Mostly Sports and Congratulations: Twitter, Facebook, and Alberta Higher Education

by

Bradley Michael Congram

A Thesis Submitted to the Faculty of Social and Applied Sciences
in Partial Fulfilment of the Requirements for the Degree of

Master of Arts
In
Higher Education Administration and Leadership

Royal Roads University
Victoria, British Columbia, Canada

Supervisor: George Veletsianos
July 2019

COMMITTEE APPROVAL

The members of Bradley Congram's Thesis Committee certify that they have read the thesis titled *Mostly Sports and Congratulations: Twitter, Facebook, and Alberta Higher Education* and recommend that it be accepted as fulfilling the thesis requirements for the Degree of Master of Arts in Higher Education Administration and Leadership:

Dr. George Veletsianos [signature on file]
Dr. Jaigris Hodson [signature on file]

Final approval and acceptance of this thesis is contingent upon submission of the final copy of the thesis to Royal Roads University. The thesis supervisor confirms to have read this thesis and recommends that it be accepted as fulfilling the thesis requirements:

George Veletsianos [signature on file]

Creative Commons Statement

Table of Contents

Acknowledgements

An incredible number of people contributed to the successful completion of this thesis and to the completion of my degree. To my supervisor and my second committee member, George Veletsianos and Jaigris Hodson respectively, thank you so much for your assistance and guidance through this process. Your feedback was fantastically helpful and professional, and challenged my view on my topic in ways that forced me to develop more in depth understanding. You encouraged me to be a better academic, rather than simply running some data through a few tests and calling it a day. To my aunt, Donna Woloschuk, thank you for sharing your experience and talking me down when I needed some place to dump my panic and frustration. To everyone at Elite Training Centre and Elite Brazilian Jiu Jitsu, thank you for your understanding when I suddenly disappeared for a year saying only "I have to write my thesis." Finally, to Persia, thanks for putting up with my nonsense. You have been unjustifiably patient with me, and you continue to make life better in every way.

ABSTRACT

One of the defining features of social media is the capability for interaction, specifically for the audience to respond to previously posted content. Higher education research to date has focused on the content published by institutions, with minimal examination of the content sent back to institutions. Previous research has often operated as though higher education's social media is homogenous, without acknowledging the variation in institutional mission. Furthermore, previous research has tended to treat institutions' social media accounts as the same, without examination of differences between social media platforms. This research examines both gaps. Twitter posts mentioning primary and secondary accounts of 25 Albertan post-secondary institutions were gathered during a two-month interval alongside messages posted to the same institutions' primary Facebook pages, to determine which topics led the audience to communicate back to institutions and to examine whether any difference existed between account types and platforms. Analysis of the data, using sentiment analysis and topic modeling, found that audiences tended to discuss institutions' sports teams, as well as events or features unique to institutions. No difference in sentiment or emotion was found between account types, with most messages being moderately positive. Facebook messages tended to include more descriptive language while the higher volume of tweets suggested that Twitter audiences appear to be more prone to interaction than Facebook audiences. Institutions may be best served by pursuing different social media strategies for Twitter and Facebook.

Chapter 1 Introduction

Introduction

More than a decade after Time magazine named “You” person of the year in 2006 in response to the rise of user-generated content, social media remains tremendously popular. According to Ryerson’s social media lab in 2017 an overwhelming majority, a full 94%, of online Canadian adults had at least one social media account (Gruzd, Jacobson, Mai, & Dubois, 2018). Social media has become a mainstream tool amongst higher education institutions (Tess, 2013) and is now considered a standard part of doing business. Individual higher education institutions have had to develop their own communications and social media policies at the same time as social media has developed as a unique industry, with new platforms frequently launching, older platforms shutting down, and a constantly changing userbase. This has produced a wide variety of social media policies across different nations (Pasquini & Evangelopoulos, 2017), and this diversity has created an opportunity for academic inquiry. Further study of the intersection between higher education and social media may provide a way to further the development of best practices within the post-secondary sector, and eventually assist resource-constrained institutions with making the best use possible of their social media presence.

The current literature examining social media and higher education has primarily had two focuses; the marketing potential of social media for higher education (Bélanger, Bali, & Longden, 2014; Constantinides & Stagno, 2011) and how institutions have used social media thus far (Beverly, 2013; D. L. Linvill, McGee, & Hicks, 2012; Peruta & Shields, 2017; Shields, 2016; Veletsianos, Kimmons, Shaw, Pasquini, & Woodward, 2017). While institutions can use social media as a mechanism to distribute content to an audience, social media also allows that audience to communicate back to the institution. Content published by an institution has the potential can induce discussion, while the content itself often has some mechanism for response which provides the specific channel for the audience to respond and discuss (Safko, 2012). Depending on the specific social media platform used there is also limited ability for institutions to filter who can view their content, allowing a broad segment of the public to communicate with the institution. These communications have yet to be the subject of extensive inquiry.

Common approaches to researching social media use in the higher education sector have included the use of surveys and focus groups, with a focus on the authors' experience and intent rather than their published content, while many other published papers have made use of content analysis to focus on the content instead of the authors (Snelson, 2016). Due to the significant variation in institutions within the higher education sector (G. Jones, 2009) there may be a risk to examining a limited number of institutional social media accounts and applying the resulting conclusions to the broader sector, as the results may be the product of characteristics unique to the limited sample. At the same time, it is also entirely plausible that content may not differ between institutions, as many institutions may share similar roles in their local communities and share similar goals for social media use. Some social media platforms have permitted access to the data published onto their platforms to encourage application development, amongst other reasons, providing access to a previously unavailable trove of social interaction data. Access to this data is not universal to all platforms, however. As an example, as access to both Facebook's and Instagram's data has been sharply limited by Facebook partly in response to privacy concerns, and partly in response to scandals stemming from controversial usage of Facebook data including 2018's Cambridge Analytica scandal (Langone, 2018).

Unlike Facebook, Twitter was originally designed with the strongest focus on communication through text, rather than through photo sharing (Arceneaux & Dinu, 2018). By design Twitter limits the length of communication by imposing a limit on post length 280 characters, originally 140 characters. As a platform originally designed with the intent of facilitating communication through text, recent popularity of pictures, small animations, and other forms of memes notwithstanding, much of the content that can be retrieved from Twitter is text and any analysis of a large amount of this data requires some method for the analysis of large volumes of written language. The analytical tools most suited to the analysis of large volumes of text data, absent significant time to support detailed manual analysis, are computational content analysis tools such as natural language processing and sentiment analysis (Grimmer, Stewart, & Michael Alvarez, 2013).

Analysis of Twitter is valuable in the context of higher education for reasons beyond Twitter's ease of access. Large scale analysis of social media data has been useful as it has allowed for the use of both quantitative and qualitative analysis, as well as the identification of different patterns or subpopulations amongst those who tweet at institutions (Kimmons & Veletsianos, 2018). On the part of the institutions, Twitter has an extremely high rate of adoption amongst United States post-secondaries (Barnes & Lescault, 2013) with at least 2,400 institutions in the United States operating an institutional Twitter account (Kimmons, Veletsianos, & Woodward, 2016). Additionally, Twitter is used by approximately 60% of online Canadian adults between the ages of 18 and 24 (Gruzd et al., 2018), the age group most commonly associated with higher education students, though there is some evidence to suggest that Twitter has a lower rate of adoption amongst university students specifically (Adams, Raes, Montrieux, & Schellens, 2018). Twitter is a worthwhile social media platform to examine given the significant proportion of Twitter users that also use another social media platform (Smith & Anderson, 2018), most notably Facebook. Currently there is mixed evidence to suggest that content posted to different social media platforms exhibits differences in civility and tone (Oz, Zheng, Chen, & Park, 2018;

Rowe, 2015) and the messages sent to the higher education sector across different platforms have yet to be examined using a comparative approach.

