material using other sources and to examine site features, such as authorship and ownership, through secondary and external means, learning more information in a shorter period of time (Wineburg & McGrew, 2019). Wineburg and McGrew (2019) note that this approach, lateral reading, resulted in the professional fact checkers identifying credibility concerns, such as a fraudulent or fictitious publisher or piece of fabricated information, far more frequently and in significantly less time than the historians and undergraduate students and suggest that lateral reading be incorporated into existing efforts to improve information curation. Lateral reading is also consistent with earlier recommendations to start with trustworthy sources, compare them to each other, and attempt to corroborate information (Meola, 2004; Ostenson, 2014). Information consumers have basically been encouraged to think like journalists and fact check information before they use it in an argument.

Unfortunately, most information consumers do not take the time to consider the source (Bråten et al., 2016; Metzger, 2007; Ostenson, 2014; Pearson, 2020), read laterally, or verify information (McGrew et al., 2019; Wineburg & McGrew, 2019), and instead rely on their networks, and shortcuts like heuristics, for information evaluation and credibility assessment (Flanagin & Metzger, 2007; Metzger et al., 2010). In fact, in a study of 259 undergraduate students, researchers found that the majority of participants ignored textual indicators of source quality (such as typos or a lack of references) and relied upon their personal experiences and opinions when evaluating a source for credibility (Bråten et al., 2016). Similarly, an experimental study of 513 students designed to elucidate the impact of source blindness, or processing a source without considering its origin, common when consuming information through curated social networks, revealed that social media users are cognitively lazy and inattentively process information (Pearson, 2020). A third experimental study (n = 2,146) which exposed participants to comparable information with differing site design heuristics (i.e. showing the information on Fox News' website versus CNN's) further revealed that, when people are exposed to attitude consistent or ideologically consonant information they tend to judge that information as more credible than cognitively dissonant information (Metzger, 2020). This reliance on networks and heuristics, such as site design or a known source, is a problem, particularly given the social, cognitive, and behavioral influences that impact network- and heuristic-based decision making when it comes to information, as discrepant views (i.e. actively disagreeing with something you read) prompt heightened attention to source evaluation (Bråten et al., 2016; Metzger et al., 2020; Pearson, 2020). Thoroughly understanding how individual information consumers approach processing and curating information, and how social, cognitive, and behavioral influences may impact these approaches, is thus crucial for understanding how information and networks that relay information are curated and how individuals engage with digital information. Given the reality that political ideology is readily performed in online spaces and reflected in networked information consumption (Havey, 2020b; Metzger et al., 2020; Pearson, 2020), understanding information curation with respect to political positions is crucial to this study.

PART II: Approaches to Processing Information

Nobel prize winning economist Daniel Kahneman's (2011) book *Thinking, Fast and Slow* is perhaps the most visible scholarship on intuitive, rapid-paced, emotional, and instinctual thinking (thinking fast) versus calculated, slower, more deliberative and systematic thinking (thinking slow) but his work extends a long history of scholarship exploring how individuals approach and process information. In *Thinking, Fast and Slow* Kahneman (2011) provides case study after case study that detail incidents of people rushing to snap judgments when triggered by a cue, bias, or other heuristic that aligns with their thinking on the subject presented and spending significantly more time and effort when these cues either did not exist or contradicted the person's priors. Existing research in credibility evaluation corroborates this (Metzger et al., 2020; Pearson, 2020). Thinking fast and thinking slow are examples of dual-process models. Prior to Kahneman, researchers in social psychology and consumer research defined, described, and experimentally validated several dual-process models which assess both a person's motive and the content they are evaluating when it comes to information processing.

Dual-Process Models

One of the earliest dual-process models developed was Chaiken's (1980) heuristic versus systematic model for information processing, which drew on both the source and message cues when evaluating how a person would react to persuasive messaging. Chaiken argued that, given information to evaluate, particularly persuasive content, people would pursue two main strategies. The first, heuristic information processing, relies upon recognizable cues that users might pick up on and react to and can be associated with Kahneman's (2011) conceptualization of thinking fast. Through heuristic information processing, information consumers are able to make quick judgments about the quality of the content they are consuming and can move forward with the assumption that this information is relevant and appropriate to their interests. Heuristics could include identifying the author of a source as trustworthy, recognizing the domain name of the website, or trusting the visual presentation of the site (Chaiken et al., 1996). The systematic model for information processing describes a more rigorous and deliberative approach. Systematic approaches to information processing involve more thorough attempts to process information through careful consideration of the source, examination of the context, deep thinking, and intensive reasoning regarding potential biases (Chaiken & Ledgerwood, 2011; Chen & Chaiken, 1999). The systematic model of information processing is more likely to be pursued when the information consumer cannot identify cues or heuristics that rapidly influence their perception of the content or when those heuristics prompt cognitive dissonance. Chaiken's (1980) systematic model of information processing can thus be thought of as Kahneman's (2011) thinking slow.

Petty and Cacioppo's (1981, 1984) elaboration-likelihood model of persuasion similarly describes these two potential routes for information processing. Describing how people are persuaded through messaging, Petty and Cacioppo explain that elaboration is the effort required on behalf of the individual to evaluate a message or content, consider their evaluation in context with their other beliefs, and decide whether to accept or reject it. Similar to Chaiken's (1980) models, their elaboration-likelihood model explains that individual information consumers can pursue two separate routes: a central route and a peripheral route. The central route requires a higher degree of elaboration and thus effort; when individuals pursue the central route they are thinking more slowly and evaluating the information presented in context and through more deliberative means. When they pursue the peripheral route, individuals are thinking fast and opting to rely on heuristics to point them towards a decision of either rejecting or accepting the information presented. Crucial to these dual-process models are heuristics.

Heuristics

Prominent among the dual-process models for information processing is the concept of the heuristic. Heuristics at their most basic level are cues that point a reader or information consumer towards a decision. In general, heuristics can be thought of as mental shortcuts or simple tricks that make decision making easier. An example of a heuristic might be identifying whether a piece of information has a listed author or byline, as information with a byline has been perceived as more credible than information presented without one (Metzger et al., 2020). When assessing information credibility online, heuristics are often utilized in lieu of slower, more deliberative assessment (Chaiken, 1980; Petty & Cacioppo, 1981, 1984, 1986; Fogg, 2002, 2003). Fogg described this in his (2002, 2003) prominence-interpretation theory. Put plainly, Fogg describes that, when presented with information, consumers will identify something such as a cue or a piece of information relevant to their reason for consuming the content, assess the prominence of that piece of information, and interpret the content accordingly. In practice, this may look like a person reading an article about climate change, noting that the author is a credentialed climate science researcher, and subsequently assessing the information presented as credible. It may also be reflected along ideological lines, with a person identifying a source, assessing that they align with that source ideologically, and subsequently choosing to process that source as credible without a deep investigation of its content (Metzger et al., 2020).

Extending Fogg's work, other researchers have identified that site features and information verification influence the perceived credibility of presented information (Flanagin & Metzger, 2007). A study of 574 participants designed to investigate whether site design influenced a person's assessment of the content located on that site revealed perceptions of credibility varied but that credibility assessments were primarily driven by website attributes such as design features, depth of content, and site complexity and that easily manipulated features central to site design can heavily influence perceived credibility (Flanagin & Metzger, 2007). Within the health environment, site design has been a crucial predictor of perceived credibility, with worse looking sites being assessed as less credible than well-designed ones (Freeman & Spyrikadis, 2004). Site design and site appearance has further been identified as a key heuristic for assessing source credibility online, specifically in regards to how well formatted a site is, whether there are typos, and what the overall degree of presentation is (Flanagin et al., 2020; Hong, 2006; Wathen & Burkell, 2002). In experimental conditions, Flanagin and Metzger (2020) identified that information consumers were likely to rely on the visual presentation of a website and whether the source of information, either a byline or the publisher of the site, was readily identifiable when evaluating the content presented. In several experiments with college students, McGrew and colleagues (2019) and Wineburg and McGrew (2019) found that students similarly relied on heuristics and were quick to assess a site as credible if there was an easily identifiable source, even if that source was a fictional nonprofit or publisher. In other studies, researchers found heuristics were used more frequently when there was a social alignment between the reader and the content (Metzger et al., 2010) and that information consumers in general rely on cues and heuristics rather than systematic evaluation of content (Flanagin & Metzger, 2020), which other researchers have noted leads to ill-informed fast thinking, source blindness, and worsened source cognition (Pearson, 2020). The predominant use of heuristics in information processing also aligns with Pirolli's (2005, 2007) information foraging theory.

Information Foraging

Information foraging theory, developed by Pirolli (2005, 2007), describes an approach to processing information that draws upon Darwinian conceptualizations of energy expenditure and reward. Put simply, information foraging theory explains that information consumers are likely to do the minimal amount of work for the maximum amount of reward (Pirolli, 2005, 2007). Experimental studies that asked students to identify and evaluate information pertaining to a particular question demonstrate that this reality is true, particularly for contemporary students who often resort to the first Google search (Bhatt & MacKenzie, 2019; Bråten et al., 2016; Hargittai et al., 2010; Julien & Barker, 2009). Specifically, people "tend to optimize the utility of information gained as a function of interaction cost" (Pirolli, 2005, p. 351). The theory was developed in response to a need to investigate strategies and technologies for information seeking, gathering, and consumption and how people adapt to the changing amount of information in an environment at any given time. The theory is centered upon the idea that, when possible, information consumers will adapt their strategies, or the structure of their information environment, to maximize their rate of information flow while minimizing their overall effort (Pirolli & Card, 1999).

Summary

There are many approaches to processing information but the literature does suggest that 1) information consumers are likely to limit the amount of effort they expend to gain access to and process information (Pirolli, 2005, 2007) and 2) that in reducing this effort information consumers are likely to opt to think fast rather than think slow (Kahneman, 2011), opting for approaches that center heuristics and other cues that minimize cognitive expenditure (Chaiken, 1980; Flanagin & Metzger, 2020; Petty & Cacioppo, 1984). This is readily supported within the empirical literature (Bråten et al., 2016; Julien & Barker, 2009; Metzger et al., 2020). Given the subsequent reality that most information processing is likely to rely upon strategies designed to encourage fast thinking, understanding what motivates the evaluation of heuristics is crucial. As more recent research (Havey, 2020b; Metzger et al., 2020; Pearson, 2020) suggests that information processing is linked to an information consumer's ideological alignment with the content they are consuming, understanding the cognitive, social, and behavioral influences on information processing is paramount.

PART III: Social, Cognitive, and Behavioral Influences on Information Processing

While the above approaches to processing information describe multiple clear pathways individuals may pursue when presented with information, they are not enough to describe how contemporary information consumers reconcile their own ideological positions, biases, and the biases of the information they are consuming when evaluating content. As a result, I also consider the cognitive, social, and behavioral influences endemic to contemporary information environments, particularly the curated information networks facilitated by social media that some refer to as “filter bubbles” (Pariser, 2011). Specifically, I describe how contemporary internet users and social media platform users, and within this study specifically student Twitter users, are influenced by the cognitive and socially driven behaviors of selective exposure and homophily when it comes to forming their social networks and consuming information.

Selective Exposure

When it comes to social media, users have the ability to curate their own networks and filter what they see and do not see (Kwak et al., 2010; Myers et al., 2014; Noble, 2018; Stedman, 2020; Steinert-Threlkeld, 2018). This can result in selective networks that only feature users and content that align with the original user’s positions, political stance, and interests (Garrett, 2009). This intentional curation of content and network through ideological alignment, specifically political alignment, is what is referred to as selective exposure. Selective exposure describes the proclivity for people to pick and choose what they consume with respect to information. With respect to this study, selective exposure is largely used to describe social media users who only consume information consonant with their existing sentiments, i.e. they will not expose themselves to cross-cutting or ideologically-opposed information intentionally.

Selective exposure has also been described as in-group siloing or out-group exclusion, the best example of which is Tatum’s (2017) book *Why Are All The Black Kids Sitting Together in the Cafeteria?: And Other Conversations About Race.* While Tatum’s research is specific to the school environment, selective exposure has also been discussed with respect to college students, most recently in Park’s (2018) *Race on Campus: Debunking Myths with Data*, in which the author describes the reality that many student subpopulations, such as white students, Greek students, and religious students are engaging in in-group siloing and isolating themselves from their out-groups, effectively engaging in on-campus selective exposure. These students are functionally choosing who they interact with and what social and cultural topics they have to care about as a result. In the online ecosystem, Barberá's (2015) massive, international study of Twitter users (n > 30,000,000) found that most users are engaging in similar social behaviors, selectively curating their networks to align with their ideological beliefs. Barberá also identified that the users he selected were fairly bimodal, mostly engaging with other like minded users, and that the degree of cross-cutting interaction (i.e. conservative to liberal, liberal to conservative), was significantly smaller for conservative users than for liberal users, suggesting a higher degree of selective exposure the further right you identify along the political spectrum.

This is a problem online, given the ease with which a user can completely ignore content that they disagree with and the consumer-driven nature of online media competition. Within the context of the online media ecosystem, competition for audience share has led to more targeted news that is designed and driven to get clicks (Munger, 2020; Metzger et al., 2020; Ribeiro et al., 2017). Most of these clicks are pushed through social media and come as a result of intra-network sharing such as a retweet, which can provide an initial or origin tweet a multiplicative reach within less than a day of the first share (Kwak et al., 2010; Pearson & Knobloch-Westerwick, 2018). This has resulted in news consumption being dominated by aggregators and social media which has in turn resulted in the creation and curation of filter bubbles and echo chambers (Barberá, 2015; Garrett, 2009; Himelboim et al., 2013; Pariser, 2011) in which people intentionally select, follow, and consume information from media outlets and other users whose beliefs are consistent with theirs (Colleoni et al., 2014; Metzger et al., 2020). People increasingly trust their feeds to make choices for them (Bennett & Pfestch, 2018; Bright, 2018; Metzger et al., 2020; Pearson & Knobloch-Westerwick, 2018), meaning that many users' information networks have devolved into what Iyengar and Hahn (2009) refer to as "red media" and "blue media."

With respect to information consumption, other research has identified that selective exposure leads to further ideological retrenchment and polarization (Stroud, 2010) and that political party affiliation is a strong predictor of media selection, with political conservatives tending to select and consume more partisan media which subsequently influences their political and social positions (Price & Kaufhold, 2019). This partisan selective exposure is consistent with other research on confirmation bias, ingroup bias, and negativity bias, which describes how information consumers favor attitude-consistent messages and strive to reduce cognitive dissonance (Knobloch-Westerwick et al., 2017; Metzger et al., 2020; Nam et al., 2013). Another study of political ideology and how it influences people's efforts to avoid cognitive dissonance identified that conservatives as a group are more resistant to change and accepting of inequality due to a stronger need for order, weaker tolerance for ambiguity and threat, and higher pursuit of selective exposure than their liberal peers (Himelboim et al., 2013; Nam et al., 2013). The asymmetrical pursuit of selective exposure is also driven by conservatives' stronger relational needs and subsequently greater pursuit of in-group consensus (Himelboim et al., 2013; Stern et al., 2014). This all results in homophilous networks on both sides of the political spectrum, with tighter clustering on the conservative end.

Homophily

Homophily is a natural result of selective exposure. Homophily describes a tendency for people to seek out people with like traits. For instance, a group of white students sitting together in the cafeteria is homophilous and homogeneous when compared to a group of students that includes white students, Black students, and Asian students (Park, 2018; Tatum, 2017). Binder and Wood's (2014) ethnographic study of student political organizations suggests that in-group siloing and the pursuit of homophily is also endemic to politically driven students, such as the

Campus Republicans under investigation in their book. With respect to the contemporary online ecosystem and the information networks made possible by social network platforms like Twitter, the mutable and curatorial structure of the internet may facilitate more heterogeneity than the structure of the cafeteria, though students who are only going to sit with students who look like them are also more likely to curate their online presence in a way that preserves network homogeneity (Brundidge, 2010). Put more simply, there is markedly less cross-cutting interaction for more extremist proponents of any ideological position.

While the internet has indeed been considered and has been a vehicle for heterogeneous and diverse exposure to information (Colleoni et al., 2014), this potential for heterogeneity has not been adopted unilaterally and homogeneity persists at the ideological poles, though this is asymmetrical as conservatives have significantly higher political homophily than their liberal peers do on average (Colleoni et al., 2014; Stern et al., 2014). Further, the increasing fragmentation of what was previously called the online "public sphere" (Bennett & Pfetsch, 2018; Bright, 2018) has contributed to a greater reliance on networked information flows (i.e. getting news because someone you follow shared that news) that facilitate polarization and thus homophily. This pursuit of homophily is apparent in even the most nascent of Twitter accounts and, while networks grow, change, and stabilize over time (Kwak et al., 2010; Myers et al., 2014), they also become increasingly less reciprocated and are driven by homophily and account popularity, ensuring that "red" networks stay red and "blue" ones remain blue (Stepanyan et al., 2010). On Twitter, this makes network analyses that draw upon latent attribute inferences -- using a user's information to make other assessments about their online behavior -- easier, as user's networks tend to be predictive of their attributes (Lazarsfeld & Merton, 1954) and as assortativity remains influential (i.e. liberals are likely to connect with liberals and conservatives are likely to connect with conservatives; Al Zaman et al., 2012). Given that the majority of the data used in this study was calculated using latent attribute inference, this is crucial.

Growing homophily has also been driven by the emergence of elite-moderated discourse on social media, as platforms like Twitter have become a place for social and political elites (like former President Donald Trump) to complain and offer unmoderated critiques (Weeks et al., 2019). While users are likely to follow like-minded people and limit their networks through selective exposure, they are also likely to invite ideologically-aligned elites into their networks, a behavior which has only increased partisan fragmentation (Barberá et al., 2015; Weeks et al., 2019). The resulting sense of strong in-group identity, bolstered by elites who have a vested financial and cultural interest in maintaining clout and influence (Munger, 2020) can lead to the perception of out-group bias, subsequent emotionality around cross-cutting discourse, and further ideological retrenchment and polarization (Barberá et al., 2015; Weeks et al., 2019). Selective exposure and homophily can thus have a direct, and cyclical, influence on political polarization, subsequent access to information, and can influence how users are consuming and digesting that information.

Summary

Most people have a proclivity to avoid cognitive dissonance. With respect to information consumption and subsequent processing, this leads to selective exposure. Similarly, most people tend to want to associate with other people who share similar beliefs and interests; this leads to homophily. In the online media ecosystem, selective exposure and homophily may be driving the approaches to information processing described above, as dual-process models such as Chaiken's (1980) systematic model suggest that, should simple heuristics trigger cognitive dissonance for an information consumer, that consumer is more likely to pursue a slower and more deliberative process to evaluate the information presented (Kahneman, 2011). As Metzger and colleagues (2020) describe, biased people process information in biased ways.

PART IV: A Theory of Constrained Choice Online

As described above, students, and the general public, tend to rely upon their social networks as proxy heuristics when making choices about information curation and consumption. Given the selective exposure and homophily I described in the previous section, and the human proclivity to "think fast" (Kahneman, 2011), this can result in information consumers curating homogenous and constrained information networks and subsequently only seeing news and information they already agree with regardless of veracity. It can also result in the feeling that anything that a person disagrees with is fake news (Ribeiro et al., 2017). And as I have described earlier in this chapter, companies have a market incentive to provide and amplify content that users want. Regardless of veracity, content that is consumed is content that is profitable. The result is an information ecosystem that, at multiple points, prioritizes maintaining cognitive consonance, which creates asymmetries with respect to the quality and quantity of information available in the contemporary media environment and entrenches ideological positions.

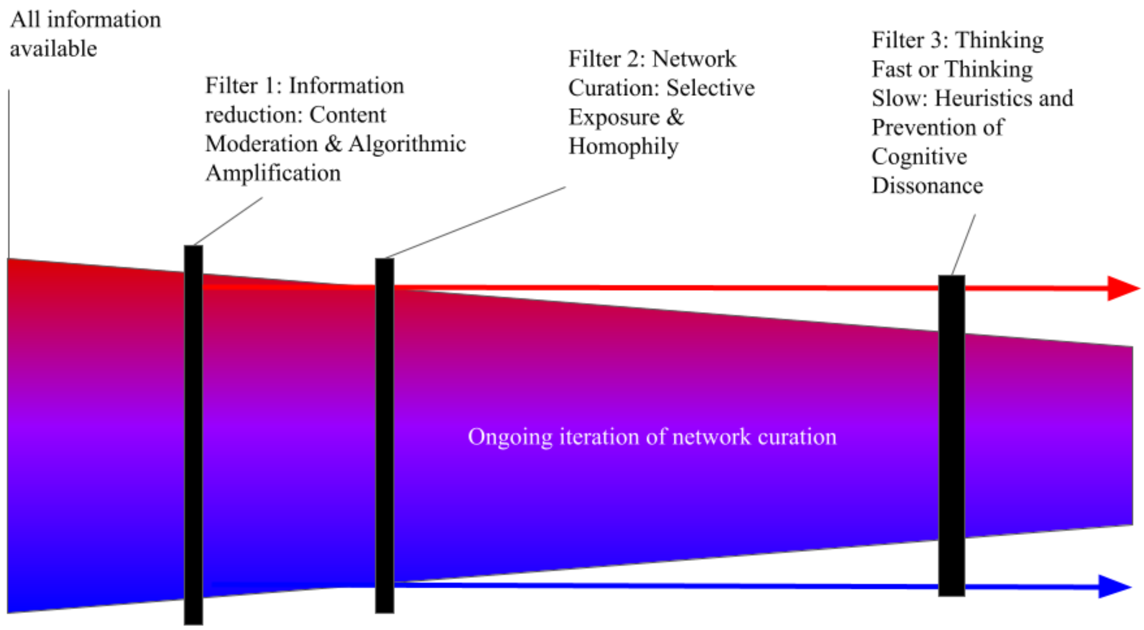
One example of the informational and ideological asymmetry caused by market incentives, selective exposure, homophily, and a reliance on heuristics for information selection is the partisan split in support for COVID-19 conspiracies on Twitter. Havey (2020b) demonstrated that conservatives were the primary discussants of most conspiracy theories and that support for these theories along partisan lines may have led to the worsening of the public health crisis and the COVID-19 pandemic. This is one small example that demonstrates the impact that political ideology can have on information and, thus, democratic society.

While this study is not explicitly about media or misinformation, ideology cannot be separated from information and digital participation has been consistently identified as a core competency for engaged citizenship in a participatory democracy (Mihailidis & Thevenin, 2013). This study thus seeks to assess how ideologically diverse the information students are accessing on Twitter are, how that information differs from their own calculated political positions, and whether there is overlap in terms of information sources across the political spectrum.

To clarify and explain the relationships between the changing information ecosystem and the market incentives that drive it, individual level consumer choices related to processing information, and the cognitive, social, and behavioral influences on those choices, I have created a diagram describing what I am calling a theory of constrained choice online. The diagram below (Figure 1) details the constriction of available information in the contemporary ecosystem towards the middle with preservation of extreme content on either end of the political spectrum. At points along the figure, I indicate filters (content moderation, algorithmic amplification, network curation, thinking fast, and thinking slow) that further constrain the information available to the modern information consumer.

Specifically, the figure suggests that, regardless of whether users in this study are thinking fast or thinking slow (Kahneman, 2011), their decision making, and the information available, is based on market incentives and responses such as content moderation and algorithmic amplification, and their own approaches to information curation. At each point, individual consumers' choices are constrained by outside features of the contemporary information ecosystem. This theory suggests that information seeking through social media, which most students engage in (Shearer, 2018), particularly on Twitter, which Kwak and colleagues have labeled both a social and information platform (Kwak et al., 2010), is driven by a desire to connect with friends and other accounts with shared interests but is mediated by factors outside the average users' control. As users continue to curate their networks, they are influenced by selective exposure and homophily as a result of an inherent desire to reduce cognitive dissonance. This likely results in highly curated information networks that reflect their in-group preferences. Once these curated networks are created, information sent through them is still processed either through thinking fast or thinking slow. Beginning with the understanding that most information consumers rely on heuristics when evaluating information (Flanagin & Metzger, 2020), the presence of positive heuristics, such as the information coming from a well-designed and known source, or negative heuristics, such as the information coming from an unknown source without a byline, influence subsequent decision making relating to information consumption. This decision making is also influenced by whether the information presented triggers cognitive dissonance (the user disagrees with the information) or cognitive assonance (the user agrees with the information), as information consumers are likely to attempt to mitigate cognitive dissonance when selecting and consuming information (Metzger et al., 2020). While the figure is presented linearly, I consider the cognitive and behavioral filters depicted to be iterative, as users may be influenced by them at multiple stages between curation and consumption. I have also included arrows depicting the preservation of extreme content within the information ecosystem, as content moderation and algorithmic amplification serve to maintain profitable, if extreme, content when it is performing well. These arrows indicate my hypothesis that, after an information consumer moves through the diagram making choices, they will likely find constrained options closer to the center alongside more extreme and polarized options at either end of their respective political pole with much of the content in inbetween each point removed via individual moderation or moderation at the company or firm level.

Figure 1: A Theory of Constrained Choice Online



Expectations of the Conceptual Approach

With respect to information processing, understanding the likelihood of both selective exposure and the pursuit of homophily in information networks is crucial. Given the changing online media ecosystem and the reality that users are functionally able to curate their information networks and feeds to exclusively feature content that they agree with, believe in, and want (Munger, 2020; Noble, 2018; Stedman, 2020; Steinert-Threlkeld, 2018), and the market incentives for content-producing outlets to respond to this user-driven demand (Munger, 2020; Noble, 2018; Roberts, 2017; 2021), knowing that users might be narrowing what they consume and facing a narrowed pool of options is imperative to answering this study’s questions. The utility of the above diagram, then, is that it merges understandings of approaches to information processing and social, cognitive, and behavioral influences on those approaches in a way that complements both strains of thinking and explains the emotional and socially-driven motivation beyond the seemingly mechanical decision making processes information consumers engage in.

Without the inclusion of selective exposure and homophily, which directly influences whether a user might think fast or think slow, understanding why an information consumer might pursue the peripheral route (here, thinking slow) rather than the central route (here, thinking fast) within the elaboration-likelihood model (Petty & Cacioppo, 1981, 1984) might be less clear. By incorporating the notion that users have biases and those biases may drive the ways that they react and respond to information, understanding the use of a quick heuristic (i.e. "That's a news source I trust!" or "That's a pundit I trust") becomes clearer, which in turn makes clearer how users of a particular political stripe might end up curating information of that same stripe.

There are some central assumptions to this approach, however. The first is that it assumes information consumers are making subconscious and conscious decisions regarding their information processing and that they are not, in contrast, making random decisions with respect to who they follow and what they consume. The second major assumption is that information consumers are trying to evaluate and process the information they are presented with in consistent ways, which, as discussed in the earlier portions of this chapter, may not be entirely accurate. The third assumption is that the first two assumptions are products of a constrained information ecosystem that is the result of content moderation and algorithmic amplification and that, regardless of a user's individual choices, the information available to them is largely controlled by market forces.

Application of This Theory

With respect to the study, the theoretical bases and theoretical diagram presented in this chapter animate research questions two, three, and five:

2) To what extent do the sources students follow on Twitter overlap ideologically?

a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?

3) How ideologically diverse are students' information sources on Twitter?

5) What institution-level features predict the ideological diversity of the information students are exposed to on Twitter?

The theory presented in this chapter also undergirds the core analysis of this study: latent attribute analysis estimation of political ideology. Understanding that the latent attributes that can be tied to a user– such as their friends or who they interact with and what they retweet, like, and share– allows us to better understand the user's own behavior and how concepts like selective exposure and homophily are influencing that behavior. Past the initial descriptive calculations central to this study, this theory will be used to explore how diverse the information students consume is with respect to their individual ideological positions, whether there is overlap between students of different political stripes, and whether there is ideological homogeneity or heterogeneity present. Specifically, I will utilize the theoretical diagram I described in this chapter regarding market incentives, cognitive and behavioral influences on processing information, and the information ecosystem to better understand how students are making choices regarding their information networks on Twitter and whether those choices are constrained by external factors. My analysis will be guided by the following questions: Are liberal students only engaging with liberal outlets? Are conservative students only engaging with conservative outlets? Is there an asymmetrical amount of crossover? Are there central users for more liberal networks? For conservative networks? Finally, I aim to determine whether there is, in fact, constrained choice online within this dataset. To that end, my analyses are also guided to answer the questions: are students able to access information that matches their own political positions? Are they choosing to? These questions, and their integration with the methodological approach of this study, are further detailed in Chapter 4 and answered in Chapters 5 and 6.

CHAPTER 4: METHODOLOGY

As discussed in the previous two chapters, political ideology can be expressed both explicitly (someone self-reporting their views) and implicitly (the same person demonstrating their political commitments through social networks and behavior). Similarly, while there are multiple approaches to processing information online and forming those social networks, understanding the company one keeps is important for understanding their behavior. To respond to concerns articulated in Chapter 2 regarding the quality of explicitly expressed political ideology, such as self-report survey data, and to highlight the social and political nature of decision making and information consumption online, I have chosen to use *in situ*, digital trace data as the primary data for this study and because it was offered outside of the primed conditions of contact with researchers, such as an administered survey or interview.

My choice to use this data stems from the belief that social networks provide relational and unobstructed data that other approaches to inquiry might not (Thomas, 2000). In qualitative work building towards this dissertation, interview participants explained to me that they consumed a wide diversity of news and information across the political spectrum (Havey, 2020a). Upon further questioning, these participants revealed that their information networks were almost unilaterally partisan and included several noncredible sources, suggesting that response bias may be an issue, given that participants are less likely to report engaging in socially stigmatized or undesirable behavior (Bradburn et al., 1978; Furnham, 1986). As I described in Chapter 2, this is consistent with other studies of students' political identities and expressed political beliefs (Bailey & Williams, 2016: Woessner & Kelly-Woessner, 2020). In response to those concerns, the digital trace data collected for this study resulted in multiple data points for each student, but primarily indicated 1) a point estimation of their political ideology based on who they interact with online and 2) an averaged value of the political ideology of the information networks they curated online.

This digital approach to understanding students' political identities and information consumption allowed me to examine the varying ways students construct their information flows and curate their social media feeds. This approach to data collection also ensured that what I collect as data reveals exactly what students see on their social media feeds, mitigating response bias with respect to what they might expect researchers to want to hear (i.e. "I consume a diverse swath of news!"; Bradburn et al., 1978; Furnham, 1986). The inclusion of secondary data (from IPEDS) allows me to explore how this localized and individual data may be associated with a variety of user, network, and institution-level features.

In summary, it is my aim to collect, analyze, and understand student-level political ideology and the collective political bend of the networks students consume and digest information from. This chapter focuses on the methodological approaches I used to explore students' political ideologies, the ideological diversity of the information they consume, and the heterogeneity or homogeneity of that information with respect to their peers. It also describes how the above-mentioned concepts varied across a variety of user and institution-level features. But first, a note on how I am operationalizing the many variables and measures I am collecting and analyzing within this study.

Connections Between Terms and Variables

In the last two chapters I introduced and explained several terms key to this study, specifically: homophily and ideological diversity. While the previous chapter described these concepts in the abstract, their connections, as well as how I am operationalizing them within the empirical part of this study, will be discussed below. I detail both terms in the next section and describe how they are being operationalized and measured within this study.

Homophily

As this study is focused on how selective exposure and homophily influence network construction and, subsequently, ideological position, it is crucial to understand how I am measuring and calculating homophily for each user. Homophily, as described in Chapter 3, is simply *like finding like* and choosing to be in community afterwards. With respect to this study, I am interested in how students pursue homophily driven by their political identity. I measure homophily in this study by calculating a student's estimated political ideology (described later in this chapter) and comparing it (calculating the difference) to the average estimated political ideology of the outlets in the student's information network (the news sources they follow on Twitter). Users with high degrees of homophily will have minute differences between their estimated political ideology and the average estimated political ideology of their network compared to students with more heterogeneous and diverse networks. The distribution of the estimated political ideologies for the outlets present in their networks (standard deviation) will also vary for students with more homogeneous or heterogeneous networks.