This research examines the content targeted at Canadian post secondary institutions' social media accounts through an examination of the content targeted towards institutions via Twitter and Facebook. To date there has been limited research examining the audience for post-secondary institutions and how the audience responds to institutions on social media. Instead, the literature to date has focused on the institutional side of social media, examining what institutions choose to broadcast (D. Linvill, Mcgee, & Hicks, 2012; Peruta & Shields, 2017). In addition, the recent literature has been primarily small in scope, where only a handful of institutions have been examined at once. There have been only a few exceptions to this trend, notably works by Beverly, (2013) and Kimmons, Veletsianos, & Woodward (2017) which both examined the higher education sector broadly and found that generally content published by American institutions onto social media was primarily that of a news feed, not intended for interactivity with the audience. Finally, higher education in Canada has not been subject to significant examination (Clark & Norrie, 2013). This is despite the Canadian higher education sector being significantly decentralized (Jones, 2014) and having a history of quickly embracing digital technologies (Dumitrica, 2017). At the same time, the Canadian higher education sector is catering to an ever-changing student demographic (Anderson, 2015; Guo & Chase, 2011) while continuing to have substantial economic impacts on their communities (Litwin, 2010). Not only does this present the opportunity for a new perspective on a well-researched topic, it does so in the unique Canadian context.

This research expands on existing research from two distinct perspectives. First it provides context around previous research into the social media use of higher education institutions. By examining the content which is directed towards institutions this research can help characterize the environment in which institutions have operated their social media accounts. An improved understanding of this environment can provide a different perspective on previous findings. Institutions electing to operate in a non-interactive manner in a negative, hostile environment may suggest different priorities, choices, or strategies than institutions which have elected to be non-interactive in a friendly or positive environment. Second, this research applies a uses and gratifications approach to the audiences for higher education Twitter and Facebook accounts, and provides additional evidence to suggest whether each of these platforms are used for different reasons by comparing the audiences for Twitter and Facebook profiles of the same institution. Previous research (Dolan, Conduit, Fahy, & Goodman, 2016; Gan & Wang, 2015; Korhan, Metin Ersoy, & Ersoy, 2016; Quan-Haase, Martin, & Mccay-Peet, 2015; Smock, Ellison, Lampe, & Wohn, 2011) has focused on why individuals use one or multiple platforms. This research takes the reverse perspective, given a single organization's Facebook and Twitter accounts it instead focuses on how their audiences differ.

Beyond academic inquiry, this research has potential for application to higher education administration. In response to previous research, higher education administrators have expressed an interest in improving their understanding of their social media presence (Agostino & Arnaboldi, 2017). Institutions and their staff must perform a broad variety of functions while, being accountable to both students and provinces, including playing a role in the community, maintaining connections with alumni, and attempting to recruit new students. Social media is a tool which enables institutions to communicate with stakeholders in all these roles, and likewise is a channel through which these stakeholders can contact the institution. The main insight of systems thinking, regardless of specific definition, as applied to higher education is that nothing any one institutional business unit does happens in isolation (Arnold & Wade, 2015). Rather all aspects of an institution are connected, and the actions of one business area will impact another. Effective administration of higher education must now include an understanding of an institution's social media profiles, their audience, and how the institution is viewed by the different social media audiences. Institutions' social media audiences can use the available interactivity to relay back to institutions important messages such as community concerns, expressions of approval, interest, or notifying institutions that they have failed to communicate clearly. Through the messages sent back to institutions, the audience will be identifying the topics that are most of interest, and in the process will be providing insight to institution about what matters to their audience. As an example, messages focused on individual faculties or events may indicate what are seen to be the defining features of an institution, while frequent concerns or questions identify where institutional communications have been ineffective or simply absent. By examining so many institutions at once, especially from a post-secondary system which features a diversity of institutions, this research establishes a sort of baseline for Canadian institutions when examining their own social media audiences. This baseline may be important for understanding an institutions' own unique audience, given their own role and specialty within a larger higher education system.

Research Questions

The overarching research question is: What public content do Albertan institutions receive on social media? The two sub questions informing it are as follows:

> Question 1: Do different types of Albertan higher education institutions consistently receive different content through both Twitter and Facebook?

Post-secondary education is not homogenous, and significant variation exists between individual institutions. Within Canada there are post-secondary institutions with student populations numbering in the hundreds, and institutions with student populations larger than small provincial capitals, with equivalent variation in services, programming, and research activity. Presently, it is unknown whether these characteristics attract different audiences, or whether the difference across institutions relates the difference in interactions between an institution and their followers. Provincial post-secondary sectors tend to be structured such that institutions of similar size, scope, and mission share classification, and this variation can be exploited. By using these classifications to sort the post-secondary sector it is possible to examine how interactions vary between institutional types. This research question is focused on Twitter and Facebook is excluded due to a richer, more comprehensive Twitter dataset being available for examination.

Question 2: Do institutions receive similar or different type of content between their Twitter and Facebook pages?

Twitter and Facebook allow for different degrees of anonymity on the part of users and there is some evidence to suggest that neither platform enables their audience to be any more or less uncivil and impolite than the other when controlled for topic and prior beliefs (Oz et al., 2018). However, when these controls are removed, there is also evidence to suggest that users on Twitter are more likely to send more impolite messages (Oz et al., 2018). Presently it is unknown whether institutions receive different types of content between social media platforms, and in turn it is also unknown whether the use of these platforms requires some difference in content, policy, or strategy. To determine whether the content of one platform can be used to predict the content of another, the analysis of Twitter content is compared against a sample of Facebook content to determine what, if any, difference exists measured in terms of both sentiment and topic.

Definition of key terms

Due to the specific context in which this research operates, there is a moderate risk of confusion surrounding terminology. Frequently used terms have a slightly different meaning than often used elsewhere, and there are uncommon terms which may require clarification. In order to alleviate this risk critical definitions are provided below.

Text Mining

Text mining is the discovery and extraction of interesting, non-trivial knowledge from unstructured text (Tan, 1999). It is an applied form of computational linguistics and encompasses a variety of methods that allow users to extract information from unstructured human language and analyze it computationally. Text mining is not intended to extract meaning from text, rather it focuses on the extraction of statistical measures like word frequency, common patterns, and pattern likelihoods (Tan, 1999).

Natural language processing (NLP)

Copestake, (2004, p. 4) defines NLP as “the automatic (or semi-automatic) processing of human language.” Rather than attempting to teach machines to interpret text, NLP relies on thorough researcher-created lexicons to map characteristics onto words. By joining characteristics such as sentiment, or positive-negative measures NLP allows for the extract and measurement of meaning stored within text (Copestake, 2004).

Token

A unit of text, measuring one or more words. It is the unit which results from the parsing of unstructured text into smaller pieces to allow for the computational analysis of the text content. Most commonly tokens will be unigrams (single words) or bigrams (pairs of words). The nomenclature “n-gram” is used to describe cases where tokens measure “n” words (Silge & Robinson, 2017). Likewise, the verb “tokenize” refers to the process of dividing bodies of text into tokens. This paper uses unigrams, single words, almost exclusively.

Interaction

For the purposes of this research, “interactions” will be used interchangeably with both “responses” and “replies” and will encompass tweets which are replies to an institution's post, and tweets which "@" or mention an institution without prompting. These are tweets published that specifically identify an institution such that it will be visible to whoever operates the institution’s Twitter account. This does not include “likes” (an option which enables the audience to demonstrate approval of the content) nor does it include “retweets” (a republishing of the same content by an audience member to their own followers.) Instead it will very specifically refer to “replies” and “mentions.” Additionally, this includes messages posted to an institution’s Facebook page. Though Facebook posts can also receive “likes” these were not recorded in the data, and instead only text replies to institutional posts made on their own Facebook pages were recorded. For both Twitter and Facebook “interactions” will be limited to the text content targeted towards institutions. These definitions are in keeping with the standard observed across the literature and are consistent with the definitions used by Kimmons, Veletsianos, and Woodward (2016).

Theoretical Framework

Before examining “how” users interact with post-secondary institutions through Twitter and Facebook, consideration must be given to the “why.” This research was guided by the uses and gratifications theoretical framework. Twitter, Facebook, and social media in general are only another, newer medium for mass communication. Uses and gratifications theory is focused on the central questions of “What do people do with mass communication?” (Klapper, 1963) and what audiences derive from their use of mass communication. Audiences select their mass communication media consumption in order to match their pre-existing preferences and obtain some form of satisfaction from this media (Katz, 1959). Uses and gratifications was identified early as a mass communications theory which was well suited for explaining internet use behaviors due to the internet’s media-like properties (Lin, 1999; Weiser, 2001), specifically that it provides an opportunity for global communication with minimal restrictions stemming from geography and timing. An idea, debate, conversation, or story can quickly be communicated with very few restrictions on the size of the audience, and that audience can quickly return for more should it match their preferences. Uses and gratifications has been previously been used as a framework to examine the gratifications which drove internet use (Stafford, Stafford & Schkade, 2004). Unlike television and radio results, Stafford, Stafford, and Schkade (2004) found that three distinct gratifications could be found for the use of the internet; process gratifications, gratification derived from the use of the media itself, content gratifications, enjoyment derived from the content consumed through mass media, and social gratifications, satisfaction from social interactions conducted through the internet. Further research would suggest that process gratifications were better considered hedonic gratifications (Van Der Heijden, 2004) as they are gratifications most often gained through the use of the internet for strictly enjoyment purposes.

Social media as a specific area of internet use has also been examined through a uses and gratifications framework by Anita Whiting and Williams (2013) who, through closer examination of social media users, further divided the three themes overarching gratifications identified in Stafford, Stafford, and Schkade (2004). Social gratifications were parsed into social interactions, expression of opinions, communicatory utility, and information sharing (Whiting & Williams, 2013). Hedonic gratifications were expanded to include passing time, convenience, and relaxation, while hedonic gratification was split into knowledge about others, information seeking, and entertainment (Whiting & Williams, 2013). Korhan and Ersoy (2016) applied a uses and gratifications framework to examine the motivations for choosing any one specific social networking platform over competitors, finding that each platform had characteristics which leant itself more towards fulfilling different gratifications. Finally, aside from application to social media as a broad category, uses and gratifications has been applied to specific social media platforms. Twitter and the microblogging platform WeChat were analyzed from a uses and gratifications perspective by Gan and Wang (2015). Very similar themes were found, with content gratification being divided into information seeking and information, social gratification being divided into social interaction and social hedonic gratification hedonic gratification being divided into entertainment and simply passing time networking (Gan & Wang 2015, p. 359). Uses and gratifications has also been applied at a feature level, suggesting that users seek different gratifications from using different features in different contexts (Smock et al., 2011).

This theoretical background is well suited to the examination of this topic, due in part to the body of literature which has applied this framework to the examination of social media already. An important detail to remember is that, much like a TV channel or radio station, social media users must choose which accounts they will follow and where they will post their content. Uses and gratifications suggests that the examination of the content published towards institutional Twitter accounts and Facebook will identify the content and topics which are most of interest to the audience, that the topics which are at the centre of the most discussion are the most discussed because they are of the most interest. Understanding that users tailor their social media use, including both what they observe and which features they use (Smock et al., 2011), a uses and gratifications framework can be used to examine what has brought users to an institution's social media content. An overabundance of discussions focused on entry requirements may indicate that the audience expects the accounts to serve some form of referral function. Likewise, seemingly inane chatter about campus landmarks would suggest that the audience's chief gratification is in the interactive component of social media. Put into layman's terms, examining the discussions around institutions online should identify what is most of interest to the audience, and may provide insight into why they are interacting with an institution's Twitter and Facebook pages.

Chapter 2 Literature Review

The purpose of social media use

Previous research into the uses of social media from an organizational perspective provides some background for understanding the higher education sector's social media use. Kaplan and Haenlein (2010) were among the first to examine how social media could be used as a tool for business, defining social media as "a group of Internet-based applications that…allow the creation and exchange of user generated content" (p. 61). The authors cautioned against attempting to maintain a presence on as many platforms as possible, while also suggesting that firms that wish to make use of multiple platforms ensure that there is alignment in content strategy across all platforms. Most importantly the authors implored firms to be active, be interesting, and to recognize that social media is a powerful public relations tool but also one that may produce profoundly negative results if misused. Safko (2012) specifically examined Twitter and other extremely short form blogging platforms, noting "the character limitations on microblogging force us to communicate in a more succinct manner" (p. 289). These articles suggest that from the perspective of the institution social media, and more specifically Twitter, is a tool which allows them to engage with their audiences in an interactive manner. This interactivity provides a means to better understand what is of most interest to an audience and the broader community around the institution. For almost any enterprise of any size there is a strong incentive maintain a positive relationship with the local community and other stakeholders through effective public relations (Wells & Spinks, 1999). Social media provides an avenue to maintain these relationships, however the openness provided by social media can produce tension with efforts to control or guide discussions (Macnamara & Zerfass, 2012). Despite this, firms are not powerless in guiding discussions (Dolan et al., 2016). Even in applications strictly focused on the public good there are challenges with promoting discussion using social media including challenges with guaranteeing the reliability of information, protecting confidentiality of discussions participants, and monitoring the discussion to ensure quality (Moorhead et al., 2013). From the perspective of higher education, social media can be effective as a communications tool, it also comes with many challenges.

As an interactive medium, the audience for social media is also of interest. Highfield, Harrington, and Bruns (2013) examined the use of Twitter by non-business users, examining whether users were using Twitter as an outlet to broadcast a public performance of belonging to a specific audience or group. Their findings suggest that Twitter is used as a way for users to express belonging to chosen audiences, a channel to demonstrate membership to a group. As part of this demonstration users will alter their language use and behaviour in order to better suit the community or group they are addressing (Marwick & boyd, 2011). Essentially users will try to appeal to the audience they imagine to be reading their online content. It is worth noting that Twitter use is not homogenous, and that a single user can simultaneously pursue multiple uses through a single account (Quan-Haase et al., 2015). The sentiment expressed in tweets by non-business users was examined by Thelwall, Buckley, and Paltoglou (2011) where the authors used sentiment analysis to examine the sentiment of Twitter content posted around popular events. There was strong evidence that negative sentiment tended to correlate with popular events while positive sentiments were less likely to line up with events. The increase in negative sentiment around events of interest suggests that tweets directed to institutions following major events may include a high proportion of messages expressing negative sentiment. This suggests that the content posted to Twitter can be influenced by external events, that newsworthy items will induce an increase in Twitter activity.

The increase in negative sentiment around events is of particular note, given that negative messages are more likely to be frequently reshared (Tsugawa & Ohsaki, 2017). It is worth noting, however, that this resharing behaviour may not be tied strictly to negative sentiment but rather to content which is interpreted as emotionally charged (Stieglitz & Dang-Xuan, 2013). A possible explanation for the resharing of negative may be a product of those who are using social media the most, with some evidence suggesting that they may be younger, extroverted, and more inclined towards negative affectivity (Correa, Hinsley, & de Zúñiga, 2010). Despite this, the resharing of content is one of the most valuable goals of social media use from a marketing perspective (Wolny & Mueller, 2013). This leaves those operating social media accounts in a tricky position, as they need to balance their goals against what will be reshared for the most, without compromising the perceived values of their brand. Further complicating this is that much social media marketing is simply communicating promotions to existing customers, rather than content which encourages lasting engagement with their audience (Schultz & Peltier, 2013).

Social media in higher education

Social media has impacted how the higher education sector has integrated the Internet (Wheeler, 2009) and can alter the dynamics in student-administration interactions (Schroeder, Minocha, & Schneider, 2010). As a result, institutions have been able to pursue institutional goals using social media. Students that follow their institutions on social media have tended towards rating their relationship with institutions as being higher quality (M. Clark, Fine, & Scheuer, 2017) while students active on social media have a more positive view of their relationship with their fellow students and tend to perceive their educational experience to be of higher quality (Rutherford, 2010). Many students seem to believe that institutions and professors that are active on social media better understand their students, and generally consider social media to have a positive impact on their education (Neier &

Zayer, 2015). Given the tendency of those most active on social media towards negativity (Correa et al., 2010) institutions with an interest in promoting a positive social media space to achieve the aforementioned benefits would be served to publish statements emphasizing the expectation of civility online, to teach critical evaluation of online material, and to monitor online communications on their own pages to ensure they do not devolve (Junco & Chickering, 2010). Institutions which have achieved benefits from social media may have done so in spite of the perception that many institutions' cultures leads them to be resistant to adopting new internet technologies (Radovanović, Hogan, & Lalić, 2015).