Ideological Diversity

Homophily can also help explain the ideological diversity of a user's network, specifically by highlighting how varied a particular user's network is from their own estimated political ideology. Ideological diversity in this study is reflected in a user network that includes information sources and outlets that span the political spectrum and vary from the user's own estimated political ideology. Given this study's focus on the ideological diversity of institutions, ideological diversity is also operationalized and explored empirically through analysis of the distribution of the point estimates for political ideology present across the dataset.

Summary

The concepts of homophily and ideological diversity are central to the questions at the heart of this study. Understanding their linkages, particularly how I am linking them within the empirical part of this study, is crucial to answering the research questions and testing the hypotheses described in the next section.

Research Questions

The purpose of this study is to explore students' individual ideological positions, and thus the ideological diversity of their institutions and the field of higher education writ large, alongside the ideological diversity of the information they are exposed to online, specifically on Twitter, and whether institution-level variables such as racial composition or selectivity influence ideological diversity. The following research questions guide the study:

1) To what extent is the political ideology of students active on Twitter skewed towards liberalism?
2) To what extent do the sources students follow on Twitter overlap ideologically?
a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?
3) How ideologically diverse are students' information sources on Twitter?
4) What institution-level features predict the ideology of students on Twitter?
5) What institution-level features predict the ideological diversity of the information students are exposed to on Twitter?

Research Design and Method

In order to answer the research questions, this study's quantitative design utilizes a multisite digital trace approach, basic descriptive statistics, and a set of predictive linear models. Passy and Monsch (2014), Thomas (2000), and other scholars have recognized the importance of social networks for examining complex behaviors and relationships. This study's focus on individual- (user) and aggregate- (institution) level data warrants a closer look at the networks that comprise both data strata. The subsequent individual-level data available is best analyzed through descriptive statistics and inferential linear models. Thus, this study utilizes a quantitative design to collect, analyze, and finally interpret user-level data in an aggregated context. The first step, described more thoroughly later in this chapter, involves identifying students across a variety of institutional types who maintain active Twitter accounts and extracting their information networks and estimated political ideologies. In the second step, students' data is collected and additional variables are calculated based on latent attribute inference and social network features. The analysis includes descriptive statistics at an aggregate level and inferential linear models incorporating both student-level data and institution-level data integrated from IPEDs. This design is ideal for a study on student social and information networks because it offers data that can make relational behaviors more plain (Thomas, 2000) and can describe the networked decision making strategies students engage in when it comes to expressing their political ideologies and curating information feeds (Shearer, 2018; Wineburg & McGrew, 2019) on a social and information platform like Twitter (Kwak et al., 2010; Steinert-Threlkeld, 2018).

In the sections that follow, I describe how I selected sites (institutions) to identify and include students from and how I both accessed the sites included in my analyses and did not access particular sites that were subsequently excluded. Next, I describe how I approached data collection, including identification and collection of primary (student-level) data, as well as the secondary data I integrated into my analyses from IPEDS. Next, I describe my approach to data analysis. I start by describing how I created and organized my data set, how I analyzed student- and institution-level network features, including descriptive statistics at the student level, how I approached aggregate, institution-level analyses, and what variables I included in my inferential linear models.

Following the presentation of my analytical approach, I describe relevant limitations and considerations for the research and close the chapter with a positionality statement that includes my own digital trace data and calculated variables.

Site Selection

Site selection for this multisite quantitative research study was informed by Binder and Wood's (2014) work detailing student political activity on a handful of campuses and my own work regarding student political organizing on three separate campuses (Havey, 2020a). Both Binder and Wood's (2014) multisite ethnography and my own multisite case study (Havey, 2020a) reveal stark differences in the strategies, behavior, and interests of politically-active students. Following these findings, I chose to include students from a variety of campuses representing a range of institutional types, student body compositions, and institutional missions. The intentional inclusion of a wide variety of institutional types responds to Binder and Wood's (2014) call to explore student political activity at a broader diversity of campuses and my recommendation to investigate qualitative trends quantitatively (Havey, 2020a).

I began site selection by considering my own campus, the University of California, Los Angeles, and the rest of the UC system. Due to the racial diversity of a variety of campuses in the UC system, as well as the varying degrees of institutional selectivity each campus boasts, I chose to include all UC campuses. I similarly chose to include the University of Southern California as a nearby, semi-selective, private institution which offered a different institutional type than the more accessible, public UC campuses. After identifying Californian campuses, I chose to include the other two sites from my previous work, the University of Arizona, a non-selective public land-grant university, and Harvard University, a highly selective, private university (Havey, 2020a). Pilot sampling indicated that it was difficult to identify students from colleges and universities whose institutions, students, and student organizations (such as the student newspaper, sororities, fraternities, campus political organizations, and the student government) lacked active social media presences or where there was not a strong institutional brand. As a result, I intentionally sampled campuses with strong institutional brands known across the United States of America and beyond. In the interest of national representativeness, I also identified campuses across the selectivity spectrum, across the nation geographically, and across a variety of institutional types, including liberal arts colleges, religious colleges, and community colleges, where I was able to identify students on Twitter. Other selected institutions include Brigham Young University, the Ohio State University, the University of Texas at Austin, Arizona State University, the University of Wisconsin-Madison, Marquette University, Clemson University, the University of Alabama, New York University, Southern Methodist University, Vassar College, and the University of Iowa. A full list of the institutions included in my analyses, as well as aggregated data for these institutions (number of students included and the average political ideology of those students) is included in Chapter 5 and Appendix B.

I sampled users from each site (the identified colleges and universities) by locating Twitter profiles that mention student affiliation at that site (i.e. “UCLA ‘23” or “Proud Bruin Class of 2022!”) and appear to be current students (have a profile photo and an active Twitter profile, including interactions with other students and student groups on Twitter). User identification is described more thoroughly in the section below detailing data collection, but was conducted through manual review of Twitter profiles associated with (following or followed by) institutional accounts (including student-run accounts that represent a student organization affiliated with the college or university).

Access to the Sites

After consultation with the University of California, Los Angeles' Institutional Review Board and careful review of Twitter's academic research policy, all data accessed for the purpose of this dissertation is public and not subject to any access concerns or institutional review board review. That being said, Twitter data belonging to a private (locked to public access) account cannot be extracted unless the researcher has been granted access to that account. While some students and student organizations have private, locked Twitter accounts (including a handful of College Republicans' Twitters and the students associated with those groups), the majority of institutional accounts (i.e. the University of California, Los Angeles' official twitter, @UCLA) are publicly accessible. While the data is theoretically accessible, actual access to students associated with each site is contingent on my ability to identify institutionally-affiliated accounts to review for users. My familiarity with UCLA and the University of Arizona, my alma mater, make this easy, but my lack of familiarity with other institutions can make this more difficult. I discuss this further within the limitations section of this chapter and describe my ethical position and the considerations I took with respect to data privacy and ethical concerns at the end of the chapter as well.

Data Collection

Data Sources

There are two data sources for this study. The first data source is primary and is drawn from identified students' Twitter accounts, in which I calculate a point estimate of their political ideology and the ideology of the media outlets they follow. These values were calculated using Barberá's (2015) R package tweetscores, which I describe more below. The second data source is secondary and comes from the United States Department of Education's Integrated Postsecondary Education Data System. The selected variables are described below.

Student Twitter Profiles

The first data source used in this study is primary and drawn from students' Twitter accounts. Students' Twitter accounts were identified through manual review of institutional and institution-affiliated Twitter accounts' followers and friends. My general workflow for student identification proceeded as follows and was designed to identify as many students per campus as possible and to maximize variance in political orientation and student experience (student journalists versus fraternity and sorority members, for instance) where possible. I also strove to include as much diversity with respect to student type as possible, sampling undergraduates, graduate students, and professional students. While I did not subsequently code for this due to inconsistencies in the data (i.e., if a student lists their graduation year, assessing their academic class standing is laden with assumptions, which is even more true for graduate students), the inclusion of graduate and professional students is a contribution and asset to this study that is absent in survey-based work that focuses exclusively on undergraduates.

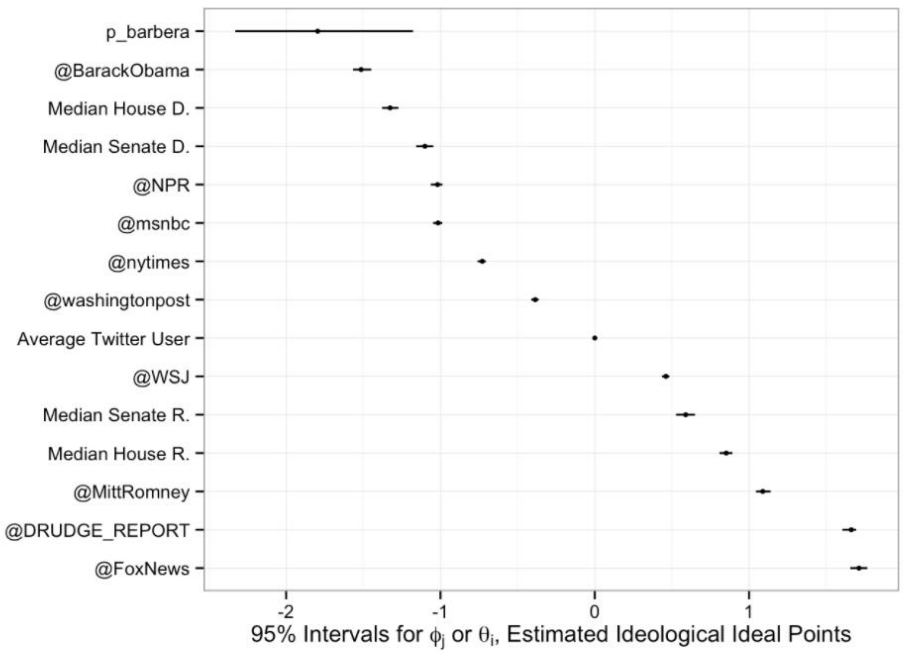
First, I identified the institution I was identifying students from (for instance, the University of California, Los Angeles). I began my search by identifying institutional and institutionally-affiliated accounts for that school, such as the school's main account (ex: @UCLA), the school's student newspaper (ex: @dailybruin), and the school's main student political groups (ex: @BruinGOP, @BruinDemocrats). After identifying an initial sample of institutionally-affiliated accounts, I reviewed the Twitter users that these accounts were following and followed by. The large accounts, such as the main UCLA institutional account @UCLA, which has over 200,000 followers, provided a broad field to extract users from so I started by reviewing who the account followed (200 users) and extracting accounts of interest, which included a variety of institutional centers and departments. These institutionally-affiliated Twitter accounts served as sampling centers from which I was able to identify individual students. Once I identified a sufficient quantity of institutionally-affiliated accounts / sampling centers, I reviewed those accounts' followers and friends (who they follow) for additional snowball sampling (Goodman, 1961) of accounts to review. I also simultaneously extracted individual student accounts who were followed by or following these accounts alongside my identification of additional sampling centers. I continued this process until I identified an adequate sample of students across a variety of experiences and political orientations. I aimed for a minimum of 100 students per campus, though some campuses in the sample have fewer students based on their overall enrollment and social media presence. Some campuses were easier to identify students for and thus boast a larger proportion of the overall sample. For instance, many community colleges are represented in the sample but in comparatively lower proportion (5-10 students versus 200-300 for larger schools). Where I could not identify any students attending a school, I removed that school from the dataset. In the interest of institutional diversity, I included these smaller subsamples in the larger dataset, though I do not plan to make any generalizable claims regarding those institutions I lack adequate data for.

After identifying a sufficient number of students for the campus, I calculated each student's Twitter User ID (an alphanumeric code unique to their account that does not change even if they alter their username) and subsequently calculated their estimated political ideologies using Barberá's (2015) tweetscores R package. Barberá's tweetscores package utilizes a Bayesian ideal point estimation approach to categorize users in comparison to a training set of politically elite users (Barack Obama, Hillary Clinton, Bernie Sanders, Glenn Beck, etc.) and assigns them a political ideology score based on their association with Twitter elites. The general workflow for the tweetscores calculation of estimated political ideology is to input a user (i.e., the author of the package, Pablo Barberá) and record the elites they follow as shown below:

```
## p_barbera follows 11 elites: nytimes maddow caitlindewey carr2n fivethirtyeight
NickKristof nytgraphics nytimesbits NYTimeskrugman nytlabs thecaucus
```

After these political elites, whose political positions are relatively stable and easily estimated based on voting records or well-documented stances in media (for some like Rachel Maddow, simply watching her show daily indicates her political position very clearly), are identified, relation to each elite is analyzed and a subsequent summary value for the initial user is calculated. This relationship is visualized below using a subsample of elites included in the package. The graph below shows Pablo Barberá's relatively liberal position with respect to the elites included in the data set. For instance, he is to the left of former President Barack Obama.

Figure 1: Sample Tweetscores Estimation



As the tweetscores tool has already been well validated in the literature, I did not engage in successive validation within this study, though I did manually review estimated political ideologies for users identified from select political student organizations (i.e. the Bruin Republicans) as well as data for myself and a handful of colleagues whose politics I know well to verify the package was working as intended. I found no discrepancies. I also used this data to construct qualitative profiles with respect to each estimated political position.

With respect to understanding what these estimated values for political ideology mean for a particular user in a broader context, for instance with regard to their daily behavior, I include sample profiles for each subset of the data (i.e., liberals; conservatives) at the start of Chapter 5. Dr. Barberá, for instance, could be presented as an example of a relatively strong liberal or member of the 'far left.' He is significantly to the left of sample elected politicians (e.g., Barack Obama, or his elected senators, such as Nancy Pelosi), to the left of what is considered mainstream liberal media (NPR), and almost the polar opposite of what would be considered mainstream conservative media (Fox News). A closer look at his Twitter profile indicates that he mostly interacts with liberal news outlets (i.e, Rachel Maddow, reporters and columnists for the *New York Times*) and that much of his online presence is directed towards discussing political topics (he is a professor whose work focuses on polarization). If I were presenting him in profile, I may simply refer to him as ‘The Liberal Politics Professor,’ whereas another example, such as a student leader from a conservative political group, might be labeled ‘The Outraged Young Republican’ when presented in profile. These profiles are more thoroughly discussed in Chapter 5 of this dissertation to more clearly contextualize the data and validate its utility.

With respect to the actual quality of the data, the construction of these profiles served as a qualitative investigation and validation of the output. Put plainly, I reviewed the estimated political ideologies of sample students who I am familiar with (for instance, the president of a club I have conducted ethnographic fieldwork on) to ensure the estimations were logical and the output I subsequently analyzed appropriate for my questions. The tweetscores package was also validated on a dataset comprised of millions of tweets and millions of unique users discussing a variety of extremely political and polarizing topics (such as elections) and more apolitical (sports) topics and has been validated by other researchers, including myself (Havey, 2020b).

In addition to my manual review of the data, I engaged in multiple methods of estimation for users and reviewed discrepancies between these methods of validation. Where association with elite users was unidentifiable for a particular student [i.e. a user failed to follow any of the politically elite users in Barberá’s training set (senators, politicians, partisan news personalities, etc.), secondary estimation of their political ideology was achieved through correspondence analysis, or analysis of a user’s discursive and relational habits (this analytical pathway leverages a user’s interactions with other users, here, students, who follow political elites for ideological estimation). Estimation using different methods, where available, revealed no statistical discrepancies (i.e., a t-test revealed that there was no significant difference between the means of the estimated political ideologies using method one [maximum likelihood estimation] and method two [correspondence analysis], $t = 0.056$, $p < 0.001$). The estimated political ideologies for student users fell on an existing scale (which includes Barberá's sample of political elites) ranging between -3 (extremely liberal) and 3 (extremely conservative) with moderates occurring roughly between -0.4 and 0.4, liberals occurring between -0.4 and -3, and conservatives between 0.4 and 3 (Barberá, 2015). Where students' political ideologies could not be estimated using either method, I assigned them a value of zero and removed them from the dataset prior to further analyses. This resulted in a sample reduction of between 8-15% per campus.

Following user identification, I extracted the user's information network (the news outlets and information sources they follow) using the Twitter API and Kearney's (2019) rtweet R package and calculated the average estimated political ideology of the network (also calculated using Barberá's (2015) tweetscores R package), the standard deviation of the estimated political ideologies of the outlets in the network, and the difference between the network mean and the student's estimated political ideology. For the purposes of this study, I am assessing the ideological diversity of a student's information network by taking the mean and standard deviation of the distribution of estimated political ideologies of each outlet the student follows. An ideologically diverse network should, given the possible range of estimated political ideologies (-3 to 3), be theoretically centered around zero and normally distributed (mean of zero and standard deviation of 1). I do not, however, expect that students' networks will be ideologically diverse or normally distributed and instead expect them to be highly skewed in the direction of their own political ideology (i.e. a student with an estimated political ideology of -

1.5 having an average network estimated political ideology between -2 and -1). To measure this, I will also record the difference between a student's estimated political ideology and the average estimated political ideology of their information network. Given the reality that conservatives are more tightly grouped, tend to pursue a higher degree of homophily (Barberá, 2015; Colleoni et al., 2014; Himelboim et al., 2013), and tend to select more partisan media than their liberal peers (Price & Kaufhold, 2019; Stern et al., 2014), I hypothesize that both the average distance between a student's estimated political ideology and their network's average estimated political ideology, as well as the standard deviation of their network's estimated political ideology (or the distribution), will be smaller for conservative students. Once I collected user-level data, I calculated the same variables for each institution in the aggregate.

Data from the Integrated Postsecondary Education Data System (IPEDS)

The other data source for this study is the Integrated Postsecondary Educational Data System (IPEDS). IPEDS is a product of the United States Department of Education and houses institution-level data collected from the institutions themselves regarding a variety of key variables. Following Binder and Wood's (2014) advice to consider how students behave differently and engaged with politics differently at different institutions, and my own work (Havey, 2020a) which suggests that conservative students at a nonselective institution behave drastically differently than their peers at a highly selective institution, I sought to incorporate institution-level data to assess whether there were significant differences for the student-level variable of interest to this study (student political ideology and their average network political ideology). To identify whether there were differences, I recorded institutional affiliation for each student included in my overall dataset and merged in IPEDS variables corresponding to their institution and its unique identifier.

I chose to include IPEDS variables for the sector of the institution (2-year public, for instance), the control of the institution (public, private), whether the institution was an HBCU or Tribal College, the size of the institution, the selectivity of the institution calculated as a percentage of students admitted versus students who applied, the total price of attendance for both in-state and out-of-state students, variables indicating the racial demographics of the institution (percent Black, for instance), and completion metrics for 4, 6, and 8 years for first-time, full-time students. I hope that the inclusion of these variables might shed light on some of the differences in data across institutions, as I informally hypothesize that some of the institutional characteristics defined by the aforementioned variables may influence students' political behaviors, ideologies, and subsequent information seeking patterns.

Data Analysis

Data analysis proceeded over the course of three main steps associated with the research questions. First, I created and organized multiple parallel central data repositories that include all information collected for outlets and information sources (including their Twitter usernames, alphanumeric Twitter user IDs, and estimated political ideologies), information collected for students (their Twitter usernames, alphanumeric Twitter IDs, estimated political ideologies, average network political ideology, standard deviation of outlet political ideology, difference between the average network political ideology and the user's estimated political ideology), and institution-level networks (the total number of students represented within the network, the average political ideology of those students, and the average network political ideology for the students representing each institution). The creation of these data repositories required student-level analyses, which I describe below. Following the creation of these central data sets, I conducted aggregate analyses to identify the overall average network political ideology and the most-followed outlets and information sources within the entire dataset. After I completed these analyses, I ran inferential linear models to identify whether institutional variables, such as racial composition, institutional type, and selectivity had any bearing on students' political ideologies and whether these variables had any bearing on students' information networks.

Creating and Organizing the Data

Outlets

The first step for this study was to create a repository of outlet-level variables for use in subsequent analyses. In my pilot analyses, I manually reviewed students' Twitter friends (who they follow) for information sources and news outlets to identify and include in my analyses. After reviewing 100 unique students, I began to stop identifying new outlets and decided to create an extraction file to compare students' friends to. Due to the saturation I faced early in manual collection, I chose to only include the 300 outlets I most frequently identified. I subsequently calculated each outlets' estimated political ideology (EPI) using the tweetscores package (Barberá, 2015). These outlets are included in Table 1 in Appendix A. The average estimated political ideology of these outlets is 0.4619, or just right of center.

Students

Following the creation of the central data set for outlets and information sources, I created a central data repository for the identified student Twitter user profiles within the study. As described above in the data collection portion of this chapter, I identified students at a variety of institutional types through manual review of Twitter profiles. After I identified the students in my overall sample, I utilized Kearney's (2019) rtweet package to extract both their Twitter user IDs and who they follow on Twitter (their networks). After extracting their overall Twitter networks, I used the outlets and information sources extraction file described above (see Table 1 in Appendix A) to pare down these networks to only reflect the Twitter accounts unique to their information networks (i.e. I removed friends and other accounts that were not identified within the extraction file). After reducing their networks to stable information networks, I merged the outlet and information source data (estimated political ideology) to each source and calculated their network's average estimated political ideology, the standard deviation of the outlets' estimated political ideologies, and the difference between the average network ideology and the user's ideology. After calculating these values, which incorporated data including multiple observations per student (who they followed), I reduced this data to reflect one observation per student and maintained the relational data for the analyses focused on news outlets, described below. Descriptive statistics regarding students, their estimated political ideologies, the homogeneity and heterogeneity of their networks, and the ideological diversity of their networks are presented in Chapter 5.

Information Networks

Following student-level analyses, I aggregated network data by institution to reflect descriptive statistics for each institution, including the distribution of student estimated political ideology, average network political ideology, average network ideological diversity (standard deviation of outlet ideology), and degree of homophily (difference between user estimated political ideology and their network's average political ideology). I then integrated variables from IPEDS corresponding to each institution's ID within the IPEDS data set. The selected variables are described below in the inferential linear models section.

Descriptive Analysis

To answer the first research question, regarding the possibility of ideological skew among students active on Twitter, I analyzed student-level data in the aggregate and at each institution individually. Specifically, I analyzed the distribution of the point estimations of political ideology for all students present in the dataset and calculated average variables for each institution. These analyses included identifying the distribution (averages and standard deviations) of estimated political ideology, average network estimated ideology, and total number of students for each institution. I also calculated categorical representations of the continuous values for estimated political ideology to more clearly represent the continuous data ranging from -3 to 3 in more relatable categories of far left, liberal, moderate, conservative, and far right. These calculations were necessary for comparison to other analyses of student political ideology, which I discuss further in Chapter 5 and Chapter 7.

One of the key facets of this study, as demonstrated by the discussion of the relevant literature and theory in Chapter 2, is how homophily, selective exposure, and network curation influence the creation and use of information networks. Given the reality that users are prone to engaging in selective exposure and homophily based on a desire to mitigate cognitive dissonance, I expect a user's own estimated political ideology to be reflected in their network. As I am not assessing user's connections to other users in this study, the potential for homophily and selective exposure to be reflected in a student's information network is limited to the outlets they are connected with. This in turn facilitates an assessment of the ideological diversity of information networks (the second research question).

To answer the second research question, regarding the extent to which students' sources overlap ideologically and the subquestion related to the ideological skew of these sources, I began by separating students into subcategories based on their continuous estimates of political ideology (a -2 became a far left liberal, for instance, based on a standardization of the data and separation of continuous values into quintiles) and analyzed which sources were present among each subgroup and identified overlap (i.e.Was Ben Shapiro only followed by political conservatives? Was the *New York Times* present in the information networks of all students?).

I also calculated the most followed outlets and information sources within the entire dataset and across each categorical value of estimated political ideology and evaluated the ranked list for ideological skew (i.e., Were the top 15 outlets in the entire dataset more liberal than they were conservative? Were they more conservative than they were liberal?).

To answer research question three, which explores ideological diversity across the networks, I calculated the standard deviations of each student network's average political ideology and the differences between those values and the user's estimated political ideology. Results for each institution's network averages, visualizations of ideological overlap, visualizations describing relationships between the variables described in the preceding paragraphs (such as the relationship between a student's individual point estimate for political ideology and the estimated ideology of the information sources they follow online), as well as the dataset's most central outlets and information sources, are located in Chapter 5.

Inferential Linear Models

To answer the fourth research question, I ran an inferential linear regression model. Specifically, using student-level estimated political ideology as the dependent variable, I ran a single linear regression model including all student-level data. I included the following variables as explanatory variables: the control of the institution, whether the institution was an HBCU or Tribal College, the size of the institution, the selectivity of the institution calculated as a percentage of students admitted versus students who applied, the total price of attendance for both in-state and out-of-state students, variables indicating the racial demographics of the institution (percent Black, percent White, etc.), a variable indicating the gender demographics of the institution (percent women), and completion metrics for 4, 6, and 8 years for first-time, full-time students. Results of the inferential linear models are presented in Chapter 6.

To answer the fifth and final research question, I ran another linear model with the dependent variable set as the average estimated political ideology of a student's information network (i.e., the average estimated political ideology of the sources and outlets they followed online). I included the following variables as explanatory variables: the student's individual-level estimated political ideology, the control of their institution, whether their institution was an HBCU or Tribal College, the size of their institution, the selectivity of their institution calculated as a percentage of students admitted versus students who applied, the total price of attendance for both in-state and out-of-state students, variables indicating the racial demographics of their institution (percent Black, percent White, etc.), their institution's endowment and assets, and completion metrics for 4, 6, and 8 years for first-time, full-time students. Results of these inferential linear models are presented in Chapter 6.

Limitations and Considerations

There are some key limitations to the data utilized in this study and the actual execution of the study as described in this chapter. First, not everyone is on Twitter. Twitter skews young, white, and educated (Kwak et al., 2010; Steinert-Threlkeld, 2018) which may have resulted in an overrepresentation of white, traditional students. It also tends to overrepresent men who are politically active (Barberá & Rivero, 2015). Similarly, a decent portion of the students I identified as conservative had private accounts, which may result in the final sample presented in this study skewing more liberal; this may be exacerbated by the general population demographics of students in the United States of America, with conservative students representing the smallest overall subgroup and moderates reining as the majority, though evaluating that skew is the purpose of this study. I also identified many liberal-leaning students (students involved in their campus Democrats clubs or marked by other more progressive activities such as a campus queer center) with private accounts, so this may not impact the sample too drastically. The sample might also overrepresent students who are already more involved than their peers purely by virtue of being on Twitter -- many of the students sampled in this study belong to student organizations and have Twitter accounts to participate in ongoing discourse central to those organizations' missions. For instance, student athletes and student journalists may be overrepresented in the sample because they were the easiest to identify during manual review of student populations on Twitter. Another student-level limitation is that, though I strove to ensure that the students I included in my sample were active on Twitter (i.e. tweeted in recently), some students may maintain Twitter accounts that they do not use as their primary social media platform. This might result in information networks and subsequent aggregate-level data that leaves out some of the other information seeking behavior students may be engaging in.

A second limitation of this work is its relative depth, which is more shallow than other studies that emphasize students' experiences, opinions, and behaviors. At an ideological level, a continuous value estimated per student based on their social networks and interactions with people in those social networks cannot provide the same level of detail and nuance that qualitative work might. For instance, a student could easily be estimated as conservative due to the overwhelming number of conservative people in their orbit, but this estimation might not indicate that that student's interactions are predicated on agreement with one facet of political ideology, such as immigration. While it is likely that interactional behavior with other accounts online (i.e., you would not necessarily follow someone who you agree with about immigration but have staunchly opposing views on abortion with) accounts for this, the actual estimation at the heart of this study leaves room for clarification and complexity that simply cannot be interrogated within the scope of this study.

A third limitation of this study regards sampling. Specifically, I could not feasibly sample every institution of higher education that exists within the United States of America nor could I easily identify every student present on Twitter. As a result, the dataset used within this study is inherently biased towards students with a heightened degree of institutional pride or at least pride for being a student and the data cannot be generalizable to every educational context in the country. That being said, data collection was designed to ensure representativeness nationally and at a variety of institutional types, as well as at a student-level, as I focused on identifying students from all walks of life (i.e., I did not solely include student journalists, for instance).

A fourth limitation of this study has to do with the outlets and information sources included in the analyses. At an outlet level, this study is limited by the availability of outlets present on Twitter. While the 300 outlets and information sources presented in the appendix offer a robust slice of the information landscape, they cannot be fully representative of how students are consuming information online. Similarly, several of the more conservative outlets identified had to be excluded from this study because they have been banned from Twitter for violating Twitter's terms of service and cannot be readily identified as being part of students' information networks. Several of the sample accounts for the estimations, such as congresspeople, have also had their personal accounts banned from Twitter for violation of the terms of service, though their congressional accounts persist. This may have limited some of the estimations.

Finally, this study exclusively looks at Twitter users, limiting its utility in describing contemporary student information seeking, as more and more students may be using other platforms such as Facebook, Instagram, and TikTok as part of their information consumption. While Twitter makes data far more accessible to academic researchers, future studies should consider qualitative approaches to investigating other platforms.

A Note on Ethics and Data Protections

As I mentioned earlier in this chapter, the data used in this study were publicly available and identified manually by me, the researcher. While this data was public (and where it was private and behind a locked account, not included), publicly accessible and available data is not always ethical to share, analyze, or use. Put plainly, the data in this study are people and people can be harmed by their inclusion in academic research like this study (boyd & Crawford, 2012; Zook et al., 2017). In the interest of protecting the students included in the dataset used for this study, I have disidentified the data and ensured that the analyses used to produce student-level variables (estimated political ideology) are slightly but negligibly different with each calculation (i.e., A student that is a -1 will probably be near a -1 on most calculations of their estimated political ideology, but will likely be a -1.01 or -1.03 each time the value is calculated, making it impossible to retroactively work out who each student included in the dataset is). The data in this study is also kept in a password-protected spreadsheet which only I, the researcher, have access to, though, in the interest of ethical data sharing, I do plan to make deidentified and un-reidentifiable data available to interested parties in the future where appropriate (Zook et al., 2017). Finally, I recognize that choices with respect to cleaning and using the data (i.e., which students to include, when to remove data that appears faulty, etc.) are subjective choices that are influenced by my positionality as a researcher, which I explain in the following section (boyd & Crawford, 2012).

Positionality of the Researcher

I close this chapter with a positionality statement. While positionality statements are less common in quantitative work like this study, I believe acknowledging my own position to the students who comprise the data in this study and my position to the research is critical in understanding my approach, my biases, and my rationale for the work (Hope et al., 2019). In my effort to present a critical quantitative positionality statement, I have included data indicating my own estimated political ideology and information network. I follow 17 information sources and outlets on Twitter that are also included in the 300 outlets included in the data for this study, including: N*PR's Consider This, The New Yorker, Teen Vogue, CNN Breaking News, CNN, The Associated Press, HuffPostEducation, HuffPost College, Unlocking Us with Brene Brown, FiveThirtyEight*, *NewsWeek*, *NPR*, *Vox*, *New York Times Opinion*, *Jezebel*, *The New York Times*, and *the Washington Post*. The average estimated political ideology of my information network is -0.2354406471. While my own estimated political ideology is -1.435039195 (solidly liberal), the average estimated political ideology of my information network is within the moderate range for the data, leaning slightly to the left, suggesting I consume a decently diverse news and information diet that is significantly more moderate than my own political identity. I do, however, not consume information from any outlets or sources that would skew my information towards the more conservative. With respect to this study, I endeavor to be cognizant of my positionality as a liberal Twitter user whose information network is markedly moderate and understand that I may encounter student political ideologies and information networks that are well-aligned with my own and also ones that are deeply dissonant. My goal is to honor the data I collect and analyze and present findings that reflect the reality of that data.