Many institutions have demonstrated a tendency to operate social media their social media presence as a one-way communication tool and to limit direct interaction (Davis, Deil-Amen, Rios-Aguilar, & González Canché, 2015). In turn this may contribute to the perception that institutions are reluctant to adopt new technologies (Radovanović et al., 2015). Operating social media as a one-way communication channel may simply may be a resourcing problem, as many surveyed institutions have reported insufficient funds and insufficient perceived value for expanded interactivity as the primary barriers to expanding social media use (Davis et al., 2015). These barriers may not be as insurmountable as first expected, as the findings of Rutter, Roper, and Lettice (2016) indicate that institutions can strengthen their brand and improve their recruitment through publishing content which encourages interactivity. The volume of published content appears to have little benefit for institutions, and they would instead be better served by publishing social media content which encourages interaction, and which the audience is likely to reshare on their own (Rutter et al., 2016). Research findings, such as those of France, Finney, and Swerdzewski (2010), suggest that this engaging content may be as simple as the promotion of institutional sports teams, encouraging the involvement of students in institutional leadership, or other content which helps strengthen feelings of attachment to the institution.

The Canadian context seems to mirror other higher education sectors in a variety of ways. Examination of Canadian institutions suggests that higher education in Canada also primarily uses social media as a news feed (Bélanger, Bali, & Longden, 2014). Unlike American institutions (Davis et al., 2015) there was some evidence to suggest that Canadian institutions will infrequently publish content designed to encourage interaction between the institutions and students (Bélanger et al., 2014). Veletsianos, Kimmons, Shaw, Pasquini, and Woodward (2017) further examined Canadian institutions' use of Twitter to determine the factors which led to the most impactful content. Canadian institutions were found to be more likely than expected to make use of hashtags and to link to external websites, differing from results gathered from other nations. Unsurprisingly, institutional accounts also tended to present an overwhelmingly positive, and thus incomplete, portrayal of the institution. As part of a broader study Pasquini and Evangelopoulos (2016) noted that when Canadian social media policy documents were compared against foreign peers there was a greater emphasis on information management, perhaps resulting in institutions being more willing, or even more aggressive, to correct misinformation. These findings suggest that Canadian institutions may be willing to encourage interaction with their institutional social media presence, a departure from the findings of previous literature (Beverly, 2013; Davis et al., 2015; Kimmons et al., 2016; D. L. Linvill et al., 2012).

Twitter and Facebook in higher education

Twitter and Facebook specifically, as two of the most popular social media platforms active today, have been studied specifically in the context of the higher education sector. Multiple articles (e.g., Dunlap & Lowenthal, 2009; Gao, Luo, & Zhang, 2012; Schroeder et al., 2010) have confirmed that post-secondary students are well-versed in the use of Twitter as a platform, are willing to use it to investigate institutions and communicate with administrations, and use it to interact with other students on campus. Twitter's role in students' decision-making during selection of school and program of study was specifically examined by Constantinides and Stagno (2011) where the authors suggested that social media is relatively inconsequential in influencing student study choices, in contrast to later, more generalized, findings (Rutter et al., 2016). Though this finding would seem to diminish the importance to Twitter it is important to remember that this paper examined students entering university, not active post secondary students. Put another way, the finding merely suggests that students are unlikely to attend an institution because they like their Twitter account, but also still use Twitter for information purposes. Students' decision to access social media for both information and interactive purposes can be contrasted against the findings of Beverly (2013) and Linvill, McGee, and Hicks (2012) which suggest that institutions are only contributing to one of those functions. Both papers' researchers found that institutional Twitter accounts were akin to institutional news feeds, with little intent of their social media presence being used as a two-way street for communication. The content published by institutions also tends to paint the experience of staff and students as exclusively positive (Veletsianos et al., 2017). Despite higher education's limited social media engagement there is some evidence to suggest that larger schools tend to generate more content, while smaller schools generated less content that was more likely to generate interaction (Peruta & Shields, 2017).

There is a financial incentive for institutions to maintain a healthy and engaged social media presence. Twitter's frequent use as a mechanism to positively highlight an institution matches nicely with research which suggests that institutions use social media to strengthen their brand (Nevzat, Amca, Tanova, & Amca, 2016). A strong social media presence may foster a sense of involvement and trust among students, encouraging increased loyalty towards the university's brand. This loyalty is particularly valuable to institutions when the findings of McAlexander & F. Koenig (2012) are considered. From the perspective that universities are marketing institutions which offer a broad array of services, the McAlexander and Koenig (2012) argued that the users of the brand (staff, students, alumni) that felt most strongly connected to the institution and to fellow alumni were most likely to donate to the university.

Facebook has also been of use for advancing institutional aims. There is evidence to suggest that Facebook as a platform can used to encourage more interaction between professors and students, can also contribute to increased collaboration between students, and can improve student awareness of campus events (Chugh & Ruhi, 2018). This is vital for institutional goals which require campus participation or simple awareness, as they can be supported through using Facebook to disseminate information (Hamid, Ijab, Sulaiman, Md. Anwar, & Norman, 2017). Likewise, similar to the findings related to social media as a whole (M. Clark et al., 2017) institutions which are interested in strengthening their brand and promoting engagement would be well served by using Facebook to pursue these goals (Brech, Messer, Vander Schee, Rauschnabel, & Ivens, 2017). These same institutions may find the most success by publishing content which they have found to promote engagement from the audience, rather than simply pursuing a larger audience (Arriaga, Andreu Domingo, & Berlanga Silvente, 2017; Brech et al., 2017). Despite similarities in application, institutions seeking to use Facebook and Twitter during what may be perceives a crisis would be well served by understanding differences between the two platforms. Social media is not homogenous, and there is evidence to suggest that Facebook may be more useful to engage in dialogue and deliver crisis messages, while Twitter may be better used as an early warning tool (Eriksson & Olsson, 2016). Despite the different applications, both platforms must be monitored during a crisis (Eriksson & Olsson, 2016).

Relation to this study

One of the defining features of effective social media use is the high level of interactivity it permits, and this has underscored how businesses are typically encouraged to use social media (Kaplan & Haenlein, 2010; Safko, 2012). Institutions have sought to use Twitter to develop their brand (Bélanger et al., 2014; Nevzat, Amca, Tanova, & Amca, 2016), using social media proactively in order to further institutional goals and to control their own image. However, at the same time institutions' other most common social media use was to publish news without the intent of receiving communication from their audience. These two uses are initiated by the institutions themselves, and neither approach requires exchanging communication with an external audience despite social media's readily available interactive capabilities. As of the writing of this research paper there has been very little examination of the content published by the audience which is targeted towards institutions. While some papers (Bélanger et al., 2014; Peruta & Shields, 2017) have indirectly examined the audience there has yet to be any research squarely focused on the communication that the audience directs back to institutions. This is the gap which this research paper seeks to fill by specifically examining the content which is sent to the higher education institutions using Facebook and Twitter. Given that the weight of the evidence suggests that the content published by institutions does not promote interaction from the audience, it would seem curious that audience would still interact through this medium. If post-secondary institutions published content does not promote dialogue, then what is the audience sending to institutions? Knowing that their audience is returning messages to the institution, should the institution use social media differently?

Chapter 3 Methodology

Methodology

This research employs a case study methodology. The purpose of the proposed research is to examine, thoroughly, examples of the communication received by institutions through their Twitter and Facebook pages. Each of the publicly funded post-secondary institutions in Alberta, Canada are included in this paper's analysis. These institutions were selected as they represent an entire provincial post-secondary system, the third largest english language system in Canada, while containing a small enough number of institutions for the analysis to be manageable. Additionally, this sector includes a wide diversity of institutional types, from small regional colleges to two members of Canada's U15 group of research universities. By collecting Twitter messages and Facebook posts that were sent to each of these institutions this research examines the topic and sentiment of online interactions directed toward higher education social media presence. This approach is appropriate given that the focus is on the content that is published to Twitter and Facebook, and not the process or decisions which generate the content. Put another way, there is no interest in some form of input to output process, rather this research is exclusively concerned with the output. Given that this research is instead examining a sample from a specific period, a case study methodology was the most appropriate avenue to pursue this research topic.