CHAPTER 5: FINDINGS (DESCRIPTIVE ANALYSIS)

The purpose of this chapter is to present the findings of this study. In an effort to keep the following chapters concise, readable, and organized around the research questions, I have separated the findings chapters, five and six, into two distinct chapters. The first of those two chapters, Chapter 5, focuses on the results of the descriptive analyses of the dataset that cover research questions 1-3. The second, Chapter 6, focuses on the inferential analyses, which cover the last two research questions. Guided by the theory I described in Chapter 3 and the methodological approach I described in Chapter 4, the descriptive results in this chapter are organized into four parts:

Part I: An introduction to the dataset constructed and used in this study

Part II: An exploration of students' political ideologies on Twitter

Part III: Sources and information diversity online

Part IV: Conclusion of descriptive findings

Guided by this study's research questions, which I have included below, Part I briefly reiterates the approach I used to collect data for this study and presents the dataset used in the study and for all analyses. Part II directly addresses the first research question, focused on students' estimated political ideologies, and provides the distribution of the estimated political ideology variable for the entire dataset, provides student profiles that contextualize the data through examples and close examinations of individual data points, provides the average estimated political ideology for select institutions alongside the number of students contributing to that average and a density distribution for all institutions included in the dataset, and presents values of the continuous estimates of political ideology represented as categorical values (far right, far left, etc.) necessary for subsequent comparison to previous studies of student political ideology using survey data. Part III presents the results that detail what sources and information students are exposed to online and how the estimated political positions of those sources reflect the political positions of the students that follow and interact with them online. Part III specifically addresses research questions two and three, which are focused on the ideological overlap of information students are exposed to and consume on Twitter and the ideological diversity, and skew, of those sources. This section includes the distribution of all sources followed by students in the dataset, overlayed distributions for each categorical subgroups' followed sources, an examination of the most prominent sources in the dataset for the categorical ideological groups calculated in Part II, and an examination of the relationships between students' estimated political ideologies and the distributions of their own information sources (the average estimated political ideology of the sources they follow, as well as the standard deviation) and the difference between these sources and their own estimated political ideologies. Part IV closes the chapter with a summary of the findings presented in this chapter and an overview of what will be discussed in Chapter 7, the discussion of these findings. The results of questions four and five, which focus on inferential linear models predicting the political ideology of students on Twitter and the ideological diversity of the information those students are exposed to on Twitter, are addressed in Chapter 6. The research questions informing both chapters are listed below.

1) To what extent is the political ideology of students active on Twitter skewed towards liberalism?
2) To what extent do the sources students follow on Twitter overlap ideologically?
a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?
3) How ideologically diverse are students' information sources on Twitter?

4) What institution-level features predict the ideology of students on Twitter?

5) What institution-level features predict the ideological diversity of the information students are exposed to on Twitter?

Part I: The Dataset

Creation of the Dataset

I constructed the dataset used in this study through the manual analysis of Twitter and the rigorous and systematic scouting of schools I described in Chapter 4. To reiterate, each student present in the dataset was identified through a manual review of the Twitter accounts associated with their designated school (for example: the University of California, Los Angeles) and confirmed through their self-representation as a student at that school (i.e., a student who followed the Daily Bruin, the University of California, Los Angeles' student-run newspaper, and identified themselves as the editor of the sports section and a member of the graduating class of 2023: "editor @dailybruinsports UCLA '23). I only included students who represented themselves as being existing students (i.e., students with graduation years earlier than 2021 were excluded from my analyses). Students who did not explicitly identify themselves as students at their designated school were not included in the sample (i.e., students who I suspected attended the school based on relationships and interactions with other accounts that explicitly claimed association with the school but did not explicitly claim that association themselves). Similarly, students whose accounts were private were not included in the sample, which I discuss in the limitations section of Chapter 4, and students who deleted their accounts since data collection were removed prior to data analysis.

Once students were manually identified, I collected their digital trace data using the rtweet package in R (Kearney, 2019) and estimated their political ideology using one of three methods available in Barberá's (2015) tweetscores R package: the first is a Metropolis-Hastings sampling algorithm, which samples the posterior distribution of the parameters of each students' data, the second a Maximum-Likelihood Estimation which is faster due to the fact that it does not sample the posterior distributions though it does result in smaller standard errors, and the third a correspondence analysis method which overcomes the limitations of the first two methods (if a user does not follow one of the sample accounts, their ideology cannot be estimated) through analysis of their networks and interactions within those networks (i.e., a student who followed no accounts in the sample used for the first two calculation methods but did follow and regularly interact with another user, or student, who did follow sample accounts and had a clearly estimable political position had their political ideology estimated through a function focusing on their peers). I chose to use the Maximum-Likelihood Estimation for the primary calculation of each students' political ideology, but employed the correspondence analysis method where the primary method was nonfunctional (the user followed none of the sample accounts necessary for calculation; Barberá, 2015). The calculation of this variable, the dependent variable at the core of this study, resulted in the removal of 8-15% of each school's student population, which is unsurprising given other studies utilizing the method (Barberá, 2015; Havey, 2020b).

Once I removed the students with incalculable ideological estimations, I recorded their estimated political ideologies (EPIs) and extracted the accounts they followed (their 'friends' in Twitter parlance) and used the outlets extraction file described in Chapter 4 to filter their friends to solely include information sources and outlets, such as *The New York Times*. Prior to student data collection, I calculated the estimated political ideologies of each outlet using the same method I used to estimate students' ideologies. Once I identified the outlets students followed, I reshaped the data to reflect individual edges (i.e., a student who followed multiple outlets, as described and demonstrated in the positionality section of Chapter 4, would have multiple observations indicating each outlet they followed) and merged in the estimated political ideologies of each outlet per student observation. Once this was done, I calculated the average of the estimated political ideologies of each outlet per student, the standard deviation of those estimated political ideologies, and the difference between the student's estimated political ideology and that of their information network. As some students did not follow any information sources on Twitter, their calculated values for their average network estimated political ideology and the standard deviation of that theoretical distribution are zero. For the analyses described in Part III of this chapter, I have removed these students from the edge dataset and the analyses focused on information sources but preserved their existence in the primary dataset focused on student political ideologies described in Part II.

Finally, I extracted data from the Integrated Postsecondary Education Data System (IPEDS) corresponding to each school included in the dataset and identified by their unique numeric identifier and merged relevant institutional data such as selectivity, racial composition, and cost to each student observation to facilitate the analyses described in Chapter 6. These variables, and their distributions, are described in Part I of Chapter 6.

The Contents of the Dataset

The dataset, comprised of student-level observations with variables for their estimated political ideology, the institutions they attend, the numeric identifier assigned to those institutions by the United States Department of Education, the average estimated political ideology of the information sources each student follows, the standard deviation of that distribution, the difference between the student's estimated political ideology and the average of their network, and institutional variables merged in from IPEDs, contains data collected from 8,554 students representing 139 schools across 43 states. The average number of students representing each school in the dataset was 61 students, with the smallest number of students (1) representing Avila College and the largest number of students (270) representing Central Michigan University. As discussed in Chapter 4, though the sample size per school is comparatively small for some of the smaller institutions (based both on total enrollment at that school and on the relative identifiability of students online– i.e., a community college with a smaller national brand and a higher degree of turnover than a prestigious four year institution was harder to locate students for), observations were kept to maximize variance and in an attempt to ensure national representativeness. Because the range of students at each individual school was large (269), I conducted sensitivity analyses to determine whether including students from institutions with smaller subsamples influenced the overall distribution of the data, specifically focusing on the distribution of the main independent variable: student estimated political ideology. To do this, I removed observations when individual schools had fewer than 50 students, fewer than 30 students, fewer than 20 students, fewer than 10 students, and fewer than 5 students contributing to the overall dataset and performed t-tests to assess whether the removal of these schools altered the distribution of the estimated political ideology of students within the dataset. After removing schools with fewer than 50 students ($n = 6,956$), there was no significant difference in means ($t = 1.409$); fewer than 30 students ($n = 7,785$), there was no significant difference in means ($t = 0.005$); fewer than 20 students ($n = 8,044$), there was no significant difference in means ($t = 0.084$); fewer than 10 students ($n = 8,426$), there was no significant difference in means ($t = 0.182$); and fewer than 5 students ($n = 8,502$), there was no significant difference in means ($t = 0.088$). As a result, I kept all of the observations. Keeping these observations in the dataset allowed for a greater overall diversity with respect to institution, state, and institutional type. Finally, the average number of students representing each state in the dataset was 198, with the smallest number of students (3) representing Mississippi and the largest number of students (953) representing California. Part II of this chapter explores this dataset.

The dataset used for this study also included edge data for each student observation based on the unique news and information outlets they followed. That edge data included variables for each individual student, the unique numeric identifiers of the outlets they followed, the Twitter username of those outlets, the official, registered name of those outlets (i.e., @NYTimes is *The New York Times*), and the estimated political ideology of those outlets. As I describe further in Parts II and III, the edge data also includes the student's unique estimated political ideology, as well as a calculated categorical version of their political ideology (i.e., a -3 on a scale ranging from extremely liberal to extremely conservative was subsequently categorized, based on the total distribution of the sample, as 'far left'). This calculation is described more thoroughly in Part II of this chapter. This categorical value for estimated political ideology, as described further in Part III, was used to identify source overlap along ideological lines, as investigation of overlap using the continuous variable for estimated political ideology was needlessly messy and difficult to interpret. As described above, the edge data reflects the reality that not all students in the dataset necessarily followed news and information outlets on Twitter. As a result of this reality, the edge data portion of the overall dataset contains 43,958 unique outlet observations representing 6,259 unique students. These students represent the same 139 schools present in the full dataset and similarly represent the 43 states present in the full dataset, albeit with smaller sample sizes per school and state (an average 45 students per school and 133 per state). The average number of outlets followed by each student is 7 and the range is 177 (a minimum of one outlet and a maximum of 178 outlets followed). Part III of this chapter explores this portion of the dataset.

Part II: Students' Political Ideologies

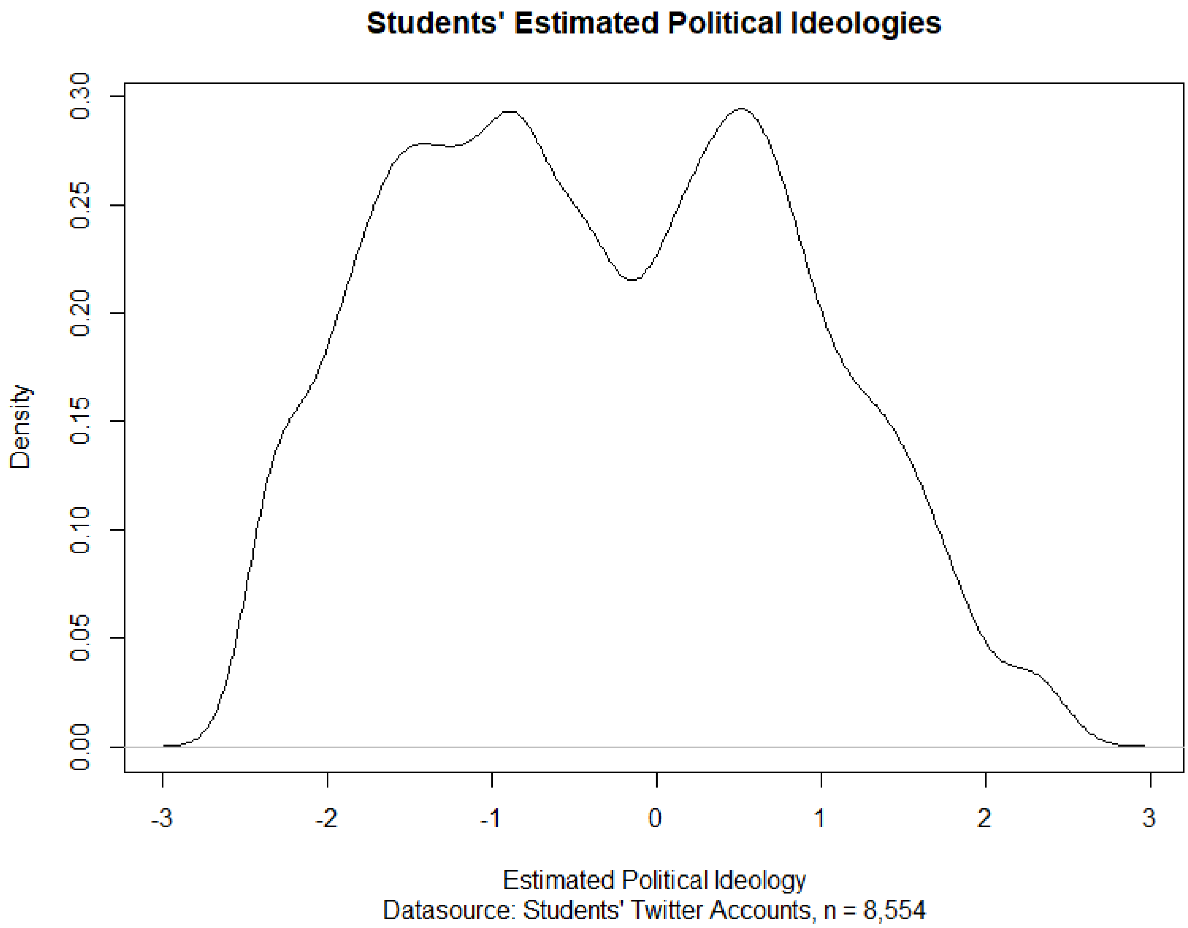
After constructing the dataset, I sought to answer the first research question: To what extent is the political ideology of students active on Twitter skewed towards liberalism? To do this, I analyzed the distribution of the variable of interest: students' estimated political ideology. In the following sections, I present and describe the distribution of estimated political ideology within the full dataset and provide a breakdown of continuous estimates of student political ideology after converting student-level values into categorical variables representing a five-point scale ranging from far left to far right. To contextualize these values, I also present student profiles for each categorical label (far left, for instance), and describe in detail what the online behavior of a sample student in that political subgroup looks like in practice. This facilitated clearer interpretation of the categorical variables I calculated to compare the data I collected in this study to other data on student political positions collected via self-report surveys which ask students to identify themselves on a five-point scale. I describe this choice more thoroughly later in the chapter.

Estimated Student Political Ideology

The full dataset of 8,554 students representing 139 unique schools and 43 unique states had an average value for estimated student political ideology of -0.337, a minimum value of -2.474, a maximum value of 2.449, and a standard deviation of 1.17. As a reminder, the range of possible values for estimated political ideology calculated using the tweetscores package lies between -3 and 3, though the majority of the estimations in this study and previous studies

(Barberá, 2015; Havey, 2020b) were between -2.5 and 2.5. With respect to categorical distributions, the 'center' lies roughly between -0.4 and 0.4, with values more negative than -0.4 being classified as liberal and values greater than 0.4 classified as conservative. The average estimated political ideology of the students in this dataset, then, lies at the slightly more progressive end of center, though the mean value is still squarely in the range considered moderate politically. There are also identifiable values for students whose political positions are at either political extreme. The distribution, as shown in Figure 2 (Distribution of Students' Estimated Political Ideology, Full Dataset), is roughly normal, with a mean centered slightly off zero and a standard deviation marginally larger than one. Visual inspection of the distribution of estimated political ideology, however, indicates that the largest peaks, both representing nearly one-third of the overall sample, lie closer to values of -1 and 1, indicating that while the average student is staunchly moderate, a good portion of the population is slightly off center politically. There also appears to be a greater number of students on the more progressive end (larger negative values) of the distribution, indicating a greater proportion of more liberal students. If the expected distribution of students' estimated political ideologies was perfectly normal, with a mean of 0 and a standard deviation of one, this dataset would show a slight skew towards the liberal end of the political spectrum.

Figure 2: Distribution of Students' Estimated Political Ideology, Full Dataset



To better understand the distribution of the sample with respect to estimated political ideology, I standardized the student-level data and created five categorical bins into which every continuous value of estimated political ideology fit ranging including far left, liberal, moderate, conservative, and far right. But first, a note on labels.

A Note on Labels Pertaining to Political Ideology Within This Study

One of the primary criticisms this study attends to is that previous analyses of students' political behaviors and ideologies have been almost entirely limited by survey data. Restricted by finite instrumentation, such as Likert-style responses that ask respondents to rate their political beliefs on a scale of very liberal to very conservative, and even sometimes in comparison to their peers, past work on student political ideology has been constrained by how respondents interpret each question and see themselves in response. Due to the reality that the majority of existing work on student political attitudes and behaviors in higher education has been conducted using survey data– specifically survey data collected by the Higher Education Research Institute (Astin, 1977, 1993; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Pascarella & Terenzini, 1991, 2005; Sax, 2008; Schiff, 1993)-- the limitations of survey data, and the labels that this data relies upon, are serious and cannot be ignored.

Specific concerns with using survey data for the evaluation of student political ideologies and behaviors revolve around the issues inherent with self-report survey questions such as "How would you rate yourself politically on a scale of 1-5 with 1 representing a liberal position and 5 representing a conservative position?" These questions are fundamentally constrained by a variety of factors such as social desirability bias, halo effects, and the formatting and clarity of items. Empirical research conducted to assess whether college students can accurately evaluate and subsequently report their own identities, learning, and development (Bowman & Seifert, 2011; Herzog & Bowman, Eds., 2011; Pascarella, 2001; Porter, 2011) indicate that these concerns are not merely theoretical but founded in the data. Self-report survey data is useful, but can also present confounding responses and data based on respondents' inaccurate interpretations of their own behavior. Political attitudes and behaviors are no exception.

Expecting students to accurately interpret what identifying as a contemporary liberal or conservative looks like in practice and asking them to map their own beliefs and actions onto that interpretation may result in imprecise and, potentially, intentionally misleading data (Bailey & Williams, 2016; Woessner & Kelly-Woessner, 2020). As I have previously found (Havey, 2020a), students often report or identify with a political orientation that is inconsistent with their beliefs. Some might even feel accurately identifying themselves as a particular political orientation, such as conservative, is akin to "academic and social suicide" (Havey, 2020a, p. 14) and may mask their beliefs on campus to avoid potential perceived repercussions.

While it may seem paradoxical given the critiques presented in the preceding paragraphs, the labels at the center of extant research on student political ideologies, such as far left, liberal, middle of the road, conservative, and far right, provide analytical and interpretational baselines for those conducting, consuming, and operationalizing the research. For instance, an individual student may not adequately interpret what it means to identify as a liberal within the contemporary political ecosystem and may subsequently over- or under-represent their own political position (i.e., a student who *believes* themself to be significantly more progressive than they actually are), but readers of aggregated data are quickly able to interpret and envision the state of a campus that self-identifies as almost exclusively liberal. And whether that interpretation, and the subsequent decisionmaking or policymaking that stems from it, is grounded in reality or adequately represents the findings of a research study, these labels provide a form of analytical stability.

This study was fundamentally designed to subvert these labels and improve upon previous work that has been limited by these labels but, in the interest of comparison and analytical clarity, the inclusion of categorical labels for political ideology was a useful tool. As a result, the following section contains calculated categorical values for the continuous data collected and analyzed within this study. The inclusion of these labels is, however, coupled with in-depth analysis and interpretation of what each individual label (i.e., liberal) represents for individual students. In this study, a label of liberal, for instance, should be interpreted as representative of online behavior that aligns with a liberal political agenda. In summary, previous studies have presented data that identifies people who have self-reported their political ideologies; this study relies upon data that reflects their lived-in, day-to-day political behavior and interactions with other people. The translation of a -2.5 with respect to a student's estimated political ideology to a label of far left, then, should be understood as an attempt to meet previous research on student political ideologies, enter into conversation with it, and expand upon it. This translability is one of the methodological strengths of this study.

The translated and contextually interpreted data for each student's estimated political ideology as a categorical variable is presented in the following section.

Estimated Student Political Ideology as a Categorical Variable

In the interest of more stable comparisons to previous, survey-based work that asked students to identify themselves politically on a scale including static positions (i.e., conservative, very conservative, very liberal. etc.), and in the interest of more easily interpretable data, I have converted the continuous variable I calculated for student estimated political ideology into a categorical variable with five, evenly spaced categories. I did this by standardizing the data and calculating quintile cutoffs, which resulted in five distinct groups I have designated as 'far left,' 'liberal,' 'moderate,' 'conservative,' and 'far right.' These five groups are separated as follows, given the full range of the dataset (-2.474:2.449), with students designated as 'far left' representing values (-2.48,-1.49], students designated as 'liberal' representing values (-1.49,-0.505], students designated as 'moderate' representing values (-0.505,0.48], students designated as 'conservative' representing values (0.48,1.46], and students designated as 'far right' representing values (1.46,2.45]. Compressing the continuous values for each student's estimated political ideologies into five categorical bins is useful for comparison to previous survey-based research, but first it is important to contextualize what these labels mean.

Student Profiles

To better understand what representation in each of these subgroups means, and to contextualize what each label means in practice, I have selected individual students, provided them pseudonyms, and presented their estimated political ideology below, alongside contextualizing information that describes why their estimated political ideology is what it is and how that information can and should be interpreted. I have also included information regarding other variables of interest in this study such as the standard deviation in a student's information network, the average estimated political ideology of sources in that network, and the difference between a student's estimated political ideology and the average estimated political ideology of their network. Later in this chapter, I use these profiles as examples to contextualize findings from analyses of the entire dataset.

Far Left: Berkeley Brian

An example of a student categorized as Far Left is Berkeley Brian. Berkeley Brian attends the University of California, Berkeley, tweets regularly about issues affecting Black and brown people, as well as queer people, and is in frequent communication with his peers. Berkeley Brian interacts with other Berkeley community members and is connected with other users who most recently have shared mutual aid requests and are vocal in their support of Palestine. Berkeley Brian specifically is very vocal about the problems with policing in the United States of America and is clearly supportive of abolition. Finally, Berkeley Brian lists his pronouns in his Twitter biography (they/he), a choice historically coded as progressive.

Berkeley Brian follows news and information outlets like *Teen Vogue* (-1.11), *Jacobin* (-2.42), *In These Times Magazine* (-2.02), and *Shadowproof.com* (-2.04), all of which are considered liberal or far left outlets within the dataset. Berkeley Brian's information network is, however, quite diverse, and features other outlets and information sources like *Politico* (1.22), *Current Affairs* (1.20), and *Russia Today* (1.22), all of which are considered conservative outlets within the dataset.

With an estimated political ideology of -2.45, Berkeley Brian is significantly to the left and is nearly two standard deviations (1.17) from the mean estimated political ideology of students within the dataset (-0.337). Berkeley Brian's average network estimated political ideology is -0.32, slightly to the right of the average user in the dataset but significantly more conservative than his own estimated political ideology, and the standard deviation of the information sources he follows and interacts with is 1.49, suggesting a broad and diverse spread of information. Finally, the distance between his estimated political ideology and the average estimated political ideology for the sources in his information network is 2.13, suggesting that, on average, Berkeley Brian consumes information that is significantly more conservative than his own political beliefs.

Liberal: East Coast Emily

An example of a student categorized as Liberal is East Coast Emily. East Coast Emily is a graduate student at Georgetown University and tweets regularly about issues impacting rural students, social mobility, and the difficulties inherent to working in the humanities. East Coast Emily is a small-town Nebraska girl at heart and regularly interacts with other graduate students, faculty members, and organizers focused on the issues facing rural and working-class communities and has written several opinion pieces regarding these issues. Her Twitter profile, in contrast to Berkeley Brian's, is particularly white and focused on issues of class. There are no mentions of race, or how the intersection of race and class influence an individual's experiences.

East Coast Emily follows significantly fewer outlets than Berkeley Brian, and gets the bulk of her information online from three sources: *NPR* (1.08), *NPR Politics* (1.15), and *The New York Times* (-1.57).

With an estimated political ideology of -1.35, East Coast Emily is less than one standard deviation away from the mean estimated political ideology of students in the dataset (-0.337). Her average network estimated political ideology is 0.22, squarely moderate, and significantly to the right of her estimated political ideology, with a standard deviation of 1.55. In short, East Coast Emily interacts with a limited number of information sources, but they are objectively diverse, covering more liberal politics alongside those that are center-right or conservative in nature.

Moderate: Just Josh

An example of a student categorized as Moderate within the dataset is Just Josh. Just Josh is a student at Yale University and his social media presence is almost entirely focused on interactions with friends and cataloging his daily life. His profile is markedly less political than Berkeley Brian's and East Coast Emily's profiles, and regularly features photos of him and his friends drinking alongside online conversations about sports. The most political feature of Just Josh's Twitter profile is his regular discussion of the difficulties inherent to being a student during the COVID-19 pandemic. Other than that, his profile is seemingly apolitical.

Just Josh follows more outlets than East Coast Emily and seems drawn to traditional, legacy media sources like the *Associated Press* (1.20), *The New York Times* (-1.57), *The Hill* (1.26), and *Politico* (1.22). Just Josh's information network is fairly diverse, with a standard deviation of 0.93 and an average source estimated political ideology of 0.33, moderate, but nearly one standard deviation to the right of the average student in the dataset.

With an estimated political ideology of 0.10, Just Josh is squarely moderate within the dataset and less than one standard deviation to the right of his average peer. His information network, however, is much more closely aligned than his liberal peers Berkeley Brian and East Coast Emily, with a difference of 0.235 between his estimated political ideology and the average estimated political ideology of the information sources in his network. While the information Just Josh is consuming on Twitter is closer to his own political beliefs, the average information source in his network is more conservative than he is.

Conservative: Southern Steve

An example of a student categorized as Conservative within the dataset is Southern Steve. Southern Steve is an undergraduate student at Clemson University, an elected student government senator, and a self-described First Amendment Advocate. Southern Steve's Twitter profile features frequent discussion about the perceived censorship of conservative voices on and offline, includes criticisms of other student government members as 'liberal snowflakes,' and boasts links to Southern Steve's contributions to *The Lone Conservative*, a national student-run opinion publication centering conservative students.

Southern Steve follows a handful of traditional, legacy media outlets like *The Hill* (1.26), *USA Today* (0.44), *NBC News* (-0.58), and *ABC News* (0.32). These outlets are all center or center-right. He also follows and interacts with more conservative and far right outlets like *Fox News* (1.50), *The Charlie Kirk Show* (1.01), and *The Ben Shapiro Show* (1.85). Southern Steve's information network is decently diverse, with a standard deviation of 0.75 and a mean estimated political ideology for sources of 0.70. This is nearly one full standard deviation to the right of the average student in the dataset.

With an estimated political ideology of 1.19, Southern Steve is squarely in the conservative category and nearly 1.5 standard deviations to the right of his average peer. His information network, unlike his more liberal peers, is slightly less conservative than he is with a difference of -0.48 between the two values. The information Southern Steve is consuming is significantly more aligned with his own political values, however, than Berkeley Brian (2.13) or East Coast Emily (1.57).

Far Right: Midwest Megan

An example of a student categorized as Far Right in the dataset is Midwest Megan. Midwest Megan is an undergraduate student at Central Michigan University and is a proud member of her sorority. Her Twitter profile prominently features posts about her sorority alongside support for Donald Trump, who she believes won the 2020 election, and information she has posted and shared regarding the dangers of the COVID-19 vaccines and masks. Midwest Megan regularly interacts with other members of her sorority and frequently retweets and replies to prominent conservative personalities like Candace Owens, Charlie Kirk, and Kayleigh McEnany. She is an active resharer of COVID-19 misinformation.

Midwest Megan only follows two information sources on Twitter: *The Ben Shapiro Show* (1.85) and *PragerU* (1.38), both of which are conservative entertainment outlets. Her average network estimated political ideology is 1.61, with a standard deviation of 0.32, and includes no moderate or liberal sources.

With an estimated political ideology of 2.39, Midwest Megan is nearly three standard deviations to the right of her average peer. Her information network, however, is significantly more moderate than she is, with a difference of -0.77 between her network's average estimated political ideology and her own, suggesting she consumes news that, while still significantly conservative (*PragerU*) and far right (*The Ben Shapiro Show*), is still slightly more moderate than her own political beliefs. That being said, her information network is more aligned with her views than her liberal peers Berkeley Brian and East Coast Emily, but less closely aligned than Southern Steve's and Just Josh's information networks.

The Political Categories

After assigning each student a categorical value based on the continuous estimate of their political ideology, I tabulated total frequency for each subgroup, and calculated the percentage each subgroup represented of the overall sample. Those results are below:

Far Left: 1721, 20.1%

Liberal: 2315, 27%

Moderate: 2069, 24.4%

Conservative: 1875, 21.8%,

Far Right: 574, 6.7%

When pushed into categorical bins, the results of this study seem to skew towards the liberal end of the spectrum, with nearly three times as many students designated as 'far left' than 'far right' and the majority of the population skewing left of center.

In comparison to some of the older research on student political ideologies, which reported stronger political centers with a clear skew towards the liberal end of the spectrum (2-3% far left, 25-29% liberal, 41-50% moderate, 22-27% conservative, and 1% far right; Dey, 1996), this study's findings indicate a reduction of the political middle and increases on both ends of the political spectrum, with a greater liberal skew. The most recent survey research (Havey, 2023), reports a slightly smaller center (35.2% moderate in 2019), with even greater liberal skew (40% liberal and 6.2% far left in 2019), and an overall reduction in students identifying as conservative and far right (17% and 1% in 2019).

In comparison to the data presented in this study, the survey data collected by the Higher Education Research Institute and analyzed by Havey (2023) underestimates conservative and far right percentages, overestimates moderates and liberals, and underestimates students on the far left. While these two populations are different and thus it is imprecise to conduct a statistical test to measure their difference, a simple t-test is illuminating. The mean of the estimated political ideologies for students in the dataset used for this study ($n = 8{,}554$) is 2.68 on a scale of 1-5 with 1 being far left and 5 being far right with a standard deviation of 1.17. The mean of the self-reported political ideologies of the students in the 2019 survey dataset Havey (2023) analyzed ($n = 11{,}160$) is 2.65 on the same 1-5 scale with a standard deviation of 0.87. A simple t-test of these sample populations ($t = 1.53$, $p = 0.93$) indicates that, while there is a visible difference in these populations, that difference is not statistically significant.

That being said, these populations are not the same and thus statistical comparisons should not be at the forefront of interpretation. The data presented in this study make clear that survey-based estimations of student political identities potentially overestimate particular ideologies, specifically moderate and liberal students, but underestimates far right, conservative, and far left students. The average student political ideology in both populations is center-left, but the distributions contributing to these means are visibly different and, as the data in this study indicate, likely undercount students at the political extremes.

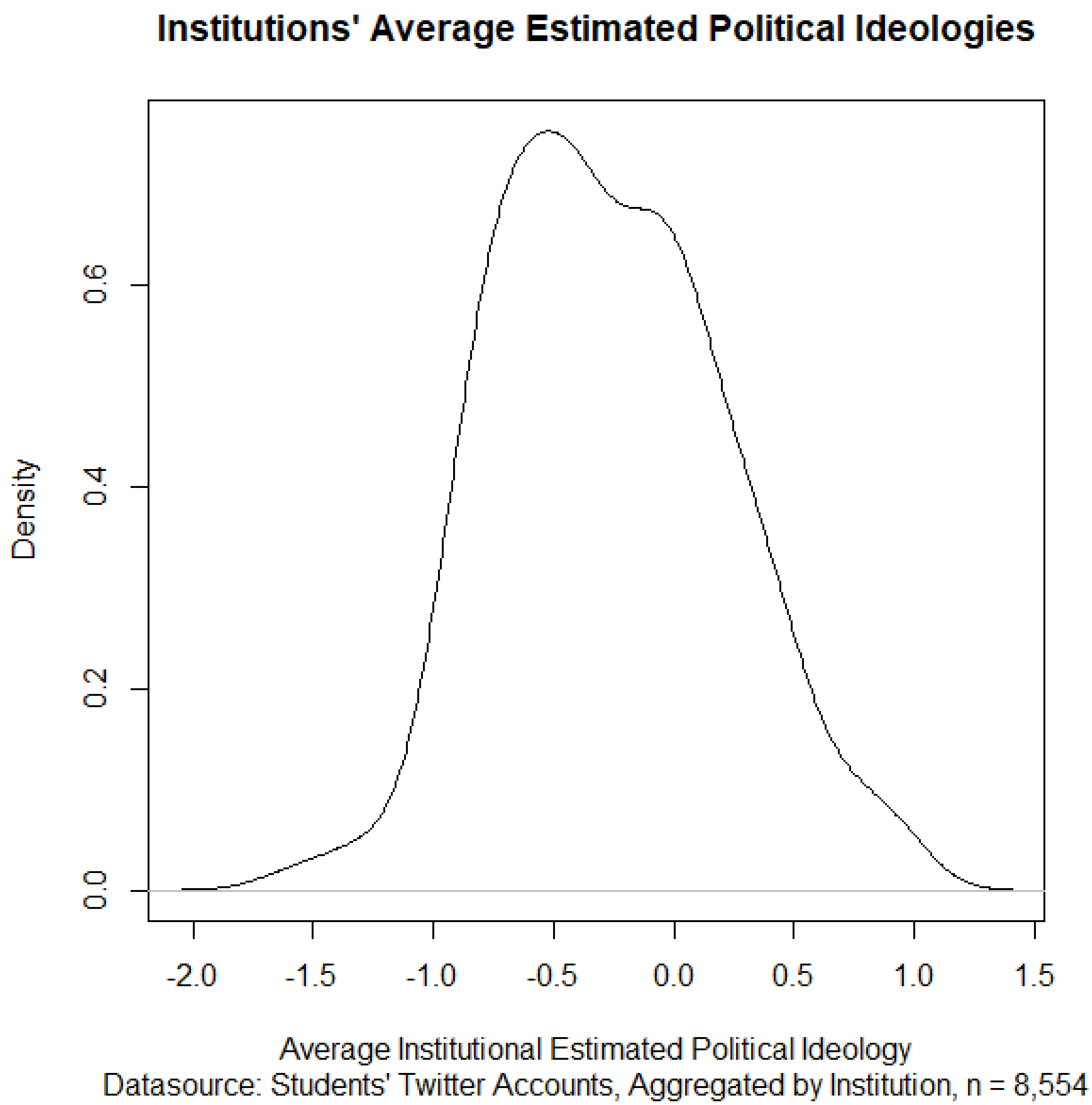
In sum, while the average student political ideology identified in this study and in other survey-based research (Havey, 2023) skews liberal, it is likely that survey-based research on student political ideologies overestimates this skew and underestimates conservative students.

The average estimated political ideology identified within this study's dataset, however, is not representative of every institution in the country and there are clear differences between schools. To consider these differences at an institutional level, I now turn to data by institution.

Average Estimated Political Ideology Per Institution

To more clearly understand how each individual institution was contributing to the distribution of the data I collected on student political ideology, I calculated an average of the estimated political ideologies of students at each institution as well as the number of students contributing to that average at each institution. In the interest of readability, I have appended the full table describing all institutions in the dataset, their average political ideology, and the number of students contributing to that ideology in Appendix B. Figure 3, Distribution of Average Estimated Political Ideology by Institution, includes all 139 institutions within the dataset. Of the 139 schools represented in the dataset, there was an average value of estimated political ideology per all students per school of -0.267, a value slightly more moderate than the average student, a minimum value of -1.56 (Colby College), squarely on the far left, a maximum value of 0.922 (El Paso Community College), squarely conservative, and a standard deviation of 0.47. The distribution of average estimated political ideology per school, then, is slightly more moderate than the distribution of students (-0.267 compared to -0.337), and shows less deviation than the total dataset (0.47 compared to 1.17) but remains, on average, squarely moderate.

Figure 3: Distribution of Average Estimated Political Ideology by Institution



In addition to analyzing the entire distribution of institutions, I have also taken a closer look at the most polarized schools within the entire dataset, acknowledging that the dataset, again, skews towards the center but does include some data points that are more liberal or conservative, respectively.

The ten most politically liberal schools range from -1.56 at Colby College, a liberal arts college in Maine, to -0.911 at the University of California, San Diego. Only Colby College can be classified as far left; the rest of these schools land squarely in the liberal valuation categorically.

The ten most politically conservative schools range from 0.922 at El Paso Community College, a community college in El Paso, Texas, to 0.429 at Harrisburg Area Community College, another community college located in Harrisburg, Pennsylvania. The least conservative of these ten colleges fall among the moderate valuation categorically, while the remaining seven fall among the conservative valuation categorically, though none meet the threshold to be classified as far right institutions.

Given the relatively small number of students contributing to the average estimated political ideologies at the average school in the dataset (average n = 61, with a minimum of 1 student for some schools), some of these values may be more skewed than they would be at institutions with comparatively larger samples. What predicts an ideological skew in one direction or the other will be more thoroughly explored in the next chapter. In the next section, I explore online information and ideological diversity within the information networks of the students in the dataset.

Part III: Online Information and Ideological Diversity

After reviewing the data to answer the first research question, I considered the edge data I collected to answer the second and third research questions:

2) To what extent do the sources students follow on Twitter overlap ideologically?

a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?

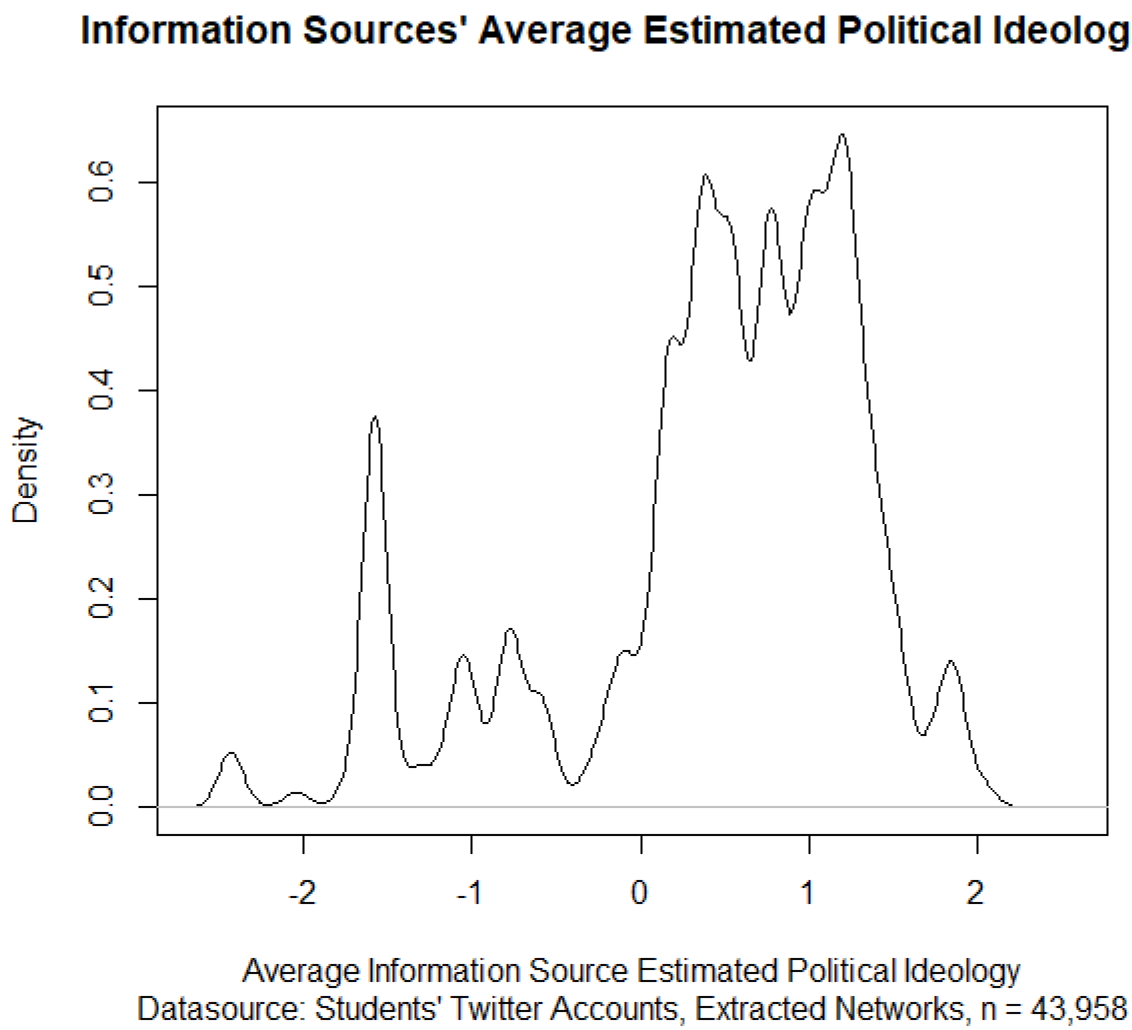
3) How ideologically diverse are students' information sources on Twitter?

To do this, I analyzed the edge data representing what sources and information students follow on Twitter and are thus exposed to and how the estimated political positions of those sources reflected the political positions of those students themselves. The following sections include the total distribution of all sources (edges) followed by students in the dataset, the overlayed distributions of the followed sources organized by the categorical subgroups ('far right,' etc.) I identified in the previous section, an examination of the most prominent sources in the dataset per subgroup, and an examination of the relationships between students' estimated political ideologies and the distributions of their own information sources (i.e., are extremely liberal students consuming extremely liberal information?), as well as an examination of the distribution of standard deviations of the sources each student follows, and the distribution of the arithmetic difference between students' sources and their own estimated political ideologies and the relationships between these variables. Each subsection serves, individually, to contribute to my understanding of ideological overlap among the sources students follow, as well as the potential skew and subsequent ideological diversity of sources present in the sample. Again, given the inextricable connection between contemporary expressions of political ideology and the information one consumes, investigation of students' information sources can provide more detail regarding their political ideology.

Estimated Source Political Ideology

To investigate the estimated political ideology of sources in the sample, which I calculated using the same tweetscores method described earlier in this chapter and more thoroughly in Chapter 4, I removed all students from the dataset who did not follow any sources. This resulted in a reduced analytic sample of 6,259 students who followed 43,958 sources. The sources ranged from -2.434, extremely liberal (*Mother Jones*), to 2.341 (*Dr. Turley Talks*), had a mean estimated political ideology of 0.4234, at the conservative edge of the middle, and a standard deviation of 0.926. Figure 4 depicts this distribution visually.

Figure 4: Distribution of Estimated Political Ideology for Sources



As the distribution reveals, the average is slightly to the right of center (0.4234) and significantly further to the right than the average estimated political ideology of the students in the dataset (-0.337), though it is fairly diverse with respect to the information sources students in the dataset follow. There is a significantly greater density of conservative sources, both continuously and categorically, and the average is drawn left primarily by one liberal source: the nearly ubiquitous 'paper of record' *The New York Times.* Other more liberal sources are significantly less followed than their more conservative peers in the media, at least by the students in this dataset, and there is a higher density of extremely conservative sources than extremely liberal ones.

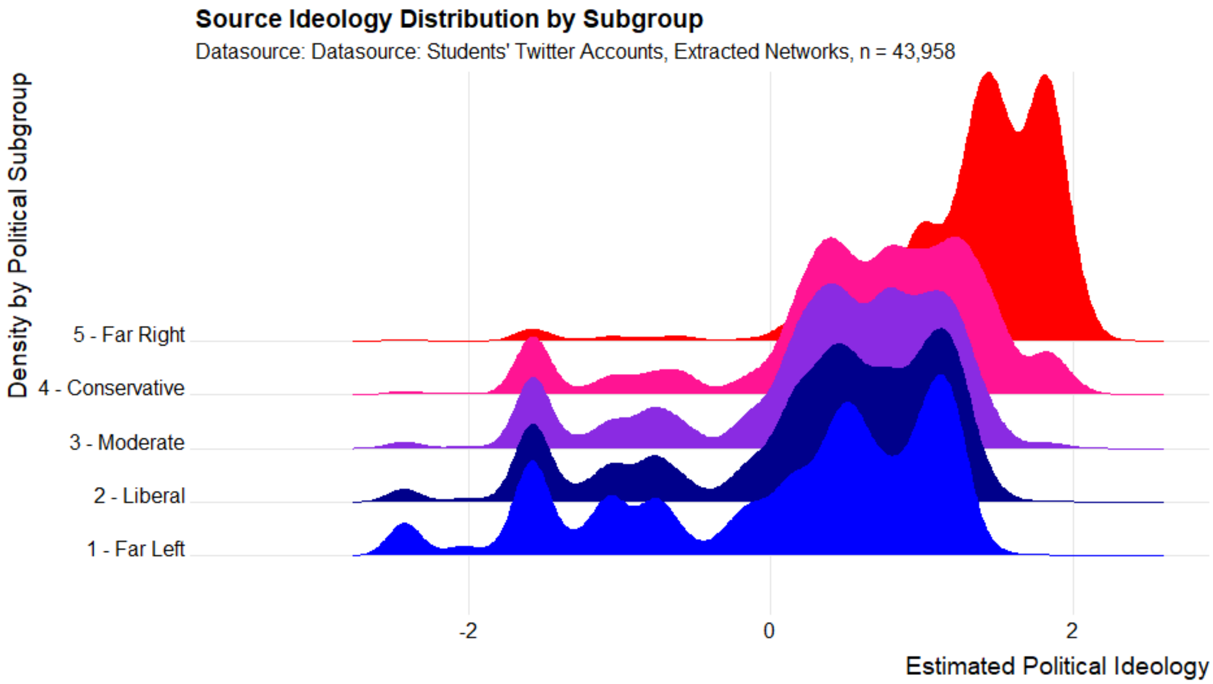
Theoretically, this distribution represents significant ideological overlap given the relatively moderate average of the estimated political ideology of sources followed in the dataset. However, understanding who follows these sources, and thus contributes to the densities represented in Figure 4, is difficult without separating them (i.e., reviewing distributions for staunchly liberal students and comparing them to staunchly conservative students). To make this comparison easier, in the next section, I present source distributions separated out into the five subgroups calculated and assigned earlier in this chapter.

Source Ideology Distribution by Categorical Subgroup

Figure 5 shows the distribution of sources followed by students separated into the five distinct subgroups listed earlier in this chapter. The subgroups are stacked vertically from Far Left at the bottom to Far right at the top and are recolored to reinforce the political subgroups (i.e., the far left subgroup is bright blue and the far right subgroup is bright red).

Figure 5 indicates that students like Berkeley Brian, East Coast Emily, and Just Josh have significantly more overlap with each other when it comes to the information sources they interact with online than they do with their more conservative peers like Southern Steve and Midwest Megan. Specifically, Southern Steve and Midwest Megan are far more likely to interact with sources that are isolated to their political subgroups and do not appear in the information networks of their more liberal peers.

Figure 5: Source Ideology Distribution by Subgroup



As Figure 5 shows, there is significant overlap among the various subgroups for students' estimated political ideologies. The most liberal students– the Far Left– show higher following densities for the most liberal of sources and share the fewest conservative sources with their peers. This trend continues across the subgroups, as liberals and moderates have roughly equivalent distributions as their more radically left peers. Conservatives and students on the far right, however, show steep drop offs with respect to more liberal sources and those comprising the Far Right subgroup follow nearly exclusively conservative sources. In sum, with respect to the second research question, there is significant overlap across most of the subgroups, though Far Right students appear to follow a significantly more polarized information base. To better understand these distributions, I now turn to reviewing the most prominent sources in the dataset per subgroup.

Most Prominent Sources in the Dataset Per Subgroup

To get a better idea of the information sources contributing to the distributions presented in the previous section and get a better understanding of ideological overlap, I have tabulated the top ten sources followed by students in each of the five categorical subgroups I have identified. Table 1 identifies these top ten lists for each subgroup.

As Table 1 indicates, there is significant overlap among the ideological subgroups with the greatest deviation, as indicated in Figure 5, occurring amongst student members of the Far Right. The predominant sources in each subgroup, save the Far Right, appear in the peer subgroups. For instance, *CNN* is in a top three spot in four of the subgroups but completely absent from the most radically conservative subgroup. The same can be said of the *Associated Press* and *The New York Times.* Other outlets, such as *BBC* and *BBC World* similarly appear in multiple subgroups. With the exception of *Fox News* and the *Ben Shapiro Show*, two conservative entertainment outlets, the most prominent outlets followed by members of the Far Right appear nowhere else in the top ten lists of every other subgroup. Here, closer examination of the most prominent sources per subgroup reveals that, while there is overlap, the clearest polarization and lowest degree of overlap is evident in the most radically conservative students' information networks. Similarly, the most prominent sources across all subgroups are fairly conservative: *CNN* (0.973), *BBC* (0.529), *Associated Press* (1.206), *NPR (*1.08), and *The Washington Post* (0.723), for instance. While some more progressive sources are prominent (*The New York Times*, -1.534; *Jacobin*, -2.422), the average source followed by students within the dataset remains a few decimal points shy of conservative, indicating that the majority of news the students in the dataset are consuming is right of center.

Table 1: The Most Prominent Sources in the Dataset Per Subgroup

	Far Left	**Liberal**	**Moderate**	**Conservative**	**Far Right**
1	CNN	BBC World	BBC	ABC News	Fox News
2	FiveThirtyEight	CNN	BBC World	BBC World	Louder with Crowder
3	Jacobin	CNN Breaking News	CNN	CNN	OAN Network
4	Lovett or Leave It	NPR	CNN Breaking News	CNN Breaking News	PragerU
5	NPR	Politico	NPR	Fox News	The Ben Shapiro Show
6	The Associated Press	The Associated Press	The Associated Press	The Associated Press	The Charlie Kirk Show
7	The Daily Show	The Daily Show	The New York Times	The Ben Shapiro Show	The Daily Wire
8	The New York Times	The New York Times	Time Magazine	The New York Times	The Joe Rogan Experience
9	The New Yorker	The New Yorker	Wall Street Journal	Wall Street Journal	The Michael Knowles Show
10	Washington Post	Washington Post	Washington Post	Washington Post	The Rubin Report

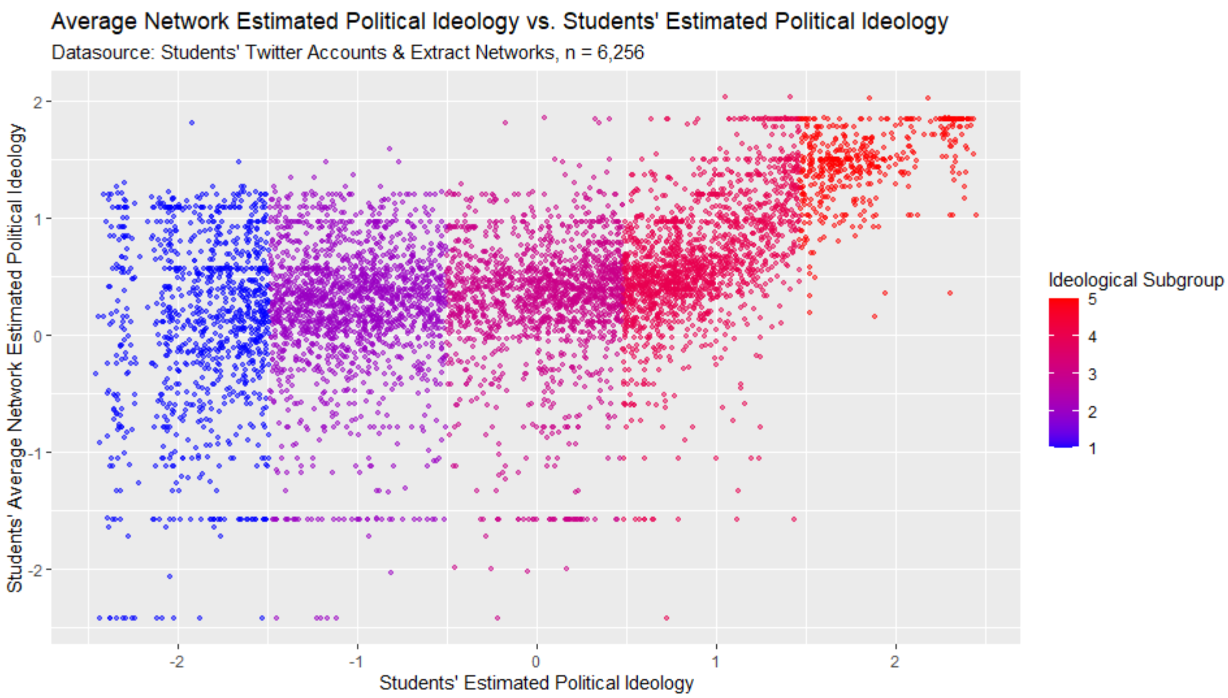
Keeping the center-right distribution of sources followed by students in the dataset in mind, and the theory of constrained choice online that I discussed in Chapter 3, I now consider how students' estimated political ideologies align with respect to the average estimated political ideologies of their information networks.

Students' Estimated Political Ideologies With Respect to Their Networks' EPIs

To investigate further how students' personal estimated political ideologies relate to the average estimated political ideologies of their information networks, and to better understand whether there is an identifiable ideological skew with respect to that information, I looked at the relationship between these two variables. Figure 6 describes this relationship between a student's estimated political ideology (EPI) and that of their information network (ANEPI). The figure is divided visually by color to depict the five categorical subgroups assessed earlier in this chapter for ease of interpretation. If we assume that students' information networks should mirror their own political ideologies, as described in Chapter 3, Figure 6 should show a clear and consistent positive correlation (i.e., students with very negative EPIs should have correspondingly negative ANEPIs). It does not show this correlation for the majority of the sample, though this positive correlation is clearly reflected as the students in question become more conservative. As Figure 6 shows, the bulk of students, regardless of their own estimated political ideology, consume information from networks that are either staunchly moderate or lean conservative. Conservative students, in contrast, consume exceedingly conservative information, which is consistent with the literature discussed in Chapter 3.

Using our example students, Figure 6 demonstrates the alignment between a student's estimated political ideology and the average estimated political ideology of sources in their network. Just Josh, for instance, had significantly more alignment between these two variables than his more liberal peers Berkeley Brian and East Coast Emily.

Figure 6: The Relationship Between Students' EPIs and Their Networks'



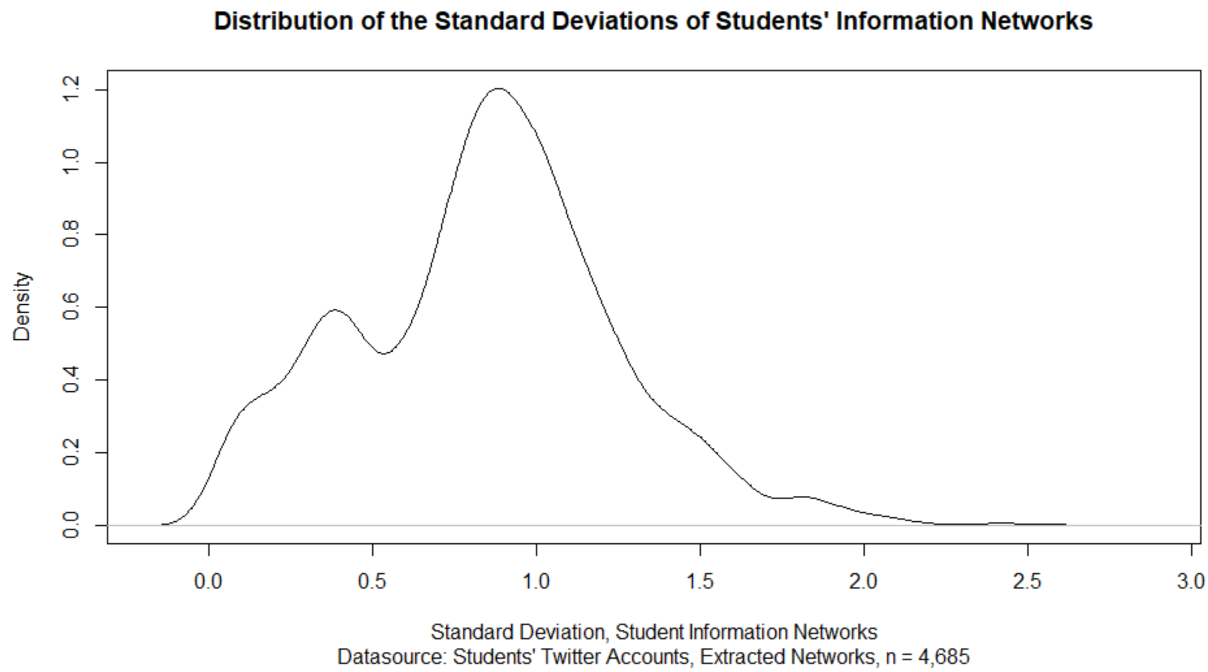
With respect to ideological skew, then, particularly in the liberal direction, the data presented here shows the opposite phenomenon: regardless of a student's own estimated political ideology, the news and information available online appears to be more moderate, or conservative, than student's own political positions until that political position is significantly more conservative than the average information available. Students with comparatively moderate or center-right political positions, then, are far more likely to consume information aligned with their beliefs than their liberal peers and students with political views significantly to the right of the average student are more likely to consume information that, while still conservative, is slightly closer to the center than their own position. With respect to the theory of constrained choice online presented in Chapter 3, it is likely that students are, in fact, presented with limited options when it comes to ideologically consonant news and that more conservative students might be seeking out this ideologically consonant information more frequently than their liberal peers. For instance, Midwest Megan may have had trouble interacting with news and information sources aligned with her far right views due to the lack of presence these sources may have on social media networks like Twitter. Far right organizations like *The Babylon Bee* have been regularly suspended from Twitter due to persistent violations of the Terms of Service and Community Guidelines and may simply be less available than outlets that comply with Twitter's standards.

To further explore the issue of skew, and ideological diversity, within the information students consume online, I analyzed the spread of students' information networks.

The Spread of Students' Information Networks (Standard Deviation)

As I have shown in the previous sections, students follow a wide range of information sources online that may match, or widely vary, their own estimated political ideologies. Understanding the average estimated political ideology of their networks, and the distribution of all sources followed by students in the dataset, however, does not clearly indicate how diverse each individual students' network is and thus, how diverse the information they are consuming is. To investigate ideological diversity, I evaluated the standard deviation of the estimated political ideology of the sources each individual student followed. Given the reality that many students followed only one source, and their standard deviations were thus 0, I removed some students from the analyses. Figure 7 represents the distribution of standard deviations across all students in the reduced analytic sample for these analyses (n = 4,685).

Figure 7: Distribution of the Spread of Students' Information Networks

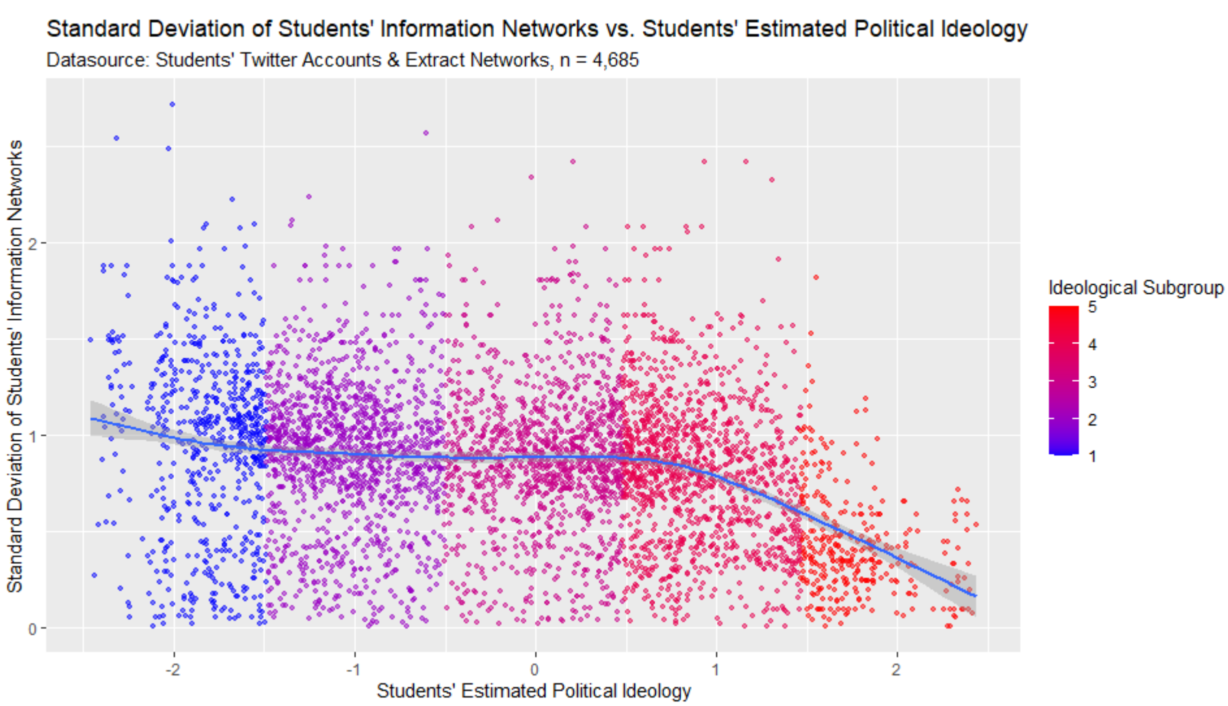


The standard deviation of students' information networks ranged from 0.0069, very little variation, to 2.71, extreme variation, with a mean of 0.847. As Figure 7 depicts, most students' information networks showed a good degree of variation with respect to the estimated political ideologies of the information sources they followed. While this implies that most students' information networks were, thus, ideologically diverse, I also plotted the standard deviation of each student's network estimated ideologies (SDANEPI) against their individual estimated political ideology (EPI) to assess whether some students' networks were more ideologically diverse than others. Again, I separated the five categorical subgroups by color for ease of interpretation and, on this figure, included a loess line indicating the trend, as it was less visible than the trend in Figure 6.

As Figure 8 shows, there is a relationship between a student's individual estimated political ideology and the standard deviation of the estimated political ideology of the sources in their network– here, a stand in for the ideological diversity of that network. The overall distribution shows a wide range of standard deviations at different student estimated political ideologies but, consistent with previous figures and earlier sections of this chapter, Figure 8 most prominently indicates that ideological diversity is fairly stable and evident across the sample for three of the five categorical subgroups: ‘Far Left’ students, ‘Liberal’ students, and ‘Moderate’ students. The other two subgroups, ‘Conservative’ students and ‘Far Right’ students show significantly less ideological diversity within their information networks. This is consistent with findings presented earlier on in this chapter, and is consistent with the theory and empirical work presented in Chapter 3. If there is a skew with respect to ideological diversity in the sample, it is more evident among conservative students, whose information networks are less ideologically diverse than their liberal peers.

With respect to the student examples presented earlier in this chapter and the most prominent outlets per subgroup presented, Figure 8 reiterates conservative ideological isolation.