Methods

As the meaning behind the written language must be imputed, and the interpretation of written language can vary substantially based on the researcher and context, the assignment of sentiment and emotion to the text is a strictly qualitative endeavour. Despite the use of qualitative methods to interpret the text, the aggregation and statistical analysis necessary for examining large volumes of data within a reasonable timeframe is quantitative. This research requires both in equal measure, and as such mixed methods are the only appropriate methods.

Examining the topics mentioned most frequently in responses to institutions and messages which mention the institution will identify the topics that are of most interest to institutions' audiences, and likewise the use of sentiment analysis will measure the tone of interactions. The topic and sentiment of text is examined to simplify the complexities of online interactions and the motivations for publishing content online. While the ideal approach would to be examine each gathered data point and perform a detailed content analysis, or alternatively to interview subjects that participated in online discussion either on institutions' Facebook pages or that mentions institutions on Twitter this becomes increasingly time consuming as the gathered data grows. Due to limitations with regards to budget, time, and scope there was no opportunity to interview subjects who publish content on Twitter and Facebook, the ethical concerns of tracking down individual users notwithstanding. In order to examine the content published online at a appropriate scale as to apply to the broader post-secondary sector it was time-prohibitive to perform a detail manual content analysis. Natural language processing and sentiment allow for the examination of the tone of conversation across different groups in a measured way, providing a consistent method to compare different groups of content. Once tone of conversation has been established, topic modelling provides a way to examine the topics most discussed amongst the same groups, again providing a consistent method to compare the topics most frequently mentioned between groups. The advantage to this approach is that it provides consistent measurement of text content along two different dimensions, and that this approach scales up to almost any size of dataset without issue.

Data Collection. There are 26 publicly funded post-secondary institutions in Alberta, listed in Appendix A. The Banff Centre was excluded from this analysis as their offered academic programming does not follow a traditional academic calendar, nor does the Banff Centre offer academic credentials in the same way as other institutions. This left 25 institutions whose Twitter accounts were examined, and each institution was sorted into one of six institutional classifications according to the organization of Alberta's post-secondary sector, published by the government of Alberta. Each institution's classification represents their offered programming, role in the Albertan post-secondary sector, and a size relative to other institutions. Institutions that offer graduate degrees and a conduct research as a part of their core institutional missions are included in the Comprehensive Academic and Research Institutions (CARI's) classification, and generally are among institutions the largest institutions in the province. Institutions which offer bachelor's degrees as their highest credential, focused on undergraduate education with research activities representing a secondary goal are combined into the Baccalaureate and Applied Studies Institutions (BASI) classification. Polytechnics are institutions focused on providing skills focused education with the intend of graduates being prepared to enter apprenticeships or other skills focused careers. Comprehensive Community Institutions (CCI's) are smaller institutions, often located in more rural communities, offering diplomas and certificates specifically with the intent of responding to local labour needs and providing an entry point for students that wish to transfer elsewhere in the post-secondary sector. Finally, Independent Academic Institutions (IAI's) are typically the smallest institutions, and offer a mix of bachelor's degrees, diplomas, certificates, and adult upgrading. IAI's were originally founded as religiously affiliated colleges, and their religious influence continues today. The primary twitter account for each institution was identified through the Twitter link present on the homepage of each institution's website. These accounts were chosen based on the logic that, by publishing these accounts on their websites, institutions are signaling which accounts they desire the audience to follow. By identifying which accounts are officially used to represent the entire institution, institutions are funneling their viewers towards the "correct" account. Secondary institutional accounts were identified through Twitter's search function, and the first three institutionally affiliated accounts were included in the analysis. To retain focus on the administrative component of the higher education sector student unions were excluded. Institutional presidents were included as valid accounts due in part to a common Twitter account naming convention of "[Institution]Prez" as well as their significant role as the head of the administration. The choice of three additional accounts is wholly arbitrary and was selected to provide a maximum of 100 Twitter accounts for examination, and to provide an equal maximum number of accounts for each institution included. Mentions of these accounts were gathered through use of the R package twitteR (Gentry, 2015) and a developer account's access keys to gather tweets through Twitter's published API. A list of search terms, in this case the usernames of institutional twitter accounts, was fed to the searchTwitter function with a declared maximum of 3000 tweets per account that could be gathered each time the query was run. This provided a maximum upper threshold of 300,000 tweets that can be gathered at once. 3000 tweets per account, per query, was selected as it is near the upper threshold of tweets that can be collected by the twitteR package in a single query as described in the twitter documentation (Gentry, 2015).

The original intent had been to collect as many tweets as possible, regardless of how broad of a time period this would cover. Due to limitations on how the standard Twitter search API operates the tweets gathered during each query run were never older than one week (Twitter, 2018). To compensate for the limited time period that could be gathered through the Twitter API the data collection plan was altered, and the data gathering script was re-run in one-week intervals for ten weeks. Each new collection of tweets was then be appended onto the previous sample using the R function "left_join," as this prevented tweets which had been gathered multiple times, due to overlapping hours, from appearing as duplicates in the dataset. Subsequently, once the ten-week data collection period was complete the finalized dataset was saved to an .rds file, a memory-efficient file format which copies how R stores objects in memory, to ensure that a fixed sample was used during analysis. Early testing of the data gathering process for a preliminary list of accounts had returned 11,000 tweets. It was assumed that the test sample was representative of the volume of tweets that would be sent to post-secondary accounts during the ten-week window, the final dataset was expected to comprise of approximately 110,000 tweets. Early manual examination of the data indicated that, for reasons currently unknown, the twitter query returned a substantial number of tweets that were unrelated to the desired subject matter. Irrelevant content was screened out using two steps. First string matching was used to ensure that one of the search terms was present in each tweet. Second, in the tweets where one of the search terms was not present, tweets were then filtered to include only tweets which were replies to one of the accounts used as a search term. This ensured that the inclusion of every tweet in the data set could be explained by identifying the relevant terms. Unfortunately, the unrelated tweets had inflated the original expected tweet count and the real number of tweets directed to the selected accounts was significantly lower than expected. To offset this unexpected drop the collection period was extended to twelve weeks.

Content was gathered from Facebook for a similar twelve-week interval, overlapping with the Twitter data gathering process. The primary Facebook page of each institution was located by following the Facebook link on each institution’s main webpage. Comments made on the institution’s Facebook page were then be transcribed manually to an excel file. The only comments which were transcribed were comments which were visible when viewing the institutions’ Facebook pages without having logged in to Facebook and as such every comment transcribed was publicly available. Individual comment authors' names were omitted, as well as any full name mentions of other users in order to preserve the privacy of commenters whose posts were recorded. If a post solely consisted of a mention of another user, the post was omitted from the data collection. Posts by the institution which did not have any comments were not recorded. Posts which were not replies were contained in a different page, marked as "Visitor Posts", and these were excluded due to not being present on every institutional page, their small number, and the extra step required to find them. In order to preserve a record of which comments were replies to which posts, posts made by the institutions themselves were recorded. Additionally, the institutional page, the institutional classification, and the date of each post was recorded.

Content Analysis. Similar to the perspective taken by Kimmons, Veletsianos, & Woodward (2016) there is significant value in examining the content of twitter traffic around the higher education sector, including any differences between messages sent to specific groups of institutions or timing of the messages. Different institutions may have different audiences, and these different audiences may be publishing different content towards these institutions. These differences may include differences in sentiment, and to estimate the sentiment of the collected tweets two lexicons were used: Saif Mohammad's nrc lexicon (2011) and the AFINN lexicon (Nielsen, 2011). The nrc lexicon was used to map eight different emotions onto the words which were extracted from the collected tweets, dividing each tweet into a collection of expressed emotions. This provided an overview of whether the published content was sad, anticipatory, or a myriad of other options. As the nrc lexicon was developed specifically through manual annotation on a crowdsourcing platform, was developed and tested for general application across a large variety of contexts, including post-secondary education, and has been thoroughly applied by Canada's National Research Council, manual confirmation seemed unnecessary.

The AFINN lexicon (Nielsen, 2011) was used to determine if the content of the examined tweets was positive, negative, or neutral in sentiment. The AFINN lexicon was selected over other lexicons due its' ranking of words' strength of sentiment, allowing for strongly negative or strongly positive content. This provided a more nuanced measurement of the mood present across the total assembled tweets. However, the AFINN lexicon had one notable limitation, in that words deemed to have strictly neutral sentiment are excluded from the lexicon, providing a scale which ranges between -4 and 4 with no 0 value. Comparing the results of these two lexicons allowed for a more comprehensive summary of what content audiences chose to send to institutions. In order to ensure that the AFINN lexicon was consistent and appropriate for this context, a sub-sample of random words were manually coded and compared against the results from the AFINN lexicon using Cronbach's alpha (Tavakol & Dennick, 2011). This was necessary due to some words such as "professor" or "university" being associated with positive sentiment in the AFINN lexicon (Nielsen, 2011) that in this context should have been omitted due to a neutral sentiment. Though Cronbach's alpha has well-examined limitations, and can be biased through the use of a large sample (Schmitt, 1996), a comparison between only two scales was a simple enough application to avoid many of Cronbach's alpha's challenges.

The AFINN lexicon was filtered down to terms that had been sent to institutions, producing a list of 832 unique words. The associated scores were removed, and then a random sample of 200 words was written out to a .csv file. This list was opened with Microsoft Excel, and each word was manually scored between -4 and 4 representing an estimation of how often and how strongly each word should represent positive or negative sentiment. With estimated scores attached, the file was read back into R and joined against the AFINN lexicon and the manual scores were compared against the original AFINN scores using Cronbach's alpha. The resulting raw alpha value of 0.955 suggests that the manual scoring was consistent with the scoring included with the lexicon. As a redundancy check, as well as to guard against any possible bias or priming before manually scoring, the manual scoring was repeated using only scores of 1 or -1. Cronbach's alpha was again calculated on this flattened scoring, resulting in a raw alpha value of 0.823. These suggest that even in the instance that the magnitude of the positive and negative scoring is inappropriate, the AFINN lexicon remains sufficiently consistent with manual classification for use on this data set.

Topic Modeling. Beyond the sentiment and tone of the tweets, the topics that drive the interaction from the audience have significant value. The topic modeling procedure for sifting through large amounts of text data was largely informed by Silge & Robinson (2017) while the inspiration for the examination of content in Twitter messages came from Robinson (2016). Differences between different institutions' audiences were examined by assigning institutions to different groups based on the research questions driving this paper. Tweets were divided based on the classification of the institution mentioned, and the specificity of the institutional account mentioned. These grouping variables were attached in order to examine differences in word frequencies and topics. Latent Dirichlet allocation (LDA) was used to determine the difference in words used between different groups. LDA is an algorithm which estimates the words that make up a topic, while also determining the proportion of each topic that composes a document (Silge & Robinson, 2017). Between sentiment analysis and topic modeling, the assembled data paints a comprehensive picture of the messages directed towards post-secondary institutions in Alberta, their tone, and their content.

To guard against the possibility that a few terms were always common across all institutions, which may mask the differences that exist between institutions, log-odds ratios were also constructed. Log-odds ratios are useful as they rank terms based on the difference in likelihood of that word appearing in one of two groups. The ranking highlighted the differences in word usage between these groups (Silge & Robinson, 2017). Comparing the different words used in tweets directed towards institutions allows insight into how audiences are choosing to interact with institutions via Twitter, and more specifically they assist with magnifying the topic differences across groups. This process essentially wipes away terms that are used at an equal frequency by both groups, and instead highlights terms which are used proportionally more often.

Words are not used in a vacuum, and the context of word usage can be used as an error check on terms identified using LDA and log-odds ratios. The association between different words can be important to understand any document in its entirety, a recommendation which comes directly from Stemler (2001). As an example, one institution may receive many tweets about recent research into the impacts of social housing, and many tweets about their student union. If context is ignored it will appear that this institution receives the same tweet as another institution which receives many tweets about "student housing." Despite the use of the same language, the two institutions have received very different messages. To account for the importance of context and association a pairwise correlation was constructed using the 50 most messaged words for each institution type. Only the 50 most messaged words were used, rather than the full list of hundreds of words, in part to eliminate words which were used very infrequently but also to reduce to full list down to a manageable size. By treating each tweet as its own section, it was possible to apply the pairwise correlation function included in the widyr package (Robinson, 2018) to calculate a Pearson correlation to estimate the association between individual terms.

Once the analysis of Twitter was complete, the results were compared against the posts made on Facebook. The operations performed on the Twitter data were repeated on data collected from Facebook. Word ranks were assembled and matched against lexicons, log-odds ratios constructed, and an LDA model assembled. These were then compared against the sentiment expressed on Twitter, with an eye towards overlaps and differences. Comparing the tone of posts as well as the topics that prompt response illustrated whether institutions received different content depending on the social media platform in use.

Limitations and Delimitations

Each of the post-secondary institutions identified as part of Alberta's publicly funded higher education system were included in the analysis, with the exclusion of the Banff Centre. Smaller institutions often do not operate multiple social media accounts and accordingly there was fewer than 100 accounts examined. This also led to an unequal distribution of the number of mentions examined for each institution. This skewed the data to include more data points from larger institutions and in turn caution is urged before the results are generalized to smaller institutions. It is also worth noting that the collected data does not include original tweets published by the institutional accounts themselves. This is an area which has already been subject to extensive examination and does not require revisiting. However, it may be the case that more content is directed towards institutions which are perceived to be more active online. The data collection used in this study does not allow for examination of this possibility. Another limitation to this study is that it is currently impossible to extract private messages between users. These messages represent a form of communication via Twitter that may be markedly different in content and sentiment than publicly viewable messages. Another notable weakness in the proposed approach of extracting tweets from Twitter based on their inclusion of search terms lies in the twitteR package. One of the challenges of gathering data for this research is a known limitation within the package which prevents it from collecting all available tweets that meet a specified criterion. This limitation stems from Twitter not indexing tweets older than a week, and as such the available tweets were severely limited.

By far the largest weakness to this approach in the inability to gather a similar volume of data from Facebook as from Twitter. Facebook limits access through their API based on the requirements of app developers and will not authorize access to public page content without receiving proof that the access is required for an app to function properly. Ideally a similar or equal volume of data would be gathered from both Facebook and Twitter, in order to ensure comparability. The data that was gathered from Facebook was too scarce for a computational approach to be wholly appropriate, and as such a more manual process was used to assist in determining differences and similarities in content.

Ethical Considerations

Despite the examined content being in the public domain, this analysis inhabits an ethical grey area as users do not know they are being observed (Kimmons & Veletsianos, 2018). However, safeguards were put in place in order to protect individual users from being identified, and to prevent identification through their published content. To guard against the possibility that an individual user could be singled out or highlighted, the analysis was only performed in aggregate; no single user was ever examined on their own. Despite this research having a mixed methods approach there were not any interviews performed, and there were no interactions between this paper’s author and anyone whose published content was included in the data for this study. As such there was no chance for any harm to anyone whose content was examined. Twitter usernames, though gathered in the data collection process, were only used for one component of the analysis. In order to provide an extra layer of protection, within the dataset Facebook users were anonymized through simply excluding the authors' identities. As well, it is worth remembering that the data collected is available to anyone that has an internet connection and an abundance of patience. In the interests of protecting the privacy of the users whose posts were collected, and to prevent them from being identified, the analysis only included aggregate information to eliminate the possibility of singling out specific posts and searching for the specific wording. Finally, the collected data was deleted upon this thesis receiving committee approval.

There was one final potential ethical challenge. At the time of writing the author of this paper was employed by Mount Royal University, one of the institutions which was examined. While the operation of Mount Royal's social media was, and remains, completely divorced from the author's position it is reasonable to interpret research into one's own employer as improper. However, no single institution was examined on its own. The finest level of analysis was by institutional classification. By operating only with aggregated data there was no opportunity for content related to Mount Royal to be treated any differently from other institutions. Isolating and examining a single institution was also counterproductive to this paper's research goals, as it would severely limit the amount of data being examined at any one time. In order to ensure integrity and impartiality of this research the R code used to gather the data and perform the analysis is available on Github, at github.com/spudtopia/thesiscode and shows exactly which processes were used to generate results. Due to ethical concerns, as well as Twitter's terms of use, and Facebook's terms of use, the data used in this paper cannot be included. However, examining the code enables other academics to review the process and to satisfy their own concerns regarding the impartiality of the analysis.