Figure 8: The Relationship Between Students’ EPIs and Their SDANEPIs



Finally, I was interested in the relative distance between students' estimated political ideologies (EPIs) and the average estimated political ideologies of their networks (ANEPIs) as, pursuant to the theory described in Chapter 3 focusing on homophily, selective exposure, and the reduction of cognitive dissonance, I would expect student's political ideologies to be fairly close to the information they consume. To this end, I calculated the difference between these two variables (ANEPI - EPI = DEPI).

The Difference Between Students' EPIs and Their Networks (Arithmetic Difference)

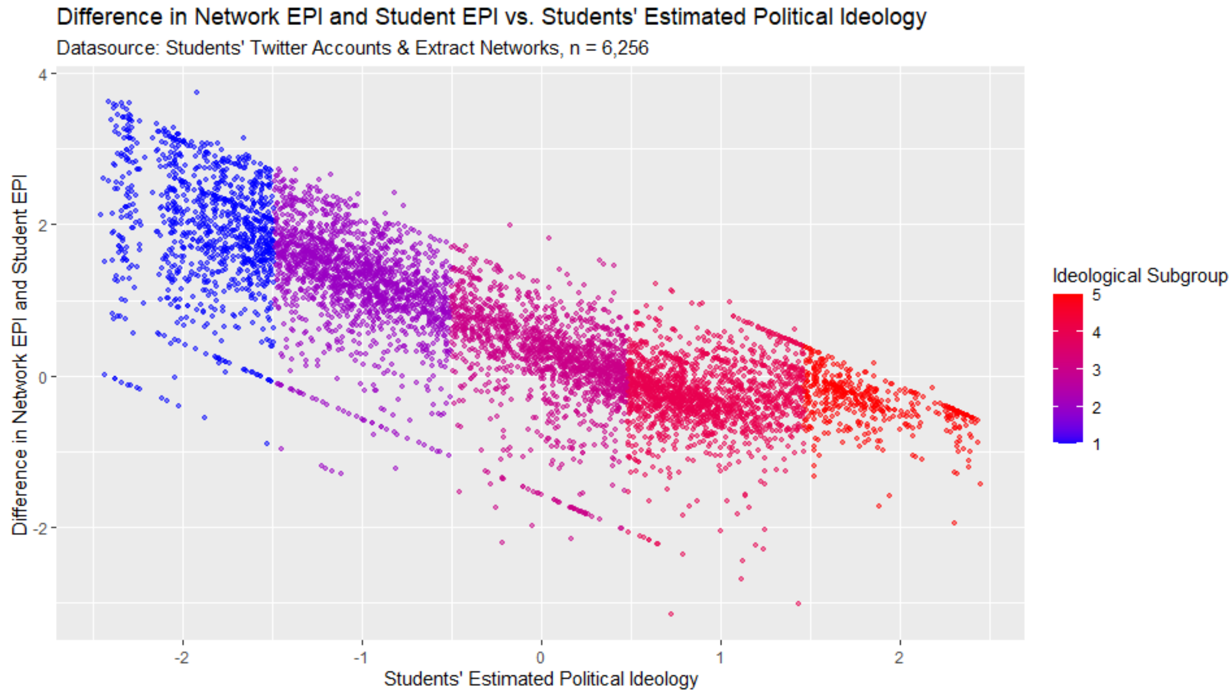
To calculate the difference between student's estimated political ideologies and the average estimated political ideologies of their information networks, I subtracted the first value from the second. The difference between students' estimated political ideologies and the average estimated political ideology of their information networks ranged from -3.14, representing a student with a network significantly more liberal than their estimated political ideology, and 3.738, representing a student with a significantly more conservative network than their estimated political ideology. The mean of the variable (DEPI) was 0.615, suggesting that the average student's information network was more moderate or conservative than their own position, a finding borne out in the findings discussed in previous sections of this chapter.

To more adequately interpret how this relationship is distributed among the various ideological subgroups in the dataset, I also plotted student's estimated political ideologies (EPI) against the difference between their average network estimated political ideology (DEPI) to ascertain whether any ideological subgroups were consuming information more aligned with their estimated political ideologies than other subgroups. The results, consistent with other figures which represent the students in the reduced analytic sample (here, n = 6,256), shows students from all subgroups separated by color (liberal to conservative indicated as a change from blue to red) and is depicted in Figure 9.

As Figure 9 indicates, there is a clear relationship between student's estimated political ideologies and the average estimated political ideologies of the information sources in their information network. Specifically, there is a clear negative relationship between student estimated political ideology and average network estimated political ideology, with more liberal students more likely to consume cognitively dissonant information than their more conservative peers, who are exceedingly more likely to consume cognitive consonant information online to a point. Students like Midwest Megan, whose political position is significantly to the right of the average information source in the dataset, may simply not have access to information that is perfectly ideologically aligned with their positions and may, by nature of source presence, consume information slightly more moderate than their own personal political position.

With respect to both ideological overlap and ideological diversity, a clear skew is evident within this data which suggests that conservative students exist in the echo chambers described in Chapter 3 far more frequently than their liberal counterparts, who are, consistent with the theory of constrained choice online I also described in Chapter 3, likely experiencing a diminished information ecosystem driven by profit incentives and sociopolitical regulation. I discuss this diminished information ecosystem more thoroughly in Chapter 7.

Figure 9: The Difference Between Students' EPIs and DEPIs



Part IV: Conclusion

This chapter was focused on answering three research questions: to what extent is the political ideology of students active on Twitter skewed towards liberalism? To what extent do the sources students follow on Twitter overlap ideologically? To what extent is the political ideology of those sources skewed towards liberalism? And how ideologically diverse are students' information sources on Twitter? In terms of practical significance, these research questions align to assess political ideology and ideological diversity in a more nuanced manner than would have been possible with a purely descriptive analysis of the ideological distribution. By adding the additional layers of analysis, the data presented in this chapter not only shows distinct and discrete ideological positions, but how those positions play out in terms of individual behavior.

To answer these questions, I manually built a dataset representing students across the country active on Twitter and calculated a variety of variables for each student based on latent attribute analysis of their political ideology online and the information they consume online. I also identified relevant institutional variables for each student, which are discussed in the next chapter. To answer the first research question, I analyzed students' estimated political ideologies as continuous variables, aggregated by institution, and assessed as recalculated categorical variables. I found that students were fairly evenly distributed across the political subgroups I identified and that students were more frequently estimated as liberals than they were as conservatives or moderates, but that previous, survey-based data on student political ideologies may overestimate moderates and underestimate conservative students and students in the far left and far right of the political spectrum.

To answer the second and third research questions, focused on ideological overlap and diversity of the information sources students interacted with online, I turned to edge data I collected from each student's individual Twitter profile. I analyzed the estimated political ideology of each source per student, analyzed the sources in the dataset in the aggregate, and explored the relationships between a variety of student-level variables (EPI, ANEPI, SDANEPI, DEPI) to ascertain whether students shared much ideological overlap when it came to the information they consumed and whether that information was ideologically diverse. I identified ideological diversity and overlap among most of the categorical ideological subgroups and indicated that, if an ideological skew was present in the data, it favored conservatives.

In sum, this study finds that, while there may be a slight liberal skew within the dataset, conservative and far right students are likely underestimated when it comes to presence on campus and are more ideologically siloed than their liberal peers. Similarly, the findings indicate that there is a great degree of ideological overlap with respect to the information consumed by students in the sample and that, if there is skew, it is more evident among conservative students, whose information networks are less ideologically diverse than their liberal peers. Finally, though the overall sample of information sources consumed by students in the sample is ideologically diverse, it is skewed slightly towards the right, as most of the sources within the sample fall to the right of center and extremely liberal students are far more likely to consume cognitively dissonant information than their conservative peers.

In the next chapter, I investigate whether there are any institutional features, such as selectivity, racial composition, academic outcomes, and financial considerations such as cost, influence the dependent variables I described in this chapter: student estimated political ideology and the average network estimated political ideology per student.

CHAPTER 6: FINDINGS (INFERENTIAL LINEAR MODELS)

The purpose of this chapter is to further present the findings of this study. This chapter specifically presents the findings of the inferential linear models which address research questions four and five, which I have included below. Again guided by the theory I described in Chapter 3 and the methodological approach I described in Chapter 4, as well as the descriptive results presented in Chapter 5, this chapter is organized into four parts:

Part I: Presentation of the Variables

Part II: Linear Regression Predicting Students' Estimated Ideology

Part III: Linear Regression Predicting Students' Information Network Estimated Ideology

Part IV: Conclusion of Inferential Findings

Guided by this study's research questions, abridged below to include only the questions relevant to this chapter, Part I describes the independent and dependent variables at the core of the analyses in this chapter and presents frequencies and distributions of each variable. Part II presents the findings of the linear regression model predicting students' estimated ideology, in which the dependent variable is individual students' estimated political ideologies as continuous values and the independent variables are the institution-level variables described in Part I. Part III presents the findings of the linear regression model predicting the average estimated political ideology of the information networks each student in the dataset has curated, in which the dependent variable is the average estimated political ideology of individual students' information networks and the independent variables are the institution-level variables described in Part I. Part IV closes the chapter with a summary of the findings presented in this chapter and an overview of what will be discussed in Chapter 7, the discussion of these findings. The research questions guiding this chapter are presented below:

4) What institution-level features predict the ideology of students on Twitter?

5) What institution-level features predict the ideological diversity of the information students are exposed to on Twitter?

Part I: Presentation of the Variables

Prior to the presentation of the inferential linear models, I present the variables included in both models, how they were coded, their minimum, maximum, and mean if relevant, and if I made any changes to the variables. I also include relevant correlations between the variables of interest and the dependent variables and describe key relationships that are explored more thoroughly in the regression models. First, I present the dependent variables.

Dependent Variables

The two dependent variables in question in this chapter are student's estimated political ideologies (EPIs), calculated using either maximum likelihood estimation or correspondence analysis as described in Chapters 4 and 5, and the average estimated political ideologies of student's information networks (ANEPIs), calculated as a mean of the estimated political ideologies of each source a student follows on Twitter. I presented the distributions of these variables in Chapter 5, but reiterate them here.

The first dependent variable, students' estimated political ideologies (EPIs), is available for all students in the dataset (n = 8,554) and ranges from -2.47, very liberal, to 2.44, very conservative, with an average value of -0.337, moderate but left of true center.

The second dependent variable, the average estimated political ideology of the sources in a student's individual network (ANEPIs), is only available for students in the dataset that followed sources on Twitter (n = 6,256) and ranges from -2.422, very liberal, to 2.031, very conservative, with an average value of 0.4297, moderate but right of true center.

Independent Variables

The independent variables included in the below models, and described in Chapter 4, include: the control of the institution, whether the institution was an HBCU or Tribal College, the size of the institution, the selectivity of the institution calculated as a percentage of students admitted versus students who applied, the total price of attendance for both in-state and out-of-state students, variables indicating the racial demographics of the institution (percent Black, percent White, etc.), a variable indicating the gender demographics of the institution (percent women) and completion metrics for 4, 6, and 8 years for first-time, full-time students. Below, I detail how each of the variables was coded within the analyses and describe the relative distributions of each variable within the dataset.

The control of the institution was included as a variable to ascertain if there was any difference in political ideology among students at public institutions versus private institutions. Of the 8,554 students in the sample 4,747 (55.5%) attended Public institutions and 3,807 (44.5%) attended Private institutions. The variable was coded as 0/1, Public/Private.

Whether an institution was an HBCU or a Tribal College was also included as a variable in the interest of determining if there was a difference in political ideology among students at predominantly white institutions (PWIs) and students at HBCUs or Tribal Colleges. Of the 8,554 students in the dataset, 8377 (97.9%) did not attend HBCUs and 177 (2.06%) did. No students in the dataset attended Tribal Colleges, so I removed this variable from the model.

The relative size of the institution was included as a variable to ascertain if there was any difference in political ideology among students at institutions varying in size. The institutions students attended were coded ordinally, with institutions ranging from 1,000-4,999 students coded as 1, institutions ranging from 5,000-9,999 coded as 2, institutions ranging from 10,000-

19,999 students coded as 3, and institutions with more than 20,000 students coded as 4. Institutions coded as 1, small institutions, represented 6.1% of all institutions attended. Institutions coded as 2, medium-sized institutions, represented 12.2% of all institutions attended. Institutions coded as 3, medium-to-large institutions, represented 31.8% of all institutions attended. Finally, institutions coded as 4, large institutions, represented 49.9% of all institutions attended. There were no continuous values for institution size.

The selectivity of the institution was included as a variable to ascertain if there was any difference in political ideology among students at institutions who accepted and enrolled comparatively fewer students than their peer institutions. Selectivity was coded as a continuous variable ranging from 0-100, with the value representing the total percentage of students who applied that were admitted to the institution. This value ranged from 4%, a highly selective institution (Stanford University), to 100% (Morehouse College), with an average value of 49.9%.

The total price of attendance for both in-state and out-of-state students was included as a variable to ascertain if there was any difference in political ideology among students at institutions charging comparatively more or less than their peers. Total price of attendance (in-state and out-of-state) was coded as a continuous variable. The total price of attendance for in-state students ranged from $11,850 (Hinds Community College) to $79,752 (Columbia University), with an average value of $46,955. The total price of attendance for out-of-state students ranged from $14,900 (Hinds Community College) to $79,752 (Columbia University), with an average value of $56,583.

Variables describing the racial demographics of each institution were included to ascertain whether the racial composition of the institution had any impact on the political ideologies of the students at that institution. I included racial composition variables for American

Indian and Alaskan Native students, Asian students, Black students, Latinx students, and Native Hawaiian and Pacific Islander students. A variable for white students was not included, as white students were the control group in the IPEDS data. Each of these variables was coded continuously as a percentage of total enrollment at that institution.

The racial demographic variable for American Indian and Alaskan Native students ranges from 0% enrollment to 13%, with an average enrollment of 0.247%. The racial demographic variable for Asian students ranges from 1% enrollment to 37%, with an average enrollment of 10.15%. The racial demographic variable for Black students ranges from 0% enrollment to 93%, with an average enrollment of 8.475%. The racial demographic variable for Latinx students ranges from 0% enrollment to 84%, with an average enrollment of 12.25%. The racial demographic variable for Native Hawaiian or Pacific Islander students ranges from 0% enrollment to 3%, with an average enrollment of 0.0347%. Finally, the gender demographic variable representing the percentage of women enrolled at the institution ranges from 0% to 100%, with an average enrollment of 54.05%.

Lastly, I have included variables reflecting each institution's completion metrics for four years, six years, and eight years, in the interest of ascertaining whether there is a difference in political ideology among students at institutions with varying degrees of academic output. These variables were coded as continuous percentages, indicating the number of students who received an academic degree versus the number of students enrolled. The variable for four year award conferral ranged from 6% to 100% with an average value of 56.46%. The variable for six year award conferral ranged from 17% to 100%, with an average value of 73.18%. Finally, the variable for eight year award conferral ranged from 17% to 100%, with an average value of 74.93%.

Correlations Between the Dependent and Independent Variables

In addition to reporting the variable distributions, I also ran correlations between each variable. For ease of interpretation and visualization, I have separated the variables into subgroups. The first subgroup includes student-specific network variables, such as the dependent variables student estimated political ideology (EPI) and the average estimated political ideology of sources in a student's information network (ANEPI), and independent variables such as the standard deviation or spread within that network (SDANEPI), and the difference between that student's estimated political ideology and their network's average estimated political ideology (DEPI). The relationships between these variables are shown in Figure 10.

Figure 10: Variable Correlation Graph: Calculated Student Variables

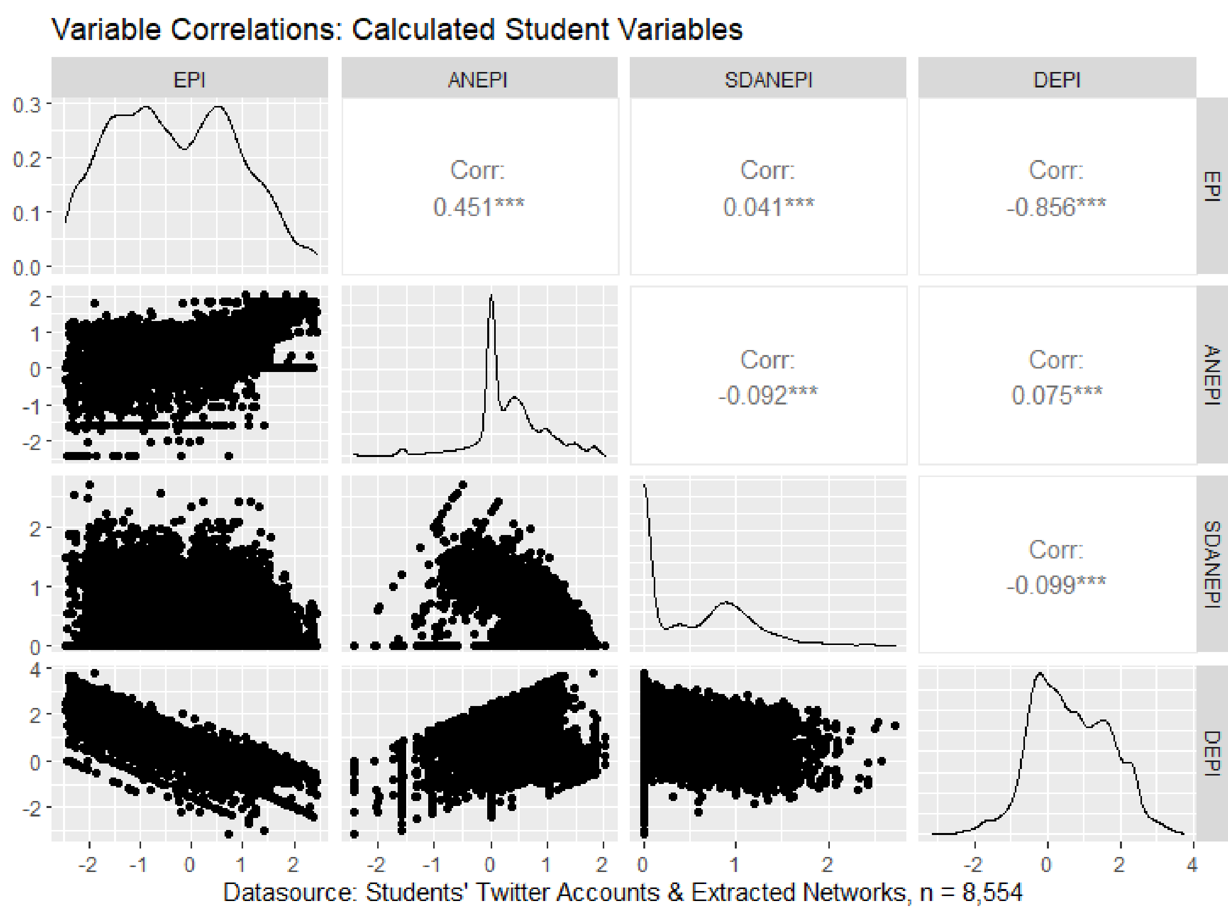


Figure 10 contains three main pieces of information for each variable: 1) that variable's distribution (the main diagonal), 2) correlation coefficients, which include markers of statistical significance, for each other variable, and 3) scatterplots which visually depict the relationship identified in data point 2. Figure 10 contains data representing each student in the overall dataset (n = 8,554), and the ranges for each variable are presented on the associated section of the X and Y axes. These ranges are consistent with the data presented in Chapter 5, and include a range of -3:3 for EPI, -3:3 for ANEPI, 0:2 for SDANEPI, and -3:5 for DEPI.

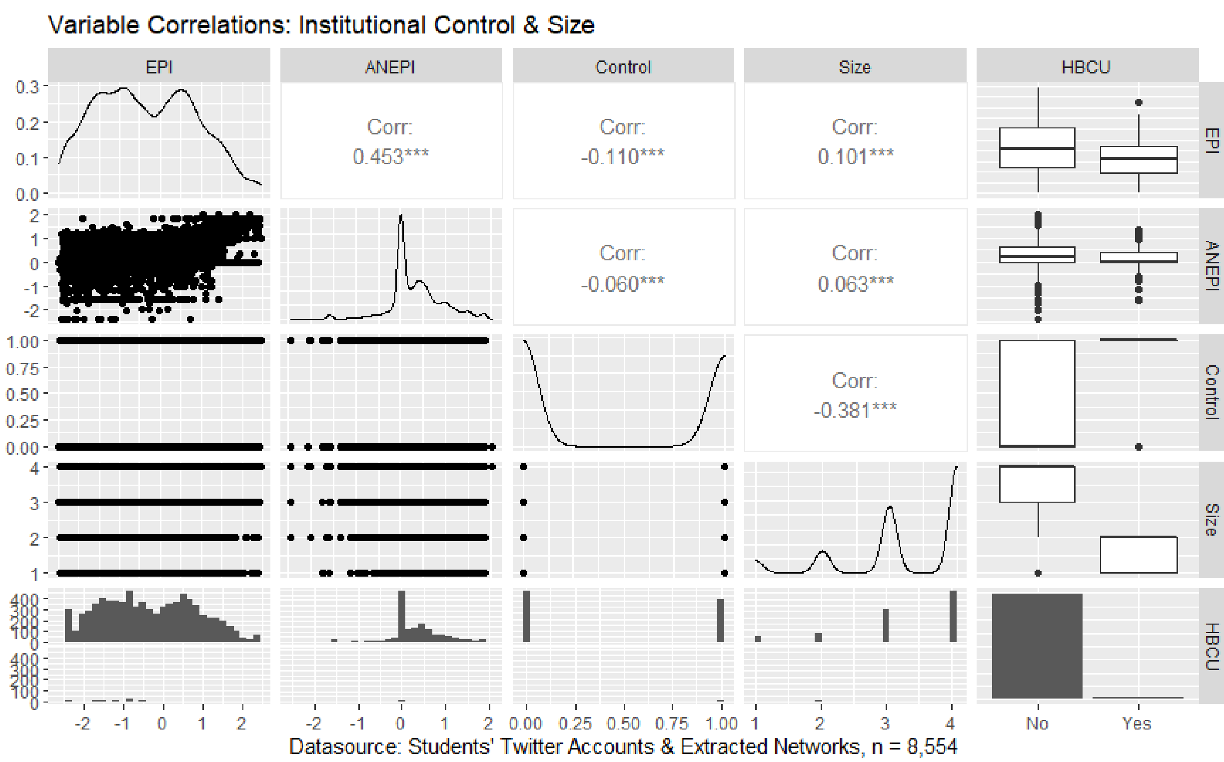
Figure 10 reiterates many of the findings presented in Chapter 5, specifically that there are clear and significant correlations between the dependent variable of interest (student estimated political ideology), and the independent variables of interest calculated from students' information network data (ANEPI, DEPI, SDANEPI). The relationship between student estimated political ideology and the average estimated political ideology of their information network is positive (0.451, $p < 0.001$). This is consistent with the finding presented in Chapter 5 indicating that students closer to the center or center right of the political spectrum are consuming information more aligned with their own political beliefs. The correlation between student estimated political ideology and the spread of a student's information network is much weaker, though significant, but indicates evidence of broad ideological diversity across the dataset that is curtailed as students get more radically conservative. Finally, the relationship between a student's estimated political ideology and the difference between that ideology and the average estimated political ideology of their information network is strong and negative (-0.856, $p < 0.001$), reiterating the finding presented in Chapter 5 that demonstrates that more liberal students are far less likely to be consuming ideologically consonant information than their more conservative peers and that students at the center and center right of the political spectrum are likely consuming the most ideologically consonant information.

With respect to the dependent variable in the second linear model for this study, average network estimated political ideology (ANEPI), there are weak but significant correlations with the spread and difference variables (SDANEPI and DEPI). The scatter plot representing the relationship between the average network estimated political ideology and the standard deviation of that network does not indicate a clearly linear relationship. The relationship between the average network estimated political ideology and the difference between a student's estimated political ideology and that of their network, however, reveals a slight positive correlation (0.075, $p < 0.001$). This correlation supports previous findings that more moderate and conservative students are likely consuming ideologically consonant information, but that is not the case for extreme cases like Midwest Megan, who simply was too conservative for much of the information available within the dataset.

The second variable group I considered included the dependent variables and variables focused on institutional control and size. After filtering the overall dataset to only include student observations with these variables, the sample size did not change (n = 8,554). These relationships are visualized in Figure 11. Figure 11 again reports three main data points for each variable: 1) the variable's distribution, 2) its correlation coefficients with the other variables in the group and whether these coefficients were statistically significant, and 3) a visualization of the correlation. With respect to the relationship between a student's estimated political ideology and the control (Public or Private) of their institution, there is a small but significant negative correlation (-.107, $p < 0.001$). There was a similarly small but positive correlation between institutional size and a student's estimated political ideology (.103, $p < 0.001$). Finally, students at HBCUs had a significantly more negative estimated political ideology (-0.85) than their peers not at HBCUs (-0.326; t = 7.32, $p < 0.001$).

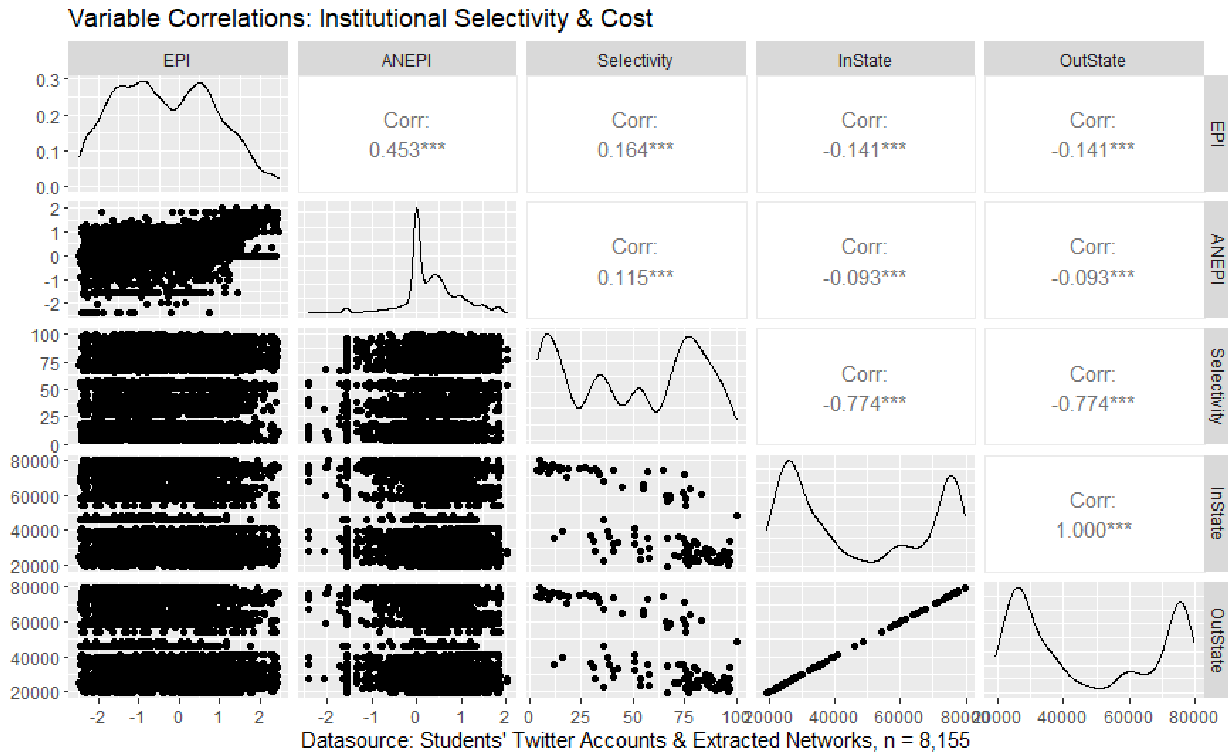
With respect to the average estimated political ideology of a student's information network (ANEPI), there was a very small negative but significant correlation with the control of a student's institution (-0.059, $p < 0.001$) and a very small but significant positive correlation with the size of a student's institution (0.061, $p < 0.001$). Similar to the relationship between a student's estimated political ideology and whether they attended an HBCU, students attending HBCUs had significantly more progressive average network estimated political ideologies (0.20) than their peers at non-HBCUs (0.32; t = 3.316, $p < 0.001$).

Figure 11: Variable Correlation Graph: Institutional Control and Size



The third variable group I considered included the selectivity and cost variables per institution (admission rate, cost of attendance in-state, cost of attendance out-of-state). After filtering the overall dataset to only include student observations with these variables, the sample size was slightly reduced (n = 8,155). These relationships are visualized in Figure 12. Figure 12 again reports three main data points for each variable: 1) the variable's distribution, 2) its correlation coefficients with the other variables in the group and whether these coefficients were statistically significant, and 3) a visualization of the correlation.

Figure 12: Variable Correlation Graph: Institutional Selectivity & Cost



With respect to the relationship between a student's estimated political ideology and the selectivity of their institution, there was a small but significant positive correlation (0.164, $p < 0.001$). There were similarly small but significant identical correlations between in-state and out-of-state cost of attendance and a student's estimated political ideology (-0.141, $p < 0.001$).

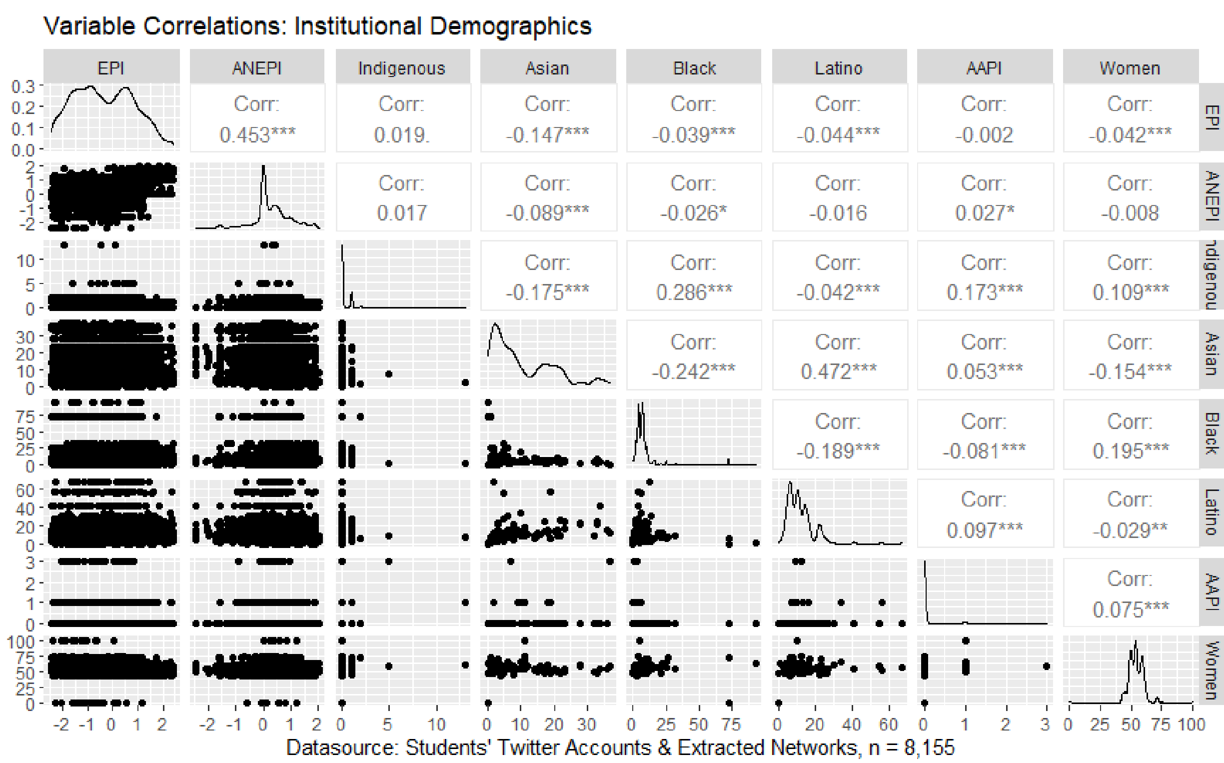
With respect to the average estimated political ideology of a student's information network (ANEPI), there was a small but significant correlation with institutional selectivity (0.115, $p < 0.001$). There were also small but significant identical correlations between ANEPI and the in-state and out-of-state cost of attendance (-0.093, $p < 0.001$). Finally, there was an unsurprising and significant negative correlation between institutional selectivity and cost of attendance (-0.774, $p < 0.001$), indicating that schools that cost more to attend are generally also more selective.