Chapter 4 Results and Analysis

Data Cleaning

Once the data collection was completed the assembled primary and sub-account data files were appended, with identifiers attached to note whether each tweet had mentioned as primary account or a secondary account. String matching was used to determine which account had been mentioned in the tweet, and in turn this matching was used to assign both institutional classifications as well as noting whether the mentioned account was a primary or secondary account. For ease of use, the date that the tweet was posted was truncated to remove the time of day. The previously developed string-matched mentions were used to screen out irrelevant tweets. Messages which were replies to one of the accounts of interest were retained. Further examination of the collected data noted that a substantial proportion of the collected tweets were retweets, repeating another account's original content. Retweets allow for additional messaging to be included alongside the original content, with the new content being separated from the original content by the text "RT." Regular expressions were used to extract only text which preceded "RT," and to remove all the repeated text which followed. Tweets which were reduced to empty strings were then deleted. Individual tweets were broken into individual words using the "unnest_tokens," function from the tidytext R package. Stop words were filtered out, as were mentions of the accounts who were used as search terms to assemble the data, and strings composed only of numbers. In order to remove links to external pages the strings "https" and "t.co" were also filtered out of the list of terms.

Descriptive statistics

The Twitter data gathering process originally returned 90,355 tweets, all published between November 1st 2018 and January 21st 2019. Upon removing the erroneously gathered tweets, and removing the many replicated retweets, this figure drops to unique 16,289 tweets, with 12,925 unique text entries. Replies made up 7,405 of the collected tweets, while after cleaning the data set retweets made up only 561 of the assembled tweets. Primary accounts were mentioned 10,481 times while secondary accounts were mentioned only 5,808 times. Only 27 of these tweets were published by one of the accounts examined. A breakdown was assembled to show the comparison between how many times primary accounts were mentioned against how many secondary accounts were mentioned, including a ratio of primary accounts vs secondary accounts, attached in table 1. The time path of Twitter activity is displayed in figure 1. As can be seen, institutional twitter accounts were mentioned consistently from the start of November until the start of December. As classes end at the beginning of December and students enter finals period, and subsequently Christmas itself, the mentions of institutional Twitter accounts plummet to nothing before rebounding during the second week of January. This is approximately when all institutions will have begun their Winter academic terms, and classes have begun again.

The data cleaning process produced a list of 19,950 unique words that had been used in the collected tweets. When split to compare account types primary accounts received 16,601 unique terms while secondary accounts received only 6,826 unique terms, indicating a broader variety of discussion amongst primary account messengers. These unique words were then split according to institutional classification, showing the number of unique words directed towards each institutional classification. This breakdown is attached in table 2. It is no surprise that the broadest variety of language was directed towards the Comprehensive Academic Research Institutions as they make up the largest student bodies, they are the most funded institutions, they are the institutions with the most fundraising activity and research partnerships, they offer the most diverse set of academic programming, and generally dominate media attention as it relates to post-secondary education.

Facebook posts were manually transcribed, with the assembled comments having occurred between November 1st 2018 and January 21st 2019. As the posts transcribed from Facebook were transcribed with the specific intent that they would be eventually be loaded into R and examined using the dplyr and tidytext packages, the data file was initially constructed in such a way as to require minimal data cleaning. The time path of Facebook posts is charted in figure 2. Though it does not exactly match the time path of Twitter posts, there are somewhat similar peaks and valleys. There is next to no activity about Christmas day, and the number of posts increases in early January as the Winter term begins. Overall there were 1,670 posts transcribed from Facebook. The breakdown of posts versus institutional classification is included in table 3. One weakness of the data that was transcribed was that a full three quarters of the data can be attributed to the CARI and CCI sectors. Worth noting is that this volume of replies was in response to only 121 initial posts by the four CARI institutions, compared to 230 initial posts by the eleven CCI institutions. As a rapid check of discussion drivers, the number of replies to each initial post were counted. All transcribed Facebook posts were necessarily responses to posts made by the institutions themselves, there was a possibility to the topic of the initial posts was what drove the post volume. Individual Facebook posts themselves were summarised in terms of the number of replies generated by each post. These were then filtered to include only posts which had at least ten replies. In turn these posts were tokenized using tidy_text's “unnest_tokens” function. This list was filtered to remove stop words, and then the terms were again aggregate to count the number of distinct posts in which each term had appeared. These were filtered again to only terms which had appeared in three distinct posts. Terms which met these criteria, as well as the sectors of the institutions that posted them, are attached in table 4. Of these resulting 23 terms, only 2 appeared in the 60 terms included in the Twitter LDA results comparing large institutions against smaller institutions: "University," and "people." The four terms which were used most frequently, each used in at least five separate posts which all generated at least ten replies, were "campus," "student," "support," and "university."

Sentiment Analysis

The terms extracted from Twitter were matched against the "nrc" lexicon, then grouped and summed according to emotion. Due to the differing volumes in tweets sent to each classification of institution, the number of sentiments expressed was converted to proportions in order to facilitate easier comparison between institutions. The proportional breakdown is attached in table 5. In order to further simplify the comparison, the proportions were converted to rankings of which emotion was most common to which emotion was least common. These rankings are attached in table 7. As can be seen, all institution classifications except CARIs and Polytechs received more words indicating anticipation than any other emotion, followed by trust, with joy being the third largest component. Disgust was the least communicated sentiment for every institution. Despite the broad difference in institutional mission, size, funding, and location, the twitter accounts for Albertan post-secondaries receive remarkably similar patterns of language. The same proportional view was applied again, only this time separating by primary versus secondary account status. The results of this roll up are presented in table 8. The same order occurs, with trust, anticipation, and joy being the three most commonly communicated sentiments. Again disgust, sadness, and anger make up the three least communicated sentiments.

The extracted terms were again matched to the “AFINN” lexicon, to estimate whether the collected terms were strongly positive or negative. Like the analysis of the “nrc” lexicon, the collected terms were examined in terms of proportions rather than gross totals due to the different volumes of tweets for each group and were also subsequently ranked. These results are available in tables 9 and 10 respectively. Again, there is little difference between different institutional classifications with all institutions receiving more moderately positive content than any other content, and with strongly negative content being the most uncommon. To get a sense of the sort of language being ranked as strongly positive or strongly negative, the 30 most commonly sent terms for each score were broken out and examined with the attached AFINN score, this breakdown is available in table 11. The same ranking analysis was completed again, this time with the mentioned accounts separated by primary and secondary account status, with the results of this breakdown included in table 12. Once again, all content directed at institutions followed a consistent pattern regardless of mentioning primary or secondary accounts. Moderately positive content was the most common, while strongly negative content was the least common.

The same as messages delivered through Twitter, the Facebook posts were tokenized to individual words, stop words were removed, and the remaining dataset was matched against both the AFINN and nrc lexicons. Results from the nrc lexicon were examined first. The summed sentiments were separated according to institutional classification and were changed from gross counts to proportions. Just as before the proportions were also ranked, with the results of this process included in tables 13 and 14 respectively. Like the results from Twitter, the three most common sentiments were joy, trust, and anticipation with these terms combined accounting for more than half of the sentiment expressed through Facebook posts. Again, the least common emotion expressed were disgust, anger, and sadness. The CARI sector had sadness ranked as the fourth most common sentiment, though this is likely due one of the CARI institutions receiving many replies on a Facebook post announcing a student suicide. Words gathered from Facebook posts were then matched against the AFINN lexicon, converted to proportions, and ranked, with the proportions and ranks attached in tables 15 and 16 respectively. Once again, the most common sentiments expressed were mildly positive, with strongly negative sentiment being very uncommon, and the upper threshold of positive sentiment being the least common. When combined with the findings of the nrc lexicon this suggests that posts on Facebook are quite often positive.

The results of this sentiment analysis are best described as homogenous. Across all sectors, account types, and both social media platforms there is remarkable consistency. When matched against the nrc lexicon joy, trust, and anticipation are the three more commonly expressed emotions in all examined data. In all examined data anger, sadness, and disgust represent the three least commonly expressed emotions. A similar effect occurs when the data is matched against the AFINN lexicon. Mildly positive sentiment was the most commonly expressed sentiment across all collected data, with strongly negative content being the most uncommon. From an emotion and sentiment perspective there appears to almost no difference in the content targeted towards post-secondary institutions no matter how the gathered data is parsed.