The fourth variable group I considered included the racial demographic variables per institution (indigenous enrollment, Asian enrollment, Black enrollment, Latino enrollment, AAPI enrollment, and women's enrollment)). After filtering the overall dataset to only include student observations with these variables, there was no change in sample size. These relationships are visualized in Figure 13. Figure 13 again reports three main data points for each variable: 1) the variable's distribution, 2) its correlation coefficients with the other variables in the group and whether these coefficients were statistically significant, and 3) a visualization of the correlation.

With respect to the relationship between a student's estimated political ideology and the racial demographics of their institution, there were several significant correlations. First, there was no significant correlation between the enrollment percentage of indigenous students and students' estimated political ideologies. There were small to exceedingly small but significant correlations between other demographic group's enrollment percentages and students' estimated political ideologies. The correlation between Asian student enrollment and estimated political ideology was negative (-0.147, $p < 0.001$), as was the correlation between Black student enrollment and estimated political ideology (-0.039, $p < 0.001$), the correlation between Latino student enrollment and estimated political ideology (-0.044, $p < 0.001$), the correlation between AAPI student enrollment and estimated political ideology (-0.002, $p < 0.001$), and the correlation between women's enrollment and estimated political ideology (-0.042, $p < 0.001$). For most of the demographic groups included in this model, increased enrollment was related with more liberal political attitudes at their institutions.

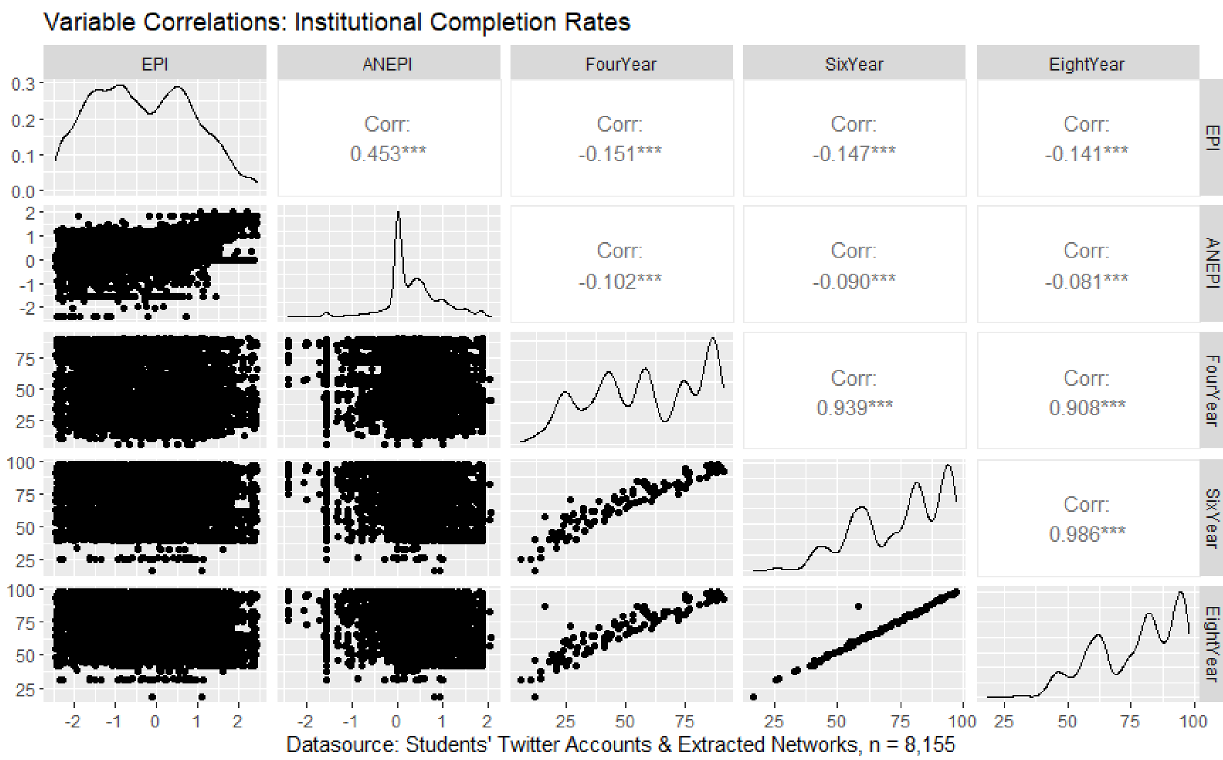
With respect to the average estimated political ideology of students' information networks, there were several significant correlations with institutional demographic variables. There was not a significant correlation between indigenous student enrollment and average network estimated political ideology or Latino student enrollment and average network estimated political ideology. There were small negative correlations between average network estimated political ideology and Asian student enrollment (-0.089, $p < 0.001$) and Black student enrollment (-0.026, $p < 0.05$). There was a small positive correlation between AAPI student enrollment and average network estimated political ideology (0.027, $p < 0.05$). Finally, there was no significant correlation between women's enrollment and average network estimated political ideology.

Figure 13: Variable Correlation Graph: Institutional Demographics



The fifth and final variable group I considered included the dependent variables and variables for institutional completion rates after four, six, and eight years. After filtering the overall dataset to only include student observations with these variables, the sample size was again unchanged. These relationships are visualized in Figure 14. Figure 14 again reports three main data points for each variable: 1) the variable's distribution, 2) its correlation coefficients with the other variables in the group and whether these coefficients were statistically significant, and 3) a visualization of the correlation.

Figure 14: Variable Correlation Graph: Institutional Completion Rates



With respect to student estimated political ideology, there were significant negative correlations between the dependent variable and the percentage of students who completed a degree in four years (-0.151, $p < 0.001$), the percentage of students who completed a degree in six years (-0.147, $p < 0.001$), and the percentage of students who completed a degree in eight years (-0.141, $p < 0.001$). The most notable trend among these variables is that institutions who graduated students more quickly were slightly more likely to have more liberal student bodies.

With respect to average network estimated political ideology, there were similarly negative significant correlations with the four year completion rate (-0.102, $p < 0.001$), the six year completion rate (-0.090, $p < 0.001$), and the eight year completion rate (-0.081, $p < 0.001$). Again, students at schools with quicker completion rates curated slightly more progressive information networks.

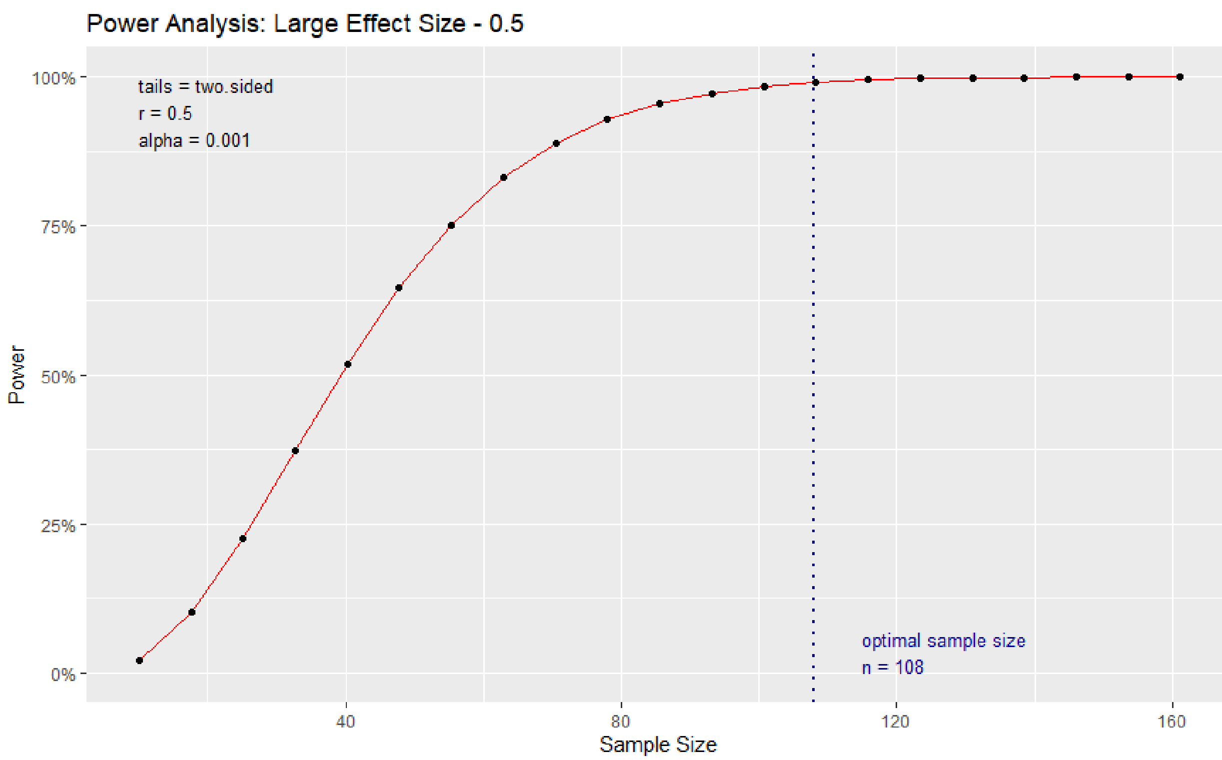
With the relationships between the main dependent and independent model variables explored, I now turn to power analyses of the data. After exploring what sample sizes were necessary to observe potential effects in the data, I discuss how the final analytic datasets for both regression models were filtered and set.

Power Analysis and Filtering of the Dataset

In the interest of identifying the maximum number and size of effects within the data analyses described in this chapter, I conducted several power analyses. Power analyses allow a researcher to investigate the sample size required to observe a specific effect size at a designated significance level and at a designated power level. To identify the requisite sample sizes for my analyses at the most conservative level of significance and highest degree of power, I set my significance level for each analysis to $\alpha = 0.001$ and I set my power level to 99%. Finally, I used the correlational data identified in the previous section to estimate potential effect sizes within the larger regression models. Given the wide range of correlation coefficients, I conservatively estimated the sample sizes necessary for effect sizes of 0.5 (a large effect size), 0.3 (a medium effect size), and 0.1 (a small effect size) to ensure my dataset was large enough to observe each effect size. These effect size examples were consistent with Cohen's predictions for effect sizes and are consistent with other effect sizes found in similar work to this study (Havey, 2023).

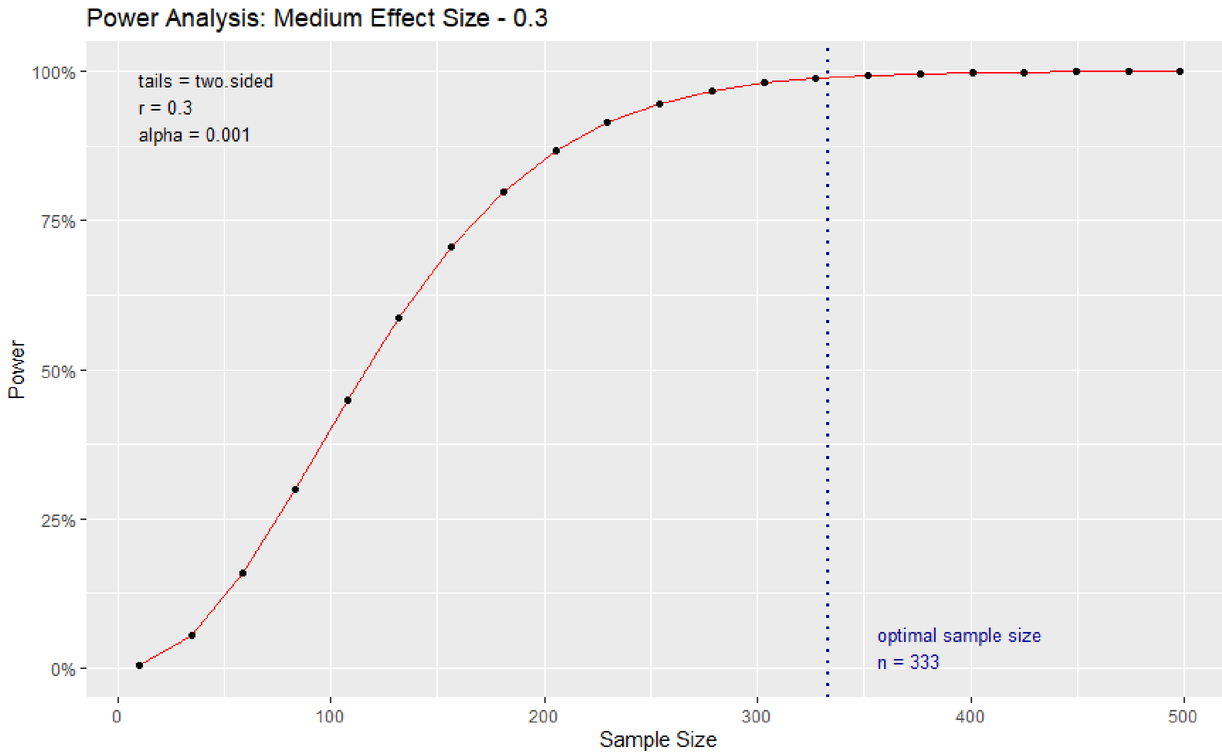
The first power analysis, identifying the sample necessary to observe a large (0.5) effect size at a significance level of 0.001 and a power level of 99%, indicated a necessary sample size of 108. This analysis is visually depicted in Figure 15.

Figure 15: Power Analysis: Large Effect Size - 0.5



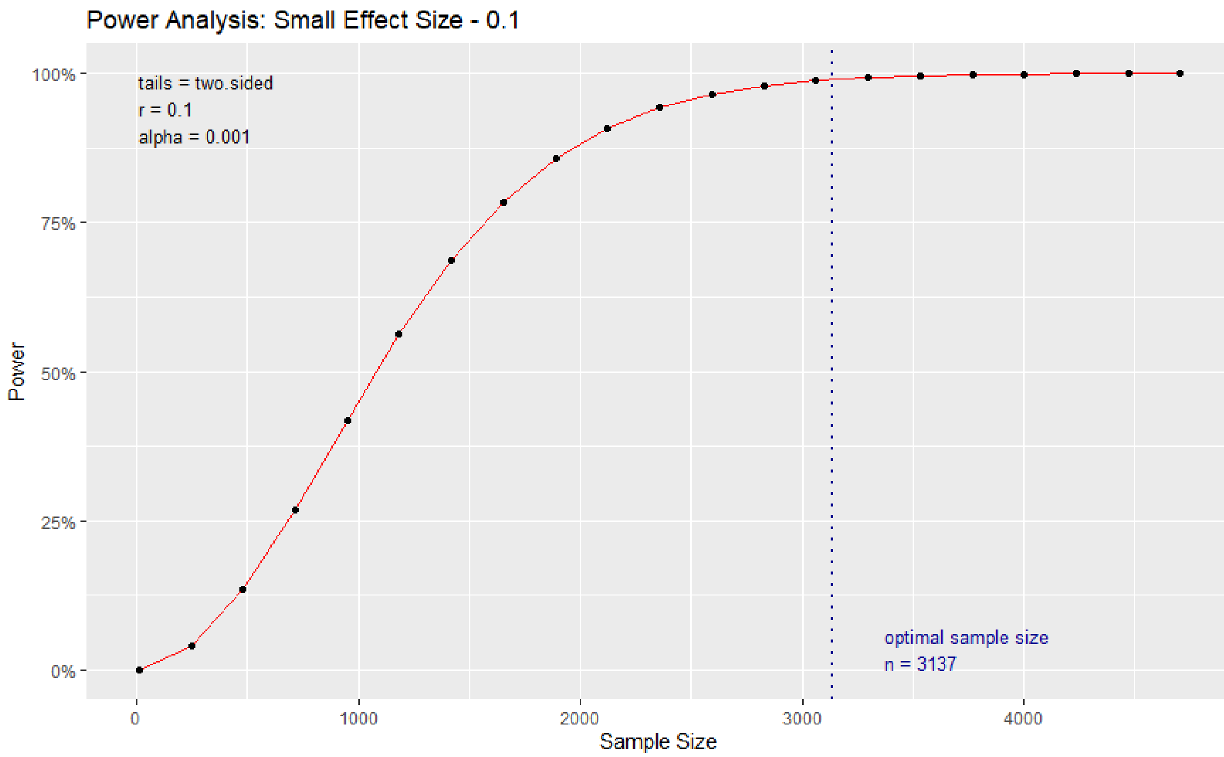
The second power analysis, identifying the sample necessary to observe a medium (0.3) effect size at a significance level of 0.001 and a power level of 99%, indicated a necessary sample size of 333. This analysis is visually depicted in Figure 16.

Figure 16: Power Analysis: Medium Effect Size - 0.3



Finally, the third power analysis, identifying the sample necessary to observe a small (0.1) effect size at a significance level of 0.001 and a power level of 99%, indicated a necessary sample size of 3,137. This analysis is visually depicted in Figure 17.

Figure 17: Power Analysis: Small Effect Size - 0.1



Given the sample sizes projected by these power analyses, all of which are dwarfed by my entire sample (n = 8,554), I filtered the dataset to remove any student observations who had missing data for any of the independent variables described above.

For the second analysis, the linear regression focused on what institutional variables predict the average network estimated political ideology for each student in the sample. To attend to this, I had to reduce the analytic sample to reflect the reality that many students in the full dataset did not follow any sources on Twitter. This resulted in a 4.6% loss in sample for the first analytic sample focused on estimated political ideology (n = 8,155), and a 26.6% loss in sample for the second analytic sample for the regression model focused on average network political ideology (n = 5,982). Both of these samples remained well above the threshold identified for all three power analyses.

The results of the inferential model predicting the average estimated political ideology of students' information networks are presented in Part III of this chapter. In the next section, I present the results of the first regression analysis focused on institutional predictors of student estimated political ideology.

Part II: Inferential Linear Model - Students' Estimated Political Ideologies

After selecting my analytic sample for this model, I ran a single, multivariate linear regression model to see what institutional variables predicted students' political ideologies. I entered all of the previously mentioned independent variables together into the model and assessed the residual standard error and adjusted R-squared to determine the fit of the model. While the median residual was small (-0.07) and the overall residuals symmetrical, the residual standard error of the model was 1.148, indicating a sizable distance between observed values and expected values within the model (notable, given the -2.5:2.5 range of the data). The adjusted R-squared was 0.049, exceedingly low for a linear model and indicative of a poor fit with respect to the variables included. Given previous research that indicates student political ideology is more likely tied to individual traits and college behaviors rather than institutional and demographic features (Havey, 2023; Havey & Schalewski, 2022), this is not surprising. The explanatory power of the model, thus, is not particularly compelling, but it does serve to confirm and complement previous findings using novel data. That being said, several of the variables still presented significant effects that should be considered. The results of the linear regression model are presented below in Table 2.

As Table 2 indicates, there were several statistically significant variables predicting the estimated political ideology of students in the dataset. While many of the variables had statistically significant correlations with the dependent variable of interest, after controlling for all other variables, these effects became insignificant or weaker. I have included all of the variables included in the model in the table, and discuss all variables per block below.

First, the institutional control variable (whether a school was public or private), had an insignificant effect on students' estimated political ideologies (-0.247, $p > 0.05$), consistent with previous research that has indicated a wide spread of political diversity at multiple institutional types (Binder & Wood, 2014; Havey, 2020a). Put plainly, institutional type does not seem to have a dramatic effect on students' political behavior. Whether a student attended an HBCU, however, did have a significant positive effect on students' estimated political ideologies (0.049, $p < 0.001$), with students at HBCUs holding slightly more positive, or closer to the center, political ideologies than their peers at other institutions. This contradicts correlational data that suggests students at HBCUs were more progressive than their peers attending other institutions, indicating that controlling for all other variables influenced this relationship.

Institutional size had a similarly moderating effect, as there was a significant positive effect with respect to the size of a student's institution. Specifically, as an institution got larger (within the subgroups), students became slightly more moderate or closer to the center (0.139, $p < 0.001$). Institutional selectivity had a similarly significant positive effect, though it was drastically smaller, with students at more selective institutions being marginally more moderate or closer to the center than students at less selective institutions (0.0022, $p < 0.05$).

In-state and out-of-state cost of attendance had negligible and insignificant effects on students' estimated political ideologies. Functionally, the cost of a school seems to have no bearing on students' estimated political ideologies, suggesting that, while other work has identified that socioeconomic status or personal wealth may be linked to more conservative political ideologies for students (Havey, 2023; Havey & Schalewski, 2022), cost of attendance may not be an adequate or appropriate proxy measure for socioeconomic status. Similarly, other individual-level variables may be responsible for potential variance driven by cost of attendance. That being said, there was a strong statistical relationship between selectivity and cost within this study, suggesting there may be more to explore with respect to socioeconomic status and student political behavior.

With respect to demographic variables, such as the percentage enrollment of historically excluded racial minority groups like indigenous students, there were few significant effects. First, the percentage of indigenous enrollment, small across the sample, did not have a significant effect on students' estimated political ideologies. This may simply be a product of the incredibly small proportion of indigenous students non-Tribal colleges and universities enroll. There was a small but significant negative effect between Asian student enrollment and students' estimated political ideologies (-0.016, $p < 0.001$), with greater Asian enrollment at an institution leading to more progressive student political ideologies. There were no significant effects on students' estimated political ideologies with respect to Black student enrollment or Latino student enrollment, though previous work has identified that the perception of a more diverse or, frankly, less white student population can lead to political polarization for white students, this study did not share that finding. Similarly, there was no significant effect with respect to Asian American and Pacific Islander students. Finally, there was a small but significant negative effect with respect to the percentage of women enrolled at an institution (-0.0083, $p < 0.001$). As the percentage of women enrolled at an institution increased, the estimated political ideologies of its students got marginally more progressive.

Finally, with respect to four, six, and eight year completion rates, there were no significant effects on student political ideology. Given recent work that has identified that college has seemingly little, if any, effect on student's political ideologies, specifically within their first college year (Havey & Schalewski, 2022) and over the course of their four college years (Havey, 2023), it is not shocking that the completion rates have minimal impact on student politics. The reality is that most students enter and exit college with the same political orientation. In the next section of this chapter, I consider how these variables, and students' estimated political ideologies, affect the average estimated political ideologies of their information networks.

Table 2: Predicting Students' Estimated Political Ideologies

Linear Regression: Predicting Students' Estimated Political Ideologies

Independent Variable	*Effect*	
Institutional Control	-0.247	
HBCU Status	0.049	***
Institutional Size	0.131	***
Institutional Selectivity	0.0022	*
In-State Cost of Attendance	2E-06	
Out-of-State Cost of Attendance	2E-06	
Indigenous Enrollment	0.047	
Asian Enrollment	-0.016	***
Black Enrollment	0.0044	
Latino Enrollment	0.00015	
AAPI Enrollment	-0.053	
Women's Enrollment	-0.0083	***
Four Year Award Rate	-0.0017	
Six Year Award Rate	-0.0091	
Eight Year Award Rate	0.0078	

*'***'= $p < 0.001$, '**' = $p < 0.01$, '*' = $p < 0.05$, $n = 8{,}155$*

Part III: Inferential Linear Model - Online Information and Ideological Diversity

After reducing the analytic sample for this model, I ran a single, multivariate linear regression model to see what institutional variables, as well as students' estimated political ideologies, predicted the average estimated political ideology of each student's network. The reduced analytic sample for this model was 5,982. I entered all of the previously mentioned independent variables, as well as the previous dependent variable, students' estimated political ideologies, together into the model and assessed the residual standard error and adjusted R-squared to determine the fit of the model. The median residual for this model was exceedingly small (-0.00706) and the minimum and maximum values were nearly symmetrical. The residual standard error of this model was 0.581, a sizable but more moderate distance between observed values and expected values than in the previous model, which is not unreasonable given the range of the data (again, -2.5:2.5). The adjusted R-squared was also improved for this model at 0.2105, still comparatively low for a linear model but indicative of better fit than the first model presented in this chapter. If the dependent variable from the previous analysis is removed, the R-squared value drops to 0.02, suggesting that much of the predictive power is due to students' estimated political ideologies, a logical finding given the relationship between the two variables reported in Chapter 5 and earlier in this chapter. Again, given previous research that indicates that student political ideology and the performance of that political ideology is more likely tied to individual traits and college behaviors, rather than institutional and demographic variables (Havey, 2023; Havey & Schalewski, 2022), the explanatory power of this model is not wholly surprising. Finally, like in the previous model, many of the significant correlations identified earlier in this chapter dissipate when other variables are controlled, suggesting weak effects and an overall limited institutional influence on student politics. The results of this linear regression model are presented below in Table 3.

As Table 3 indicates, there were several statistically significant variables predicting the average estimated political ideology of student's information networks. I have included all of the variables included in the model in the table and discuss them below.

The first statistically significant predictive variable was students' estimated political ideology, the dependent variable in the previous regression. As Chapter 5 demonstrated, there is a clear relationship between a student's political position and the average estimated political ideology of their network. Here, there is a statistically significant effect showing that, as students' estimated political ideologies become more positive, or move towards the more conservative end of the political spectrum, the average estimated political ideologies of their networks simultaneously move to the right (0.231, $p < 0.001$). This finding is consistent with the findings in Chapter 5, which presented this relationship without other variables, and the correlational analyses presented earlier in this chapter.

With respect to institutional variables, there were several statistically significant predictive variables. There was no significant effect between institutional control and the average estimated political ideology of a student's information network, reiterating the finding from the first model in this chapter that suggested institutional control has a limited influence on students' political behaviors. Whether a student attended an HBCU similarly had no significant effect on that student's information network, which is unsurprising given the relatively weak effect HBCU attendance had on students' estimated political ideology and the strong effect students' estimated political ideology had on the average estimated political ideology of their network. Similarly, though institutional size did appear to drive some variance in students' estimated political ideologies, there was no effect on the average estimated political ideologies of their information networks, again suggesting that a students' own estimated political ideology is the most important driver of the political positions of the information networks they curate. This is consistent with the theory presented in Chapter 3 and previous findings suggesting institutional variables have limited, if any, influence on student political behavior (Havey, 2023; Havey & Schalewski, 2022). Institutional selectivity did, however, have a small but significant positive effect on the average estimated political ideology of students' information networks (0.00182, $p < 0.001$). While individual socioeconomic status was not measured within this study, using selectivity as a proxy for the socioeconomic distribution of a school (i.e., schools that are considered more prestigious with lower acceptance rates, such as Stanford, also tend to be more expensive than their more inclusive and accessible peers, such as community colleges; indeed, within this study there was a strong positive correlation between cost of attendance and institutional selectivity (0.65, $p < 0.001$)) that may suggest that socioeconomic status drives both students' estimated political ideologies and the estimated political ideologies of their information networks. While investigating that relationship is beyond the scope of this study, a positive finding would be in line with previous research that indicates socioeconomic status as a driver of political ideology (Havey, 2023; Havey & Schalewski, 2022).

With respect to the actual variables representing cost of attendance for in-state and out-of-state students, there were no significant effects on average network estimated political ideology. The exceedingly small but insignificant positive effect between the independent variables representing cost of attendance and the dependent variable in this model, as well as the clear relationship between cost of attendance and institutional selectivity, may, however, provide additional credence to the possible relationship between a student's socioeconomic status and their political ideology within this study and identified within previous research (Havey, 2023; Havey & Schalewski, 2022).

With respect to variables representing the racial and gender composition of each institution, there were no significant statistical effects between the enrollment percentages of different racial groups, such as Black students, Asian students, and Latino students, and the average estimated political ideology of students' individual information networks. There was, however, a significant but relatively small positive effect between the percentage of women enrolled and students' average network estimated political ideology (0.00269, $p < 0.05$). While the previous model indicated that a greater percentage of women enrolled in an institution generally predicted a more progressive student body, this finding suggests an opposite effect: that the percentage of women enrolled in an institution may moderate the average estimated political ideology of students' information networks. This effect is, however, quite small, and may simply be a product of the reality that most students' information networks were more moderate or conservative than their behavior suggested they were. I discuss this possibility of a constrained information ecosystem more thoroughly in Chapter 7.

Finally, with respect to the relationships between the variables representing institutional completion rates at four, six, and eight years, there was one statistically significant effect on average network estimated political ideology. There were no significant effects between four year and six year completion rates and the average estimated political ideology of students' networks, but there was a significant but small positive effect between the dependent variable of interest and eight year completion rates (0.00728, $p < 0.05$). Given the relative size of this effect and the limited relationships between the other completion rates and the dependent variables, and the relationship between selectivity and eight year completion rates (schools with higher eight year completion rates were comparably more selective), this finding is not wholly surprising and may be an interesting path to pursue in future research.

Table 3: Predicting Average Network Estimated Political Ideologies

Linear Regression: Predicting Average Network Estimated Political Ideologies

Independent Variable	*Effect*	
Student Estimated Political Ideology	0.231	***
Institutional Control	0.0356	
HBCU Status	0.116	
Institutional Size	0.0152	
Institutional Selectivity	0.00182	***
In-State Cost of Attendance	0.0000006	
Out-of-State Cost of Attendance	0.0000006	
Indigenous Enrollment	0.008	
Asian Enrollment	-0.00097	
Black Enrollment	-0.00142	
Latino Enrollment	0.00057	
AAPI Enrollment	0.0211	
Women's Enrollment	0.00269	*
Four Year Award Rate	-0.00157	
Six Year Award Rate	-0.00353	
Eight Year Award Rate	0.00728	*

‘***’= $p < 0.001$, ‘**’ = $p < 0.01$, ‘*’ = $p < 0.05$, $n = 5{,}982$

Part IV: Conclusion

This chapter was focused on answering two research questions: what are the predictors of students’ estimated political ideologies and what are the predictors of the average estimated political ideologies of students’ information networks? The two models further contribute to the phenomena under investigation in this study and, while the actual explanatory power of these models is comparably low, the models provide nuance regarding some of the relationships demonstrated in Chapter 5 and in the correlation analyses earlier in this chapter.

With respect to the institutional predictors of students’ estimated political ideology, it is unsurprising that institutional variables such as selectivity and racial composition were not wildly predictive of students' estimated political ideologies. Given previous research that indicates that students' political positions are more often the result of behaviors and experiences than institutional features (Havey, 2023; Havey & Schalewski, 2022), the lack of strong relationships in this model is reasonable. That being said, there were some interesting and statistically significant relationships predictive of students' estimated political ideology that square with previous research, as well as the theory that guided this study, that I will discuss later and more thoroughly in Chapter 7.

With respect to the average estimated political ideologies of students' information networks, a stronger model than the model describing students' estimated political ideologies, the relationship between students' estimated political ideologies and the estimated political ideologies of their information networks is perhaps the clearest and aligned with the theory presented in Chapter 3. This relationship is also clearly demonstrated in Chapter 5. Outside of that relationship, the other institutional variables were only marginally predictive of the average estimated political ideology of students' information networks, though present.

In the next chapter, I discuss the findings of Chapters 5 and 6 in more detail and grounded in the empirical research that came before this study. I briefly outline several key themes that can be drawn from these findings and discuss the research questions in order before discussing them in concert with each other. Specifically, I discuss how the findings presented in the previous two chapters speak to three main points: 1) the reality that surveys are not entirely trustworthy measures of student political behavior, 2) that polarization exists across the entirety of the political spectrum but that there is clustering and increased siloing in particular segments of that spectrum, and 3) that the nature of the internet has constrained choice and access with respect to information and that this constrained information ecosystem has driven polarization. Finally, I offer some implications of these findings and suggestions for future work.

CHAPTER 7: DISCUSSION & IMPLICATIONS

This study was driven by the consistent criticisms levied against institutions of higher education, specifically that they lack ideological diversity, are liberally skewed and thus inherently liberalizing and biased and that that is counterproductive to the purposes of higher education. This narrative is evident across the history of higher education in the United States of America, with any threat to the status quo, such as the fights for coeducation (Thelin, 2011), integration (Biondi, 2012), and divestment (Soule, 1997), characterized as attacks on the standards of a theoretically merit-based system.

Early champions of this critical position, such as William F. Buckley, as well as contemporary standard bearers of conservative political though like Dinesh d’Souza (1991), David Horowitz (2009), and Ben Shapiro (2010), have quickly and often positioned diversity at the core of this criticism. Arguing that colleges and universities are watering down educational standards for the illiberal purposes of diversity, equity, and inclusion (d’Souza, 1991), that schools are sites of indoctrination that inhibit academic freedom and are a threat to democracy (Horowitz, 2009), or that the liberal bias they perceive is so strong as to suggest brainwashing (Shapiro, 2010), conservative critiques of higher education have been persistent and cloaked in strategically defensible arguments about merit, standards, and representation.

These critiques are not without impact, as public higher education in the United States of America has faced persistent disinvestment and restructuring, often driven by partisan policymaking (Cantwell & Taylor, 2020), the whims of radical partisan lawmakers (Dar, 2012; Dar & Lee, 2014), and white racial resentment (Taylor et al., 2020). Reshaping higher education, specifically to be in line with dated and regressive political policies is, and has been, the goal (Trachtenberg, 2019).

As detailed in Chapter 2 of this study, these criticisms have not been unaddressed within higher education research. Numerous studies have focused on the accused liberal skew of institutions of higher education, explored whether the college experience itself is politicizing, and described the social and political demographics of this nation's colleges and universities. This research has been almost as persistent as the criticisms that have seemingly driven it, but it has also been stagnant with respect to the data examined and the methodological approach used to examine it. This study was designed to enter into conversation with the extant research literature on student political ideologies, behaviors, and the potential ideological skew of institutions of higher education and extend that conversation using a contemporary and novel data source– digital trace data– and a methodological approach designed to limit the biases and issues inherent to survey research described in Chapters 2 and 4 of this study. This study was guided by the following research questions:

1) To what extent is the political ideology of students active on Twitter skewed towards liberalism?
2) To what extent do the sources students follow on Twitter overlap ideologically?
a) To what extent is the political ideology of the sources students follow on Twitter skewed towards liberalism?
3) How ideologically diverse are students' information sources on Twitter?
4) What institution-level features predict the ideology of students on Twitter?
5) What institution-level features predict the ideological diversity of the information students are exposed to on Twitter?

Using a multistage and multisite quantitative approach to collecting and analyzing the data, I created a large dataset representing 8,554 students across 43 states and 139 institutions of higher education. The rigorous manual production of this dataset ensured that all students identified matched study criteria and that, through latent attribute analysis of students' social media data extracted from Twitter, the calculated variables of interest reflected students' actual observed behavior, rather than the individual perception and interpretation of that behavior evident in survey research. As described in Chapters 2, 4, and 5, this is a major strength of the data collected and analyzed in this study and extends previous survey-driven work.

This strength is mutually reflected and demonstrated in the guiding theories of this study, described in Chapter 3. With respect to the latent attribute analysis that undergirded the calculation of the variable at the heart of this study– student's estimated political ideologies– and the subsequent interpretation of that variable through the student profiles presented in Chapter 5, it is clear that homophily and selective exposure play a role in the curation of students' information networks and their subsequent performances of political ideology.

As I described in Chapter 3, humans are driven by a variety of social, emotional, and cognitive factors that push them down one of two routes: thinking fast or thinking slow (Kahneman, 2011). Both routes exemplify the theory of information foraging Pirolli (2005, 2007) developed, and suggest that information consumers are generally focused on maximizing output while minimizing input. Membership in particular social groups, such as the ones outlined by political ideology in this country, influence individuals' inputs, leading them to ask reflective questions like "who am I in conversation with?", "who do I want to be in conversation with?", and "what information do I actively want to avoid?" that drive their choices on and offline. In theory, concepts such as homophily and selective exposure explain such behavior. In general, people seek out members of their in-group (homophily) and work to limit exposure to outgroups or information that may be dissonant with their reality (selective exposure).

In practice, this may look like Berkeley Brian only associating with other liberal Berkeley students online, or Midwest Megan sticking to socializing exclusively with the other women in her sorority, who may have been selected or self-selected into homophilous houses on the basis on socioeconomic status, ethnic identity, or grades (Armstrong & Hamilton, 2013; Nguyen et al., 2020). People, like the students that comprise my dataset, make choices based on emotions and ease, a reality reflected in this study.

The rest of this chapter explores these choices, responds to the research questions posed at the start of this study and reiterated throughout, and identifies the implications of this work. As I discussed at the end of Chapter 6, the findings of this study can be grouped into three main ideas: 1) that surveys are not entirely trustworthy measures of student political behavior, 2) that polarization exists across the entirety of the political spectrum but that there is a clustering and increased siloing in particular segments of that spectrum, and 3) that the nature of the internet has constrained choice and access with respect to information and that this constrained information ecosystem has driven polarization. I discuss each of these points individually alongside the research questions that drove this study and follow that discussion with implications for future research, practice, and policy.

On Data, Its Limitations, and Its Ability to Impact Research, Policy, and Practice

As I described in Chapter 2, all of the previous research conducted on student political ideologies and whether college is inherently politicizing has been informed by survey data (Astin, 1977, 1993; Bailey & Williams, 2016; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Pascarella & Terenzini, 1991, 2005; Schiff, 1993; Woessner & Kelly-

Woessner, YEAR). In fact, the majority of this research (Astin, 1977, 1993; Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022; Schiff, 1993) has been informed by the exact same survey data, collected, managed, and made available by the Higher Education Research Institute at the University of California, Los Angeles.

These studies have presented inconsistent findings, with some researchers identifying a consistent liberalizing influence of college on students (Astin, 1977, 1993; Elchardus & Spruyt, 2009; Hanson et al., 2012; Pascarella et al., 2012; Pascarella & Terenzini, 1991, 2005), and others identifying multidirectional and moderating effects (Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022). The effects of college on students' politics have been, where identified, relatively weak and driven by student interactions rather than institutional features. For instance, it is unsurprising that students socialized into a particular field, such as sociology or political science, might reflect the sociopolitical contours of that field (Elchardus & Spruyt, 2009) or that students who choose to attend a liberal arts college are themselves more liberal on average than their peers at other institutional types (Hanson et al., 2012). They are also dated and, as Dey (1996, 1997) recommended in his own studies of student politics, regular assessment is necessary to reflect the changing political tides that are inherent to our national culture.

These studies are also limited by the biases inherent to survey-based research described in Chapters 2 and 4 of this study, namely that research relying on self-reported data can be compromised and limited by self-selection, response bias, and interpretation bias (Sax et al., 2003). Compounding these issues, studies that have focused on whether college students can even accurately assess and self-report their own political ideologies have indicated that self-identification (as liberal, for instance) and articulated support for actual policies (such as affirmative action or abortion) are infrequently aligned within student populations (Bailey &

Williams, 2016; Woessner & Kelly-Woessner, 2020). In short, students are not the best assessors of their own political positions and survey research relying on student self-report data is limited in terms of both accuracy and its ability to inform research, practice, and policy. This study attends to that reality and presents data, and findings, that are driven by the everyday behavior of students and is informed by their actions, not what they choose to report.

Do Campuses Lack Ideology Diversity?

The short answer to this question, which mirrors the first research question of this study, is no. This study found minimal support for claims that campuses are skewed towards liberalism and lack ideological diversity. As I described in Chapter 5, the average estimated political ideology of students in this nationally representative dataset (n = 8,554) was -0.337, slightly left of center but squarely within the moderate designation calculated within this study. The data in this study present a smaller middle (i.e., fewer moderates) than in previous, dated studies (Astin, 1977, 1993; Dey, 1996, 1997) and consistent with more recent survey-based studies of student politics (Havey, 2023; Havey & Schalewski, 2022). There is broad ideological representation across all institutional types and schools themselves are almost unilaterally moderate. This finding is supported by the regression analyses reported in Chapter 6, which identify institutional size as a moderating influence on students' estimated political ideologies, describe that increased minority enrollment has a very modest liberalizing effect, and demonstrate that most of the variables contributing to across-institution variation (i.e., student body demographics, cost of attendance, selectivity) have negligible if any effects on student politics. Overall, while the descriptive power of the regression analyses in Chapter 6 is comparably low for linear models, this is mostly a product of the inclusion of institution-level variables to predict student-level outcomes. With respect to this study, this may suggest a propensity for students to gravitate towards more ideologically consonant institutions (Hanson et al., 2012). Given the presence of ideological diversity evident across the institutions in this dataset, these findings may also indicate that the institution itself is not particularly relevant to student political ideologies. This is consistent with past research which identifies student choices and interactions as the driving forces behind political and ideological changes (Dey, 1996, 1997; Havey, 2023; Havey & Schalewski, 2022).

The smaller middle identified in this study also suggests that previous research may overestimate liberal student populations and underestimate conservative ones, artificially producing liberal skew where it does not necessarily exist. There is less variance with respect to student political ideology in recent, survey-based work (SD = 0.87, Havey, 2023) than within this study (SD = 1.17), and the data within this study present a potentially more accurate spread of student political ideologies. Given the potential for interpretation bias on surveys (Sax et al., 2003), and past work on students' political identities and their support for policies (Bailey & Williams, 2016; Woessner & Kelly-Woessner, 2020), it is possible that many students who identified themselves on surveys as moderate or liberal are not indeed moderate or liberal and may instead more accurately find themselves reflected somewhere else on the political spectrum.

One of the core strengths of this study is that the data analyzed is, effectively, presented in context. The student profiles presented in Chapter 5 are an example of this contextualization, as each profile describes what each students' estimated political ideology means in practice and allows for a more nuanced interpretation of what a -1.5 (liberal) means. The data collected via survey that drives other research on student political ideology fails to present as clear of a picture of each respondent in context. They may, for instance, identify as moderate but, in practice (measured in other studies through support of actual policies; Bailey & Williams, 2016;

Woessner & Kelly-Woessner, 2020) more clearly align with conservatives. Flattening that context and simply reporting that a self-reported two on a scale of one to five is a two, regardless of whether they may actually be closer to a three or a four, is a weakness of past studies that is accounted for in this one. This refinement of past work is evident in the greater diversity of student estimated political ideologies reported in this study.

Does the Information Students' Consume Lack Ideological Diversity?

One of the arguments at the core of criticisms of institutions of higher education as lacking ideological diversity is that students exist with echo chambers and campus bubbles. Within these bubbles, critics argue, students only interact with peers that they already agree with and are not exposed to a diversity of viewpoints and are instead indoctrinated into liberal groupthink (Horowitz, 2009; Shapiro, 2010). As I described in Chapter 3, there is a great degree of homophily and selective exposure online with respect to studies of the general population. This study predominantly focused on evaluating the ideological diversity on college campuses but, as described in Chapters 1, 2, and 3, evaluating that ideological diversity would not be complete without an investigation into the online information behaviors of the students in question. To explore the ideological diversity of the sources students follow on Twitter, guided by the concepts of homophily and selective exposure, and to assess whether there was a particular ideological skew within those sources, I extracted the 43,958 sources the 8,554 students in the dataset followed on Twitter and analyzed them with respect to their political positions and the political positions of the students following them.

With respect to the first two research questions, focused on the potential political skew of students and their sources, the findings of this study indicate broad representation across the political spectrum. Specifically, information sources of all political stripes are represented within the dataset and the average political position of these sources is center right politically. There is also significant overlap among the students in the dataset, with students from the far left, liberal, moderate, and conservative categories sharing many of the same information sources (visualized in Figures 5 and 6). The ideological diversity of individual students' information networks, as described by standard deviation and average network estimated political ideology, was strong, though the variance in the information networks of students on the far right of the political spectrum was nearly negligible, indicating their networks are not ideologically diverse. This was not true of their more liberal peers, who were more likely to consume information that was ideologically dissonant (i.e., Berkeley Brian, a member of the far left, still consumed information classified as conservative) than their more conservative peers (the alignment between Just Josh and Southern Steve's political positions and the information they consumed was much closer than that of Berkeley Brian and East Coast Emily). The regressions presented in Chapter 6 further substantiate this finding, as conservative students consume conservative media at a far greater rate than their liberal peers. Further, there are information sources followed by members of the far right that do not appear in the information networks of any other students, indicating that, if one end of the political spectrum is particularly siloed, polarized, and engaging in selective exposure, it is conservative students.

This finding is consistent with previous studies of the general public on Twitter, which have identified greater clustering and homophily among conservative users (Barberá, 2015; Colleoni et al., 2014; Himelboim et al., 2013; Stern et al., 2014), greater selective curation of information networks (Weeks et al., 2019; Pearson et al., 2018), and greater polarization with respect to information sources (i.e., news that is more extreme) among conservatives (Price & Kaufhold, 2019).

Overall, even though the majority of students in the dataset leaned left, there were significantly more conservative information sources in the edge dataset and less ideological alignment between a students' political position and the political position of the average source in their information network for students on the left than students on the right. As I described in Chapter 1 and theorized at the end of Chapter 3, this may be the result of a constrained information ecosystem online. I explore my theory of constrained choice online, and how the findings of this study support it, in the following section.

Is The Online Information Ecosystem Constrained?

As I described in Chapter 3, the information available online is not merely the product of chance or even popularity: it is a carefully curated result of heavy moderation, business choices, and amplification of particular perspectives. Platforms like Twitter are subject to intensive content moderation (Roberts, 2016, 2017), an incredibly opaque practice that is often intentionally obfuscated in an effort to preserve corporate profits (Roberts, 2018). As Munger (2020) has described, social media platforms like Twitter act as amplifiers for news outlets and information sources and, with the rise of clickbait media and the profit incentive that drives it, what is presented to users on these platforms cannot be taken at face value as the most useful, accurate, or desirable information available. As Noble (2018) has so thoroughly documented, it is often the information that is most profitable, regardless of its veracity, relevance, or potential for harm that appears on and is amplified by these platforms

We know that social media is increasingly the source of information for most people (Gottfried & Shearer, 2016; Shearer, 2016), and that information consumption on social media platforms like Twitter is exceedingly partisan (Price & Kaufhold, 2019; Licari, 2020). Given these realities, I explored how students' curated their information networks on Twitter.

The average student in this study was just left of center. The average information source was just right of center, nearly one standard deviation more conservative than the average student. With respect to the information in students' information networks, and the potential skew of that information, most students were exposed to and consumed sources more moderate than their own political positions. For instance, students like Midwest Megan, a student on the far right who was exceedingly more conservative than the majority of students in the dataset, the information available online is simply not radical enough to match her political position. This may be due to the reality that radical, far right news and entertainment outlets regularly violate the Twitter terms of service and are promptly banned from the platform, but it is important to note. Similarly, a student like Berkeley Brian, at the other end of the political spectrum from Midwest Megan, has limited but moderate access to information that aligns with his political beliefs but is significantly more likely to be exposed to and consume information that is more than two standard deviations to the right of his own politics. Extreme political positions exist within the dataset with respect to both students' estimated political ideologies and the estimated political ideologies of the information sources they interact with, but the fact of the matter is that they are simply not the predominant sources available. Moderate, center right sources are. Across the political spectrum, students are either unable, or unwilling, to access information that is completely aligned with their positions unless they are squarely moderate.

Put plainly, even if students wanted to exist explicitly in echo chambers online, they would be hard pressed to do so unless they identified as moderates. However, given the worse alignment between political position and the politics of information sources for left-leaning students, and the lack of ideological diversity with respect to information for right-leaning students, the constrained information ecosystem online may be disproportionately impacting students at the extremes and may be leading to the discrepancies between survey data and the data collected in this study. Specifically, a lack of accurately aligned information may cause students to interpret themselves as more politically extreme, or moderate, than they are.

Summary of Discussion

This study utilized student-level Twitter data to assess students' individual ideological positions, the ideological diversity of their institutions and the field of higher education writ large, the ideological diversity of the information they are exposed to and consume online, and to investigate whether skew existed. Using a multistage, multisite quantitative approach and latent attribute analysis of students' digital trace data, I analyzed the estimated political ideologies and information networks of 8,554 students following 43,958 sources across 43 states and 139 institutions. The major findings of this study were: 1) that surveys may not be the most accurate way to measure student political ideology and behavior, 2) that campuses do not lack ideological diversity, and 3) that, where extreme polarization and siloing exists, it is predominantly localized to conservative students. I now shift to discussing the implications of these findings.

Implications

The findings of this study present significant implications for research, practice, and policy, and the study itself represents an innovative data source and approach to understanding student and institution level variation with respect to political behavior. In the sections that follow, I present considerations for research, practice, and policy, respectively.

Considerations for Research

One of the key considerations for research from this study should be to consider new approaches to measuring student behavior and evaluating it in context. As I have detailed throughout this dissertation, the majority of extant research on student political behavior and the political influences of college on students has been localized to survey research and, with minimal deviation, been localized to survey data collected by one institution: The Higher Education Research Institute at the University of California, Los Angeles. There are, of course, many benefits to survey research but, as I have demonstrated within the last few chapters, there are also limitations. One of these limitations is that survey research relies on self-report data, which is subject to a variety of selection, response, and interpretation biases. One of the strengths of this study is thus methodological: the data collected within this study rely on latent attribute analysis of digital trace data and represent students' actual behavior rather than their interpretation and perception of that behavior. While research using digital trace data and Twitter data in particular is not particularly novel or new in other academic fields (Kwak et al., 2010; Steinert-Threlkeld, 2018), education research using Twitter data in particular has been limited to text analyses of cultural phenomena (Black in the Ivory, for instance, and other topics that have been broadly discussed online). Future research on student political behaviors, and on student behavior in general, could benefit from the incorporation of digital trace data.

Another strength of this study and implication of its findings is the comparative data that it provides researchers. Specifically, the data collected and analyzed in this study could be used as a contemporary and nationally representative baseline of students' political ideologies and may serve as comparison data, or even be included in models, for future studies of campus politics focused on students, faculty, and administrators. For instance, institution-level data at the aggregate could be used as a comparison and independent variable for analyses of faculty-level digital trace data describing political ideologies and social organization online.

Finally, one of the core findings of this study was that, where ideological skew and a lack of ideological diversity exists, it appears to be more concentrated among conservative students.

Given the relatively little research documenting the behavior, and experiences, of conservative students on college campuses (Binder & Kidder, 2022; Binder & Wood, 2014; Havey 2020a, 2021, 2023; Havey & Schalewski, 2022), it is my hope that this study prompts renewed, and increased, interest in conservative students and their experiences in higher education.

Considerations for Practice

At the heart of this study is my concern for higher education and the negative effects of partisan policy making, often driven by critiques of institutions of higher education as lacking ideological diversity, and how those policies impact students, staff, and faculty. That being said, there are also clear practice implications from this study.

First, this study provides nationally representative data for higher education institutions in the United States of America and point estimates of students' estimated political ideologies that, when presented in the aggregate, can be used as institution variables (i.e., a school can be assigned an average political ideology, as well as a standard deviation and additional information about the contours of the student population that comprise that average). These institution-level data points could be used by concerned administrators as starting points, and supporting data, when reacting to partisan policy making such as the recent Florida bill that mandates surveys focused on ensuring ideological diversity. Given the moderate nature of all institutions in the dataset, as well as the general lack of extreme values within the institutional dataset, the data collected and analyzed for this study could serve as a baseline for informing practice and policy at an institutional, state, and federal level.

Further, the findings presented in this dissertation could be used to advocate for greater interaction across ideological boundaries in support of specific college outcomes. Diversity experiences have long been associated with reducing prejudices and improving cognitive development in the college context (Antonio et al., 2004; Bowman, 2010; Chang, 1999; Chang et al., 2004, 2006; Denson & Chang, 2009), and ideological diversity experiences can be just as beneficial. As Luo (2021) demonstrated using survey data, college ideological diversity experiences benefit skill development and postcollege outcomes such as earnings and life satisfaction. Recent research has also identified that ideology strongly influences postcollege outcomes, such as job and social satisfaction, and that exploring political ideology and being politically engaged in college benefits students' civic and emotional development (Johnson & Ferguson, 2018). In essence, there are clear benefits to 'breaking the bubble.'

Considerations for Policy

Finally, there are some clear implications for policy from this study. First, as discussed in Chapter 1, there has been historic disinvestment in higher education in the United States of America. This disinvestment has been the result of decades of partisan policy making (Dar, 2012), increasing political polarization and a conservative emphasis on austerity measures that have resulted in slashes to federal and state education budgets (Dar & Lee, 2014), and growing political division (Cantwell & Taylor, 2020) that has resulted in heightened partisanship, white racial resentment, and reduced state support for higher education (Taylor et al., 2020).

As described in Chapter 1, this historic disinvestment and strategic targeting of higher education has come as a result of political extremes (Cantwell & Taylor, 2020; Dar & Lee, 2014; Parker, 2019). One of the key findings is that students in higher education are, at least politically, not extreme. There is representation along the most extreme portions of the political spectrum but students and their institutions are exceedingly moderate and not lacking for ideological diversity. Given the relative marginality of political extremists in higher education, and recent research that indicates that higher education institutions are not themselves very politicizing, much less liberalizing (Havey, 2023; Havey & Schalewski, 2022), this study can hopefully inform future policy by grounding arguments for disinvestment and increasingly draconian regulation in actual, empirical data rather than anecdotal evidence and nonrepresentative surveys.

CHAPTER 8: CONCLUSION

This study focused on evaluating student politics using an innovative, social media approach to assess ideological skew on college campuses. Driven by persistent criticisms of higher education as liberally skewed and the material impacts that perception has had on institutions of higher education, I collected digital trace data on 8,554 students across 43 states and 139 institutions. After collecting this data, I used latent attribute analysis to estimate individual students' political ideologies and collected their information networks, the news sources they followed on Twitter, to investigate whether the information students were consuming was skewed or lacking in ideological diversity.

Using this nationally representative dataset, I explored whether there was a clear ideological skew towards liberalism on college campuses and if there was a lack of ideological diversity on those college campuses. I identified that the average college student is moderate, but leans left, and that there is no lack of ideological diversity on college campuses in the United States of America. In comparison to other, survey-based research on this topic, the findings of this study indicate a potential overrepresentation of liberal students in past research as well as an underrepresentation of conservative students. Consistent with other research, institutional variables such as cost of attendance, selectivity, and control were not significantly predictive of students' estimated political ideologies. Given previous research that identifies on-campus interactions as the driving force of student politics, and changes to those politics, the lack of variance explained by institutional variance is neither surprising nor concerning. Overall, the findings presented in this study suggest significantly more variance with respect to student political ideology, and a smaller political middle, than previously identified in empirical research on the topic. College campuses, though regularly perceived as politically extreme, are not.

With respect to students' information networks, the findings indicate that students consume a diverse swathe of information online. The average estimated political ideology of the sources in students' networks was significantly more moderate than the average student, and students on the left were less likely to find ideological alignment between the information they consumed on social media platforms like Twitter and their own political positions. Students' information networks were ideologically diverse, but became less diverse as students themselves became more conservative. Finally, the majority of sources students followed on Twitter were center right and significantly more moderate than their own political ideologies. Given this reality, this study found support for the theory of constrained choice online presented earlier in this study.

There were clear implications from this study for research, practice, and policy, but the major takeaways are 1) that digital trace data can and should be incorporated into research on student populations, particularly given the increasingly online nature of the contemporary student, 2) that college campuses are not as politically extreme as they are perceived and criticized to be, and 3) that homophily and selective exposure is more prevalent among conservative students, who are more likely to be in an ideological bubble than their liberal peers.

Finally, I hope this study is primarily used as evidence that college campuses are not places of indoctrination or political extremes and that institutions of higher education in general are not lacking for ideological diversity. Select students may exist in filter bubbles, or be more likely to spend time with students whose political stripes match theirs, but this is no different than in the general public, where homophily and siloing is most prevalent among conservatives, and it is certainly not cause for alarm or literal and figurative disinvestment from higher education as a whole.

APPENDIX A

Table 1: Outlets and Their Estimated Political Ideologies

Twitter User ID	Outlet / Name	EPI
538politics	538 Politics	0.557396
ABC	ABC News	0.329786
abcnews	ABC News	0.733631
AFP	Agence France-Presse	1.022981
afpfr	Agence France-Presse	1.022981
ajc	Atlanta Journal-Constitution	0.99905
AJENews	Al Jazeera	-0.21633
AJEnglish	Al Jazeera	-1.01959
alfranken	The Al Franken Podcast	-1.5658
ALNewsNetwork	Alabama News	0.811984
AlterNet	AlterNet	-1.56037
amconmag	The American Conservative	1.341312
AmericanThinker	American Thinker	1.789981
AmerIndependent	American Independent	-0.3101

amspectator	The American Spectator	1.568222
AP	The Associated Press	1.206434
AppleNews	Apple News Today	0.899744
ArmchairExpPod	Armchair Expert	-1.18858
ArmyTimes	Army Times	-2.30035
arstechnica	ARS Technica	-1.19898
ASlavitt	In the Bubble with Andy Slavitt	-1.72348
ATLBlackStar	Atlanta Black Star	-0.64327
axios	Axios	1.078187
AxiosReCap	Axios Today	1.163997
azcentral	Arizona Central	1.012451
baltimoresun	Baltimore Sun	0.353372
BBCBreaking	BBC	0.529514
BBCWorld	BBC World	0.796901
BearingArmsCom	Bearing Arms	1.685754
beforeitsnews	Before It's News	1.447207
benfergusonshow	The Ben Ferguson Pod	1.473155

benshapiro	The Ben Shapiro Show	1.850335
BGOV	Bloomberg Government	1.042919
billboard	Billboard	-0.78799
BostonGlobe	Boston Globe	0.836247
bostonherald	Boston Herald	1.022836
BreitbartNews	Breitbart	2.025166
BreitbartTech	Breitbart Tech	2.338333
BreneBrown	Unlocking Us with Brene Brown	-1.32957
BulwarkOnline	The Bulwark	2.057575
business	Bloomberg News	0.812297
BusinessInsider	Business Insider	0.842065
BuzzFeed	BuzzFeed	0.116836
CBNNews	Christian Broadcasting Network	0.120134
CBNOnline	Christian Broadcasting Network	1.594766
CBS	CBS News	1.298063
CBSLA	CBS LA	0.599713
CBSNews	CBS News	0.561052

CFO	CFO	0.63728
CharlieKirk11	The Charlie Kirk Show	1.021501
chicagotribune	Chicago Tribune	1.483578
CNBC	CNBC	0.784996
CNET	CNET	0.533933
CNETNews	CNET	0.739251
CNN	CNN	0.973557
cnnbrk	CNN Breaking News	0.385415
CollegeFix	The College Fix	1.787595
coloradodaily	Colorado Daily	-0.51993
commondreams	Common Dreams	-1.58749
Consortiumnews	Consortium News	-0.37114
CR	Conservative Review	1.874443
crookedmedia	Crooked Media	-2.24414
crooksandliars	Crooks and Liars	0.52978
csmonitor	Christian Science Monitor	0.953158
CTmagazine	Christianity Today	1.105942

curaffairs	Current Affairs	1.209213
DailyCaller	Daily Caller	1.378049
dailydot	Daily Dot	0.580197
dailykos	Daily Kos	-2.06565
DailyMail	Daily Mail	0.42426
DailySignal	Daily Signal	1.510382
dallasnews	Dallas Morning News	1.1561
davidaxelrod	The Axe Files (with David Axelrod)	0.482316
dcexaminer	Washington Examiner	1.326815
DEADLINE	Deadline	1.009075
DeadlineWH	Deadline	1.009075
defense_news	Defense News	1.28995
democracynow	Democracy Now	0.535568
denverpost	Denver Post	0.842916
DeseretNews	Deseret News	1.814912
DNewsOpinion	Deseret News	1.226862
DrTurleyTalks	Turley Talks	2.340149

economics	Bloomberg Economics	0.682349
EconomistRadio	The Economist (Podcast)	0.179752
engadget	Engadget	0.299701
EpochTimes	Epoch Times	1.477463
FAIRMediaWatch	FAIR	-1.79508
FDRLST	The Federalist	1.479967
financialbuzz	Financial Buzz	0.678021
FiveThirtyEight	FiveThirtyEight	-1.05336
Forbes	Forbes	0.55705
ForeignPolicy	Foreign Policy	0.164173
FortuneMagazine	Fortune	0.33246
FoxNews	Fox News	1.506728
FreeBeacon	Washington Free Beacon	1.283718
freep	Detroit Free Press	0.916478
freespeechtv	Free Speech TV	-1.69231
FT	Financial Times	0.394507
glennbeck	The Glenn Beck Program	1.437909

goodnewsnetwork	Good News Network	-1.18944
guardian	The Guardian	0.343094
hartfordcourant	Hartford Courant	-0.42657
HillReporter	Hill Reporter	0.553383
HoustonChron	Houston Chronicle	0.80959
HuffPost	HuffPost	1.056025
HuffPostCollege	HuffPost College	0.359168
HuffPostEdu	HuffPostEducation	-1.31522
Independent	The Independent	-0.04464
indystar	Indianapolis Star	0.912859
insideclimate	Inside Climate News	0.425604
inthesetimesmag	In These Times	-2.02446
IQ2US	Intelligence Squared US	1.130849
jacobinmag	Jacobin	-2.42208
Jezebel	Jezebel	-2.03129
joerogan	The Joe Rogan Experience	0.927301
jonlovett	Lovett or Leave It	0.412332

journalsentinel	Milwaukee Journal Sentinel	1.030057
JudicialWatch	Judicial Watch	1.59061
kairyssdal	Make Me Smart	1.147055
KCStar	Kansas City Star	0.700452
laconiadailysun	Laconia Daily Sun	1.600627
latimes	Los Angeles Times	0.786271
LifeNews	Life News	0.50599
LifeZette	LifeZette	1.622148
MailOnline	Daily Mail	1.196916
Marketplace	Marketplace (Podcast)	0.889297
MarketWatch	Marketwatch	1.288964
marklevinshow	The Mark Levin Show	2.031643
martinepowers	Post Reports	-0.56663
mashable	Mashable	0.176562
mediaite	MediaIte	1.1007
megynkelly	The Megyn Kelly Show	1.298299
MegynKellyShow	The Megyn Kelly Show	1.491943

mercnews	Mercury News	-0.09385
michaeljknowles	The Michael Knowles Show	1.763786
mikiebarb	The Daily	0.963674
mollywood	Make Me Smart	0.275176
monthly	Washington Monthly	-1.36911
MotherJones	Mother Jones	-2.43486
myfairobserver	Fair Observer	0.915647
NatCounterPunch	Counterpunch	-1.63482
NatEnquirer	National Enquirer	0.55947
NationalFile	National File	2.030604
NBCLA	NBC LA	0.444003
NBCNews	NBC News	-0.58378
neutralnews	Neutral News	1.265621
NewAbnormalPod	The New Abnormal	0.347996
NewAmericanMag	The New American	1.865105
newrepublic	New Republic	0.533944
newsbusters	News Busters	1.765539

newsday	NewsDay	0.542141
Newser	Newser	0.696895
newsmax	NewsMax	1.61212
NewsNationNow	NewsNation Now	-0.52122
newsone	NewsOne	-1.39126
Newsweek	NewsWeek	0.05397
Newsy	Newsy	0.708822
NewYorker	The New Yorker	-0.78714
njdotcom	NJ.com	0.489625
NOLAnews	NOLA.com	0.381592
novapbs	Nova PBS	0.046341
NPR	NPR	1.08607
npratc	Consider This (from NPR)	0.046379
NPRCodeSwitch	Code Switch	-0.69059
nprfreshair	Fresh Air	-0.40905
NPRNewsNow	NPR News Now	1.08607
nprpolitics	NPR Politics Podcast	1.159208

NRO	National Review	1.530468
NYDailyNews	New York Daily News	0.375823
NYMag	New York Magazine	0.480357
nypost	New York Post	1.351933
nytclimate	NYT Climate	-1.53751
nytimes	The New York Times	-1.5741
nytimesworld	NYT World	0.563206
nytopinion	New York Times Opinion	0.458477
OANN	OAN Network	1.423893
OccupyDemocrats	Occupy Democrats	-0.08171
ocregister	Orange County Register	1.262389
onthemedia	On the Media	0.778989
Oregonian	The Oregonian	0.704493
orlandosentinel	Orlando Sentinel	0.311647
OWHnews	Omaha World-Herald	0.886418
ozy	Ozy	-1.17234
patribotics	Patribotics	1.12402

PBS	PBS	-0.18421
PBSDS	PBS	-0.44995
PBSSoCal	PBS	-1.00592
petersuderman	The Reason Roundtable	1.072177
PittsburghPG	Pittsburgh Post-Gazette	0.801011
PJMedia_com	PJ Media	1.373369
planetmoney	Planet Money	-1.21145
PodSaveAmerica	Pod Save America	-2.2925
politico	Politico	1.226439
politicususa	Politicus	-2.00371
POPSUGAR	Popsugar	-1.26183
PostStandard	Syracuse Post-Standard	0.84296
prageru	PragerU	1.387027
PreetBharara	Stay Tuned with Preet	0.836057
ProjectLincoln	The Lincoln Project	0.496148
propublica	ProPublica	-0.1553
Quillette	Quillette	0.249397

qz	Quartz	-1.27277
RadioTimes	Radio Times	-0.57404
Rasmussen_Poll	Rasmussen Reports	1.900267
RawStory	Raw Story	0.534789
RealCandaceO	The Candace Owens Show	1.404249
RealClearNews	RealClear Politics	1.20304
realDailyWire	The Daily Wire	1.862288
reason	Reason	1.160656
RedState	RedState	1.950823
Reuters	Reuters	0.329842
reviewjournal	Las Vegas Review Journal	1.186867
RightWingWatch	Right Wing Watch	1.168313
rollcall	Roll Call	1.228421
Roughly	Rough Translation	-1.93051
RT_com	Russia Today	1.220437
RubinReport	The Rubin Report	1.414917
Salon	Salon	0.685338

SCMPNews	South China Morning Post	0.649584
SCrowder	Louder with Crowder	1.7128
seattlepi	Seattle PI	0.78793
SecondNexus	Second Nexus	-1.75661
sfchronicle	San Francisco Chronicle	-0.06142
sfexaminer	SF Examiner	-1.33672
SFGate	SF Gate	0.84829
sfindependent	Independent Journal	-0.37331
shadowproofcom	Shadowproof	-2.04707
Slate	Slate	0.063597
SlateGabfest	Political Gabfest	1.06057
sltrib	Salt Lake Tribune	1.182547
snopes	Snopes	0.23481
Sojourners	Sojourners	0.389041
SputnikInt	Sputnik International News	0.027663
starsandstripes	Stars and Stripes	0.638487
StartHereABC	Start Here	1.002761

StarTribune	Star Tribune-Minneapolis	0.676742
SunSentinel	Sun Sentinel	1.066508
Suntimes	Chicago Sun-Times	0.779888
SykesCharlie	The Bulwark Podcast	1.281244
SYSKPodcast	Stuff You Should Know	-0.89381
TB_Times	Tampa Bay Times	0.98586
teamtrace	The Trace	0.285624
TechCrunch	TechCrunch	0.880202
TeenVogue	Teen Vogue	-1.11886
Tennessean	Tennessean	1.118627
TheAdvocateMag	Advocate	-1.6401
theamgreatness	American Greatness	1.940906
TheAspenTimes	Aspen Times	0.165323
TheAtlantic	The Atlantic	0.1664
theblaze	The Blaze	1.564098
thedailybeast	The Daily Beast	0.717455
TheDailyShow	The Daily Show	0.564617

thedispatch	The Dispatch	2.099772
TheEconomist	The Economist	0.179752
TheFiscalTimes	Fiscal Times	1.184618
TheGrayzoneNews	The Grayzone News	-0.5509
theGrio	TheGrio	-1.63579
thehill	The Hill	1.268027
theinquisitr	Inquisitr	1.392829
theintercept	The Intercept	-0.7333
TheLastRefuge2	The Last Refuge	1.572432
thenation	The Nation	-0.28512
theprogressive	The Progressive	0.579812
TheRightScoop	The Right Scoop	2.245264
TheRoot	The Root	0.11318
theskimm	The Skimm	0.362405
TheWeek	The Week	0.855985
thinkprogress	ThinkProgress	-1.36242
ThisAmerLife	This American Life	-1.05143

TIME	Time Magazine	0.203706
TMZ	TMZ	0.158379
townhallcom	Townhall	1.762433
TPM	Talking Points Memo	0.35744
trish_regan	Trish Intel	1.332012
truthout	Truthout	-1.05841
TucsonStar	Arizona Daily Star	1.143182
TwitchyTeam	Twitchy	1.907385
UpFirst	Up First	-1.51354
UPI	UPI	0.665729
Upworthy	Upworthy	-0.84939
USATODAY	USA Today	0.443197
usnews	US News and World Report	1.205705
VanityFair	Vanity Fair	0.04879
Variety	Variety	1.084381
VICENews	VICE News	-0.95415
VOANews	Voice of America	0.442889

voxdotcom	Vox	-0.08375
washingtonpost	Washington Post	0.721329
WashTimes	Washington Times	1.566724
weatherchannel	The Weather Channel	0.664555
WestJournalism	Western Journal	1.657436
WhatsNewsWSJ	What's News	1.1857
Wonkette	Wonkette	-1.52707
worldnetdaily	WND	1.945055
WSJ	Wall Street Journal	1.370485
WSJPodcasts	The Journal	1.370485
wvgazettemail	Charleston Gazette-Mail	0.470223
zerohedge	ZeroHedge	1.15548
wvgazettemail	Charleston Gazette-Mail	0.470223
zerohedge	ZeroHedge	1.15548

APPENDIX B

Table 1: Average Estimated Political Ideology Per Institution

Institution	Average Estimated Political Ideology	Students
American University	-0.745655945	169
Arapahoe Community College	0.146464516	3
Arizona State University	0.077739406	232
Arkansas State University	0.317717771	45
Auburn University	0.861494306	56
Augsburg University	-0.073744722	42
Ball State University	-0.278811083	133
Bellevue University	-0.325165701	15
Bergen Community College	-0.175040344	2
Boise State University	-0.208175101	50
Boston College	-0.488635076	97
Boston University	-0.632322249	118
Bowling Green State University	-0.547560945	139
Brigham Young University	0.126106773	78

Broward College	0.770067013	2
Brown University	-0.64155351	85
Carnegie Mellon University	-0.631714293	70
Central Michigan University	-0.472042457	270
Century College	-0.984162214	11
Citrus College	0.347633781	7
Clark College	-0.747993059	21
Clemson University	0.152976126	200
Cleveland State University	-0.368177471	86
College of DuPage	0.094984598	15
College of William and Mary	-0.547499893	135
Colorado State University	-0.400868002	69
Columbia University	-0.624067855	285
Cornell University	-0.448479438	140
Dartmouth College	-0.747387106	165
Davidson College	-0.702504901	103
DePaul University	-0.501934361	64

Des Moines Area Community College	0.089209548	10
Drake University	-0.524231689	201
Drexel University	-0.515079102	56
Duke University	-0.439938041	79
East Carolina University	0.301414011	110
East Tennessee State University	0.390443232	38
Eastern Washington University	0.210033195	7
El Camino Community College	0.568878471	47
El Paso Community College	0.922482215	2
Florida International University	-0.320806208	39
Florida State University	-0.232129569	64
George Mason University	-0.93078274	46
George Washington University	0.326215969	79
Georgetown University	-0.395923226	141
Georgia Southern University	0.044983067	95
Harrisburg Area Community College	0.429566694	2

Harvard University	-0.661403182	143
Hinds Community College	-0.391177421	3
Howard University	-0.895869634	130
Iowa State University	-0.033035506	237
Kansas State University	0.441583164	40
Kennesaw State University	-0.195064254	44
Kent State University	-0.645529071	84
Liberty University	0.913530292	63
Louisiana State University	-0.421848538	115
Marquette University	-0.166018326	187
Marshall University	-0.090672507	78
Miami University-Oxford	0.50364102	32
Middlebury College	-0.809666935	47
Montclair State University	-0.328715859	44
Nassau Community College	0.662845484	3
New York University	-0.602413407	148
North Dakota State University	0.092117277	21

Old Dominion University	0.314725025	47
Pierce College	-0.103594756	19
Plymouth State University	-0.36810383	12
Portland State University	-0.74911351	19
Post University	0.483765202	2
Princeton University	-0.744587883	87
Rhode Island College	-0.825973203	9
Rowan University	-0.107703472	61
Saint Cloud State University	-0.000787473	19
Salt Lake Community College	0.028616074	11
Seattle University	-0.467988493	23
South Dakota State University	0.076892611	21
Stanford University	-0.310637062	142
The Ohio State University	-0.165282343	245
UC Berkeley	-0.784832127	119
UC Davis	-0.672696104	76
UC Merced	-0.529226398	42

UC Riverside	-0.459740194	68
UC San Diego	-0.911693478	150
UC Santa Barbara	-0.352413241	32
UC Santa Cruz	-0.975302083	55
UCLA	-0.438587729	77
University of Alaska Anchorage	-0.026641301	12
University of Alaska Fairbanks	-0.773551057	3
University of Arizona	0.008497078	70
University of Arkansas	-0.116379984	213
University of Idaho	0.099721254	36
University of Maine	-0.381744585	27
USC	-0.290346815	119
UT Austin	-0.362107099	164
Vanderbilt University	-0.649055516	157
Wake Forest University	-0.325081482	131
Washington State University	0.012085403	69
Wichita State University	0.224886313	48

Wilmington University	0.339900807	25
Yale University	-0.498164198	146

REFERENCES

Alwin, D. E, Cohen, R. L., & Newcomb, T. M. (1991). *Political attitudes over the life span.* University of Wisconsin Press.

Alwin, D.E., & Krosnick, J. A. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology*, 97(1), 169-195.

Antonio, A. L., Chang, M. J., Hakuta, K., Kenny, D. A., Levin, S., & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science*, 15(8), 507-510.

Armstrong, E. A., & Hamilton, L. T. (2013). *Paying for the Party*. Harvard University Press.

Astin, A. (1977). *Four critical years*. Jossey-Bass.

Astin, A. (1993). *What matters in college? Four critical years revisited*. Jossey-Bass.

Astin, A., & lising-antonio, a. (2012). *Assessment for Excellence: The Philosophy & Practice of Assessment & Evaluation in Higher Education*. Rowan & Littlefield Publishing Group.

Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, 43(3-4), 465-482.

Bailey, M., & Williams, L. R. (2016). Are college students really liberal? An exploration of student political ideology and attitudes toward policies impacting minorities. *The Social Science Journal*, 53(3), 309-317.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23(1), 76-91.

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of Twitter users. *Social Science Computer Review*, 33(6), 712-729.

Benenson, J., & Bergom, I. (2019). Voter Participation, Socioeconomic Status, and Institutional Contexts in Higher Education. *The Review of Higher Education*, 42(4), 1665-1688.

Bennion, E. A., & Nickerson, D. W. (2021). Decreasing Hurdles and Increasing Registration Rates for College Students: An Online Voter Registration Systems Field Experiment. *Political Behavior*, 1-22.

Berry, J. M., & Sobieraj, S. (2013). *The outrage industry: Political opinion media and the new* *incivility.* Oxford University Press.

Bhatt, I., & MacKenzie, A. (2019). Just Google it! Digital literacy and the epistemology of ignorance. *Teaching in Higher Education*, 24(3), 302-317.

Binder, A. J., & Wood, K. (2014). *Becoming right: How campuses shape young conservatives* Princeton University Press.

Binder, A. J., & Kidder, J. L. (2022). *The Channels of Student Activism: How the Left and Right Are Winning (and Losing) in Campus Politics Today*. University of Chicago Press.

Biondi, M. (2012). *The Black revolution on campus.* University of California Press.

Birnbaum, R. (1983). *Value of different kinds of colleges. In J. L. Bess (Ed.), Foundations of American higher education.* Simon & Schuster.

Bloom, A. (2008). *Closing of the American mind*. Simon and Schuster.

Bowen, W. G. (1977). Economic problems confronting higher education: An institutional perspective. *The American Economic Review*, 67(1), 96-100.

Bowman, N. A. (2010). College diversity experiences and cognitive development: A meta-analysis. *Review of Educational Research*, 80(1), 4-33.

Bowman, N. A., & Seifert, T. A. (2011). Can college students accurately assess what affects their

learning and development? *Journal of College Student Development*, 52, 270-290.

boyd, d. (2008a). Can social network sites enable political action. *International Journal of Media and Cultural Politics*, 4(2), 241-244.

boyd, d. (2008). Why youth (heart) social network sites: The role of networked publics in teenage social life. In D. Buckingham (Ed.), *Youth, Identity, and Digital Media* (pp. 119–142). MIT Press.

boyd, d., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.

Boyte, H. & Hollander, E. (1999). *Wingspread declaration on renewing the civic mission of the American research university.* http://www.compact.org/civic/Wingspread/Wingspread.html

Buckley, W. F. (1986). *God and Man at Yale: The Superstitions of 'Academic Freedom'.* Regnery

Publishing.

Cantwell, B., & Taylor, B. J. (2020). Political Rancor and Educational Inequality: Why Building Consensus Is Necessary to Renew American Higher Education. *Change: The Magazine of Higher Learning*, 52(3), 68-72.

Chang, M. J. (1999). Does racial diversity matter? The educational impact of a racially diverse undergraduate population. *Journal of College Student Developmen*t, 40(4), 377–395.

Chang, M. J., Astin, A. W., & Kim, D. (2004). Cross-racial interaction among undergraduates: Some consequences, causes, and patterns. *Research in Higher Education*, 45(5), 529-553.

Chang, M. J., Denson, N., Saenz, V., & Misa, K. (2006). The educational benefits of sustaining cross-racial interaction among undergraduates. *The Journal of Higher Education*, 77(3), 430-455.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal* *of Communication*, 64(2), 317-332.

Daniels, J. (2018). The algorithmic rise of the “alt-right”. *Contexts*, 17(1), 60-65.

Dar, L. (2012). The political dynamics of higher education policy. *The Journal of Higher Education*, 83(6), 769-794.

Dar, L., & Lee, D. W. (2014). Partisanship, political polarization, and state higher education budget outcomes. *The Journal of Higher Education*, 85(4), 469-498.

Davies, J. C. (1965). The family's role in political socialization. *The Annals of the American Academy of Political and Social Science*, 361(1), 10-19.

Davis, J. A. (1992). Changeable weather in a cooling climate atop the liberal plateau: Conversion and replacement in forty-two General Social Survey items, 1972-1989. P*ublic Opinion Quarterly*, 56, 261-306.

Denson, N., & Chang, M.J. (2009). Racial diversity matters: The impact of diversity-related student engagement and institutional context. *American Educational Research Journal*, 46(2), 322-353.

Dey, E. L. (1988). C*ollege impact and student liberalism revisited: The effect of student peers.* Paper presented at the meeting of the Association for the Study of Higher Education, St. Louis, MO.

Dey, E. L. (1996). Undergraduate political attitudes: An examination of peer, faculty, and social influences. *Research in Higher Education*, 37(5), 535–554.

Dey, E. L. (1997). Undergraduate political attitudes: Peer influence in changing social contexts. *The Journal of Higher Education*, 68(4), 398-413.

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, 102(3), 690-755.

Donovan, J., & Boyd, D. (2021). Stop the presses? Moving from strategic silence to strategic amplification in a networked media ecosystem. *American Behavioral Scientist*, 65(2), 333-350.

Doyle, W. R. (2007). Public opinion, partisan identification, and higher education policy. T*he Journal of Higher Education*, 78(4), 369-401.

Doyle, W. R. (2010). US Senator's Ideal Points for Higher Education: Documenting Partisanship, 1965–2004. *The Journal of Higher Education*, 81(5), 619-644.

d'Souza, D. (1991). *Illiberal education: The politics of race and sex on campus.* Simon and Schuster.

Dugan, J. P. (2015). The measurement of socially responsible leadership: Considerations in establishing psychometric rigor. *Journal of Educational, Cultural, & Psychological Studies*, 12, 23-42. doi:10.7358/ecps–2015-012-duga

Dunkel, C. S., & Decker, M. (2012). Using identity style and parental identification to predict political orientation. *Current Psychology*, 31(1), 65-78.

Eagan, K., Stolzenberg, E. B., Ramirez, J. J., Aragon, M. C., Suchard, M. R., & Hurtado, S. (2014). *The American freshman: National norms fall 2014.* Higher Education Research Institute, UCLA.

Eagan, M. K., Stolzenberg, E. B., Zimmerman, H. B., Aragon, M. C., Whang Sayson, H., & Rios-Aguilar, C. (2017). *The American freshman: National norms 2016.* Higher Education Research Institute, UCLA.

Ediger, D., Jiang, K., Riedy, J., Bader, D. A., Corley, C., Farber, R., & Reynolds, W. N. (2010, September). *Massive social network analysis: Mining twitter for social good*. In 2010 39th International Conference on Parallel Processing (pp. 583-593). IEEE.

Ehrlich, T. (Ed.). (2000). *Civic responsibility and higher education*. Greenwood Publishing Group.

Elchardus, M., & Spruyt, B. (2009). The culture of academic disciplines and the sociopolitical attitudes of students: A test of selection and socialization effects. *Social Science Quarterly*, 90(2), 446–460.

Ellison, N. B., & boyd, d. (2013). Sociality through social network sites. *The Oxford handbook of internet studies*, 151-172.

Feldman, K. A. (Ed.) (1972). *College and student.* Pergamon.

Feldman, K. A., & Newcomb, T. M. (1969). *The impact of college on students.* Jossey-Bass.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9(2), 319-342.

Gándara, D., & Ness, E. C. (2019). Ideological think tanks and the politics of college affordability in the states. *The Journal of Higher Education*, 90(5), 717-743.

Gil de Zúñiga, H., & Chen, H. T. (2019). Digital media and politics: Effects of the great information and communication divides. *Journal of Broadcasting & Electronic Media*, 63:3, 365-373, DOI: 10.1080/08838151.2019.1662019

Gillespie, T. (2010). The politics of 'platforms'. *New Media & Society*, 12(3), 347-364.

Gillespie, T. & Aufderheide, P. & Carmi, E. & Gerrard, Y. & Gorwa, R. & Matamoros-Fernández, A. & Roberts, S. T. & Sinnreich, A. & Myers West, S. (2020). Expanding the debate about content moderation: scholarly research agendas for the coming policy debates. *Internet Policy Review*, 9(4). DOI: 10.14763/2020.4.1512

Gottfried, J., & Shearer, E. (2016). *News use across social media platforms 2016.* Pew Research

Center.

Gross, N. (2013). *Why are professors liberal and why do conservatives care?* Harvard University Press.

Gross, N., & Simmons, S. (Eds.). (2014). *Professors and their politics.* JHU Press.

Gurin, G. (1971). *The impact of the college experience. In S. Withey (Ed.), A degree and what else? Correlates and consequences of a college education*. McGrawHill.

Gutmann, A. (1987). *Democratic Education.* Princeton University Press.

Gutmann, A., & Ben-Porath, S. (2014). Democratic education. *The Encyclopedia of Political Thought*, 863-875.

Gutmann, A., & Thompson, D. F. (1998). *Democracy and disagreement*. Harvard University Press.

Hanson, J. M., Weeden, D. D., Pascarella, E. T., & Blaich, C. (2012). Do liberal arts colleges make students more liberal? Some initial evidence. *Higher Education*, 64(3), 355-369.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4(1), 468–494.

Havey N. (2020a) "Radicalized on Campus? (Un)Coded Whiteness as Campus Social Movement", J*ournal of Critical Thought and Praxis,* 10(1). https://doi.org/10.31274/jctp.11566

Havey, N. (2020b). Partisan public health: how does political ideology influence support for COVID-19 related misinformation?. *Journal of Computational Social Science*, 3(2), 319-342.

Havey, N. (2021). The New Right: Conservative student political repertoires and intragroup conflict. *Education Policy Analysis Archives*, 29(134). https://doi.org/10.14507/epaa.29.5669

Havey, N. (2023). The "Liberal" Campus: An Investigation of Changes to Student Political Ideology. *In Press. The Journal of College and Character.*

Havey, N., & Schalewski, L. (2022). Political Development Away From Home: Understanding How the First Year Influences College Students' Political Orientation Change. *Journal of The First-Year Experience & Students in Transition*, 34(1), 27-42.

Herzog, S., & Bowman, N. A. (Eds.). (2011). *New Directions for Institutional Research: No. 150. Validity and limitations of student self-report data*. Jossey-Bass.

Himelboim, I., McCreery, S., & Smith, M. (2013). Birds of a feather tweet together: Integrating network and content analyses to examine cross-ideology exposure on Twitter. Journal of *Computer-Mediated Communication*, 18(2), 154-174.

Hindman, M. (2008). *The myth of digital democracy*. Princeton University Press.

Horowitz, D. (2009). *Indoctrination U.: The left's war against academic freedom.* Encounter Books.

Howe, E. C., & Fosnacht, K. (2017). Promoting democratic engagement during college: Looking beyond service-learning. *Journal of College and Character*, 18(3), 155-170.

Hyman, H. H., & Wright, C. R. (1978). *The enduring effects of education on values.* University of Chicago Press.

Johnson, M. R., Dugan, J. P., & Soria, K. M. (2017). " Try to See it My Way": What Influences Social Perspective Taking Among College Students? J*ournal of College Student Development*, 58(7), 1035-1054.

Johnson, M. R., & Ferguson Jr, M. (2018). The role of political engagement in college students' civic identity: Longitudinal findings from recent graduates. *Journal of College Student Development*, 59(5), 511-527.

Johnson, M. R., & Peacock, J. (2020). Breaking the bubble: Recent graduates' experiences with ideological diversity. *Journal of Diversity in Higher Education*, 13(1), 56.

Jost, J. T., Nosek, B. A., & Gosling, S. D. (2008). Ideology: Its resurgence in social, personality, and political psychology. P*erspectives on Psychological Science,* 3(2), 126-136.

Judis, J. B. (2001). *William F. Buckley, Jr.: Patron saint of the conservatives.* Simon and Schuster.

Jurkowitz, M. (2014) *Small digital news sites: young, lean and local.* Pew Research Center.

Kissel, A. (2020). *Campus Free Speech: A Cultural Approach. Sketching a New Conservative Education Agenda.* American Enterprise Institute.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. *In Proceedings of the 19th international conference on the World wide web* (pp. 591-600).

Labaree, D. F. (1997). *How to succeed in school without really learning: The credentials race in American education.* Yale University Press.

Laukaitis, J. J. (2013). The academy on the firing line: William F. Buckley, Jr.'s God and Man at Yale and the modern conservative critique of higher education. *American Educational History Journal*, 40(1/2), 129.

LeVine, R. A. (1966). *American college experience as a socialization process. In T. M. Newcomb & E. K. Wilson (Eds.), College peer groups: Problems and prospects for research.* Aldine.

Licari, P. R. (2020). Sharp as a fox: Are Foxnews. com visitors less politically knowledgeable?. *American Politics Research*, 48(6), 792-806.

Luo, J. (2021). Interaction across Ideological Boundaries and College Outcomes. *The Journal of Higher Education*, 92(1), 56-83.

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114-133.

McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2017). Using Twitter for demographic and social science research: Tools for data collection and processing. *Sociological Methods & Research*, 46(3), 390-421.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413-439.

Metzger, M. J., Hartsell, E. H., & Flanagin, A. J. (2020). Cognitive dissonance or credibility? A comparison of two theoretical explanations for selective exposure to partisan news. *Communication Research*, 47(1), 3-28.

Mihailidis, P., & Thevenin, B. (2013). Media literacy as a core competency for engaged citizenship in participatory democracy. *American Behavioral Scientist*, 57(11), 1611-1622.

Miller, W., & Gunnels, K. (2020). *Engagement in Higher Education: Building Civil Society through Campus Activism.* In *Civil Society and Social Responsibility in Higher Education: International Perspectives on Curriculum and Teaching Development.* Emerald Publishing Limited.

Morgan, D. L., & Davis III, C. H. (Eds.). (2019). *Student activism, politics, and campus climate in higher education.* Routledge.

Munger, K. (2020). All the news that's fit to click: The economics of clickbait media. *Political Communication*, 37(3), 376-397.

Nemeth, J. (2020). The Passion of William F. Buckley: Academic Freedom, Conspiratorial Conservatism, and the Rise of the Postwar Right. *Journal of American Studies*, 54(2), 323-350.

Nguyen, Q., Poquet, O., Brooks, C., & Li, W. (2020). *Exploring Homophily in Demographics and Academic Performance Using Spatial-Temporal Student Networks*. International Educational Data Mining Society.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism.* NYU Press.

O'Connor, K. M., & McEwen, L. (2021). *Real World Learning Through Civic Engagement: Principles, Pedagogies and Practices.* In *Applied Pedagogies for Higher Education* (pp. 63-89). Palgrave Macmillan.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.

Parker, K. (2019). *The growing partisan divide in views of higher education.* Pew Research Center, August, 19, 2019.

Pascarella, E. T. (2001). Using student self-reported gains to estimate college impact: A cautionary tale. *Journal of College Student Development*, 42, 488-492.

Pascarella, E., & Terenzini, P. (1991). *How college affects students: Findings and Insights from Twenty Years of Research.* Jossey-Bass.

Pascarella, E., & Terenzini, P. (2005). *How college affects students (Vol. 2): A third decade of research.* Jossey-Bass.

Pascarella, E. T., Salisbury, M. H., Martin, G. L., & Blaich, C. (2012). Some complexities in the effects of diversity experiences on orientation toward social/political activism and political views in the first year of college. *The Journal of Higher Education*, 83(4), 467-496.

Pearson, G. D., & Knobloch-Westerwick, S. (2018). Perusing pages and skimming screens: Exploring differing patterns of selective exposure to hard news and professional sources in online and print news. *New Media & Society*, 20(10), 3580-3596.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, 116(7), 2521-2526.

Pike, G. R. (1999). The constant error of the halo in educational outcomes research. *Research in Higher Education*, 40, 61-86.

Porter, S. R. (2011). Do college student surveys have any validity? R*eview of Higher Education*, 35, 45-76.

Pryor, J. H., Hurtado, S., Saenz, V. B., Santos, J. L., & Korn, W. S. (2007). *The American freshman: Forty year trends*. Higher Education Research Institute.

Religion in America: U.S. Religious Data, Demographics and Statistics. (2022). Retrieved 13 January 2022, from https://www.pewforum.org/religious-landscape-study/compare/party-affiliation/by/state/

Rhoads, R. A. (1998). *Freedom's web: Student activism in an age of cultural diversity.* Johns Hopkins University Press.

Riddle, S., & Apple, M. W. (Eds.). (2019). *Re-imagining education for democracy*. Routledge.

Roberts, S. T. (2016). Commercial content moderation: Digital laborers' dirty work. *Media*

Studies Publications. 12.

Roberts, S. 2017. "*Content moderation*," In: Schintler, L & McNeely, C (Eds.). Encyclopedia of big data. Springer International.

Roberts, S. T. (2018). Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday*.

Roberts, S. (2021). *Your AI is a Human*. In Mullaney, T. S., Peters, B., Hicks, M., & Philip, K. (Eds.). (2021). Your Computer is on Fire, 51-70. MIT Press.

Sax, L. J. (2008). *The gender gap in college: Maximizing the developmental potential of women and men*. Jossey-Bass/Wiley.

Sax, L. J., Gilmartin, S. K., & Bryant, A. N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44(4), 409-432.

Schiff, T. W. (1993). *Political identification and political attitudes of American college students* (Doctoral dissertation, University of California, Los Angeles).

Schneider, G. L. (1999). *Cadres for Conservatism: Young Americans for freedom and the rise of the contemporary right.* NYU Press.

Shapiro, B. (2010). *Brainwashed: How universities indoctrinate America's youth.* Thomas Nelson.

Shearer, E. (2018.). *Social media outpaces print newspapers in the U.S. as a news source.* Pew Research Center.

Smith, E. (2020). *Don't Read the Comments.* Inkyard Press.

Soule, S. A. (1997). The student divestment movement in the United States and tactical diffusion: The shantytown protest. *Social Forces*, 75(3), 855-882.

Steinert-Threlkeld, Z. C. (2018). *Twitter as data.* Cambridge University Press.

Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2014). “Ditto heads” do conservatives perceive greater consensus within their ranks than liberals?. *Personality and Social Psychology Bulletin*, 40(9), 1162-1177.

Taylor, B. J., Cantwell, B., Watts, K., & Wood, O. (2020). Partisanship, white racial resentment, and state support for higher education. *The Journal of Higher Education*, 91(6), 858-887.

Terenzini, P. T. (1994). Educating for citizenship: Freeing the mind and elevating the spirit. *Innovative Higher Education*, 19(1), 7–21.

Thelin, J. R. (2011). *A history of American higher education.* JHU Press.

Trachtenberg, B. (2019). The People v. Their Universities: How Popular Discontent Is Reshaping

Higher Education Law. *Kentucky Law Journal*, 108, 47-62.

Twenge, J. M., Honeycutt, N., Prislin, R., & Sherman, R. A. (2016). More polarized but more independent: political party identification and ideological self-categorization among US adults, college students, and late adolescents, 1970-2015. *Personality and Social Psychology Bulletin*, 42(10), 1364-1383.

Weeks, B. E., Kim, D. H., Hahn, L. B., Diehl, T. H., & Kwak, N. (2019). Hostile Media Perceptions in the Age of Social Media: Following Politicians, Emotions, and Perceptions of Media Bias. *Journal of Broadcasting & Electronic Media*, 63(3), 374-392.

Weidman, 1. (1989). *Undergraduate socialization: A conceptual approach. In J. Smart (Ed.), Higher education: Handbook of theory and research* (Vol. 5). Agathon.

Weil, F. D. (1985). The variable effects of education on liberal attitudes: A comparative-historical analysis of anti-Semitism using public opinion survey data. *American Sociological Review*, 50 , 458-474.

Wilder, C. S. (2013). *Ebony and ivy: Race, slavery, and the troubled history of America's universities.* Bloomsbury Publishing USA.

Wineburg, S., & McGrew, S. (2019). Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information. *Teachers College Record*, 121(11), n11.

Woessner, M., & Kelly-Woessner, A. (2020). Why College Students Drift Left: The Stability of Political Identity and Relative Malleability of Issue Positions among College Students. *PS: Political Science & Politics*, 53(4), 657-664.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., ... & Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLoS computational biology*, 13(3), e1005399.