Topic Modeling

The first approach used to estimate the subjects which generate Twitter replies to specific institutional types was Latent Dirichlet Allocation (LDA.) As mentioned previously, LDA is a computational method which estimates which words make up a given topic, as well as the proportion of each topic represented by each word. In order to estimate these topics, the algorithm requires that the user provide the number of distinct topics they expect to be present in the data. Intuitively in the case of institutional classifications mentioned was an expectation of six categories, one for each institutional classification used by the provincial government. The results of the algorithm were then truncated to the 15 terms which were most expected to compose each topic, and these truncated results are attached in table 17. Mentions of institution sports team seem to interfere with the ability of the algorithm to separate topics cleanly. Topic 1 includes mentions of the twitter account for the Alberta Colleges Athletic Conference, the organizing body for college sports in Alberta, the twitter account for the Lakeland Rustlers, as well as the words “game,” “team,” and “win.” The second topic, as estimated by the algorithm, seems to have more of a recruitment angle due to the inclusion of words like “students,” “campus,” “check,” and “join.” The third topic appears to be mostly composed of tweets direct at Mount Royal University. Mount Royal University professors Duane Bratt and David Finch, both frequent commentators and columnists for local media, are mentioned alongside of the Mount Royal University Student’s Association (“samrubuzz,”) as well as the sports teams for MRU, the Cougars. Like the second topic, topic 4 appears again focus on sports teams whose games are organized by the ACAC, with mention made to multiple teams including the NAIT Ooks, the Old College Broncos, and the Keyano College Huskies, as well as the terms “game,” “lead,” and “win.” Despite again mentioning the ACAC sport account, the fifth topic appears to have a mix of sports mentions and topics specific to southern Alberta. The terms “acac_sport,” “game,” and “win,” are again clear references to collegiate athletics, and the terms include the twitter accounts for Medicine Hat College’s, the University of Calgary’s, and Red Deer College’s teams. The other terms reference the Calgary Public Library, Calgary Arts Development, WinSport (also known as the Calgary Olympic Park,) the National Music Centre, and the Nakiska ski resort which is also located near Calgary. Taken together this suggests the algorithm is returning some sort of southern Alberta focus for the fifth topic. The sixth topic possibly to had more of an academic focus, including mention of the Canadian Tri-Council agency “nserc_crsng,” the term "research," as well as mentions of University of Calgary law dean Ian Holloway ("lawdeanholloway") and University of Alberta professor Tim Caulfield ("caulfieldtim.") An alternative explanation for both Caulfield and Holloway being grouped together as a topic was that they were present in the media during the time of collection due to providing commentary for events of provincial interest. These results do not appear to cleanly align with the mission and functions of different institutional classifications within the Albertan post-secondary sector, and instead the groupings seem to be based on geography and events in the province during the data collection far more than the institutions themselves.

This process was again repeated to compare primary accounts versus secondary accounts, to determine if any difference existed. For this comparison the LDA function was told that there was only two distinct topics present in the data collected data. As there were only two topics being estimated from the same number of terms the results were trimmed to the top twenty terms most strongly associated with each topic. These results are included in table 18. These two topics show a clearer separation. The first topic is composed of strongly positives words, including "congratulations," "congrats," "amazing," and "awesome," while also having terms more squarely focused on the academic mission of institutions, including "dr," "research," and "nserc_crsng." By comparison the second topic is composed almost exclusively of terms related to college sports. There are terms for both women's volleyball and men's volleyball ("wvb," and "mvb," respectively,) as well the twitter account for ACAC Sports and the terms "team," and "game." Unlike the previous grouping, separating terms by two topics seems to correctly sort terms into categories which resemble topics relevant to primary institutional accounts and secondary institutional accounts.

Log-Odds Ratio

Log-Odds ratios are used to compare the relative likelihoods of terms appearing in one of two bodies of text. Colloquially this may be thought of as magnifying the differences in content between two different samples. One complication of this approach is that terms which are mentioned only a few times in one category appear as incredibly likely to appear in one category, and tremendously unlikely in the other, based solely on their infrequency. To remove these terms from the analysis the terms were filtered to include only those that had been used at least 25 times during the sampling period. Unfortunately, this removed the Specialized Arts and Cultural Institution category entirely, as it had not received enough Twitter mentions to break this threshold. Additionally, due to the number of classifications, six plus a mixed mentions category, it was not feasible to compare every classification against every other classification as this would have required repeating the analysis 21 different times. In order to pare down the resulting analysis to an understandable volume, this analysis was restricted to the four comparisons of highest interest based on the role of different types of institutions in the Albertan post-secondary sector and how those roles overlap. To further assist with ease of interpretation each log-odds ratio table is reduced to the first 30 and the last 30 words, representing the 30 words most likely to appear in each group. As the tables produced from the log-odds calculation ranged from 350 rows to 1040 rows this simplification was necessary.

The first comparison performed was between the CARI sector and the BASI sector. This comparison is of interest due to two institutions that comprise the BASI sector being in the same cities as the two largest institutions that comprise the CARI sector. While the CARI sector is expected to conduct research and operate graduate studies programs alongside their bachelor’s programs, the BASI sector is instead directed to focus on bachelor’s level education. Both sectors operate in the same cities with an overlapping body of students and provide overlapping academic programming. The results of the log-odds comparison are attached in table 19. A few of the terms' association with each section make immediate intuitive sense. Professors Tim Caulfield, Ian Holloway, and Tamzin Blewett ("drtblew,") are all employed by institutions in the CARI sector, and as such are mentioned by twitter users mentioning those institutions. Likewise, Duane Bratt, David Finchy, and Lesley Brown ("lesleyabrown") are employed by the BASI sector. Other terms appear to be strongly related to the institutions' unique roles within the post-secondary sector. The CARI sector received more mentions of "dr," "researchers," "science," "award," and NSERC. As well, the CARI sector received mentions of the faculties offering programs only available at their institutions, including the University of Alberta Faculty of Medicine and Dentistry ("ualberta_fomd,") and the University of Calgary School of Medicine ("ucalgarymed.") By comparison almost all the terms sent mostly to the BASI sector are sports related. The sole terms which do not have a sports association are "mru," and "weekend." Curiously the premier of Alberta, Rachel Notley ("rachelnotley,") is much more likely to be mentioned in messages sent to the CARI sector, while former minister for the post-secondary education Thomas Lukaszuk ("lukaszukab") is more likely to be mentioned in messages sent to the BASI sector.

The second comparison was between the CARI sector against the CCI sector. Unlike the previous comparison, there is no overlap in mandate between these two sectors. CCI's deliver certificates, diplomas, and other career-focused education programs primarily in smaller communities throughout the province. The results of this comparison are attached in table 20. Many of the terms strongly related to the CARI sector are unchanged from the previous comparison, including the references to specific professors, medicine schools, and research. Unlike before the top thirty terms now include the terms "innovation," "sustainability," and "health." Like the previous result, CCI's were significantly more likely to receive messages which included terms related to their sports teams. Multiple sports teams that had not been included in the search terms were mentioned, as were a variety of sports and the ACAC again. Also mentioned frequently was the term “stmarysu,” which appears to be a mistaken effort to mention Calgary’s St Mary’s University, however this term relates to the twitter account for St Mary’s University in San Antonio, Texas. The sole non-sports related terms included in the 30 terms most strongly attached to the CCI sector are Albertan MLA Danielle Larivee (“daniellelarivee,) and Albertan entrepreneur Mike McCready (“mikemccready.”)

The CCI institutions were then compared against the Polytechnics. Neither sort of institution is expected to conduct research, nor does either institution awards bachelor’s degree. Though there are two colleges located in the same cities as polytechnics, CCI institutions are in much smaller urban centres than polytechnics. Despite these differences, both groups of institutions are tasked with providing career-focused education meant to provide a training to meet local labour force demands. The comparison between CCI’s and Polytechnics is included in table 21. For both categories sports related topics continue to make up a significant number of terms. When sports related terms are ignored, the terms more strongly associated with the CCI sector include “daniellelarivee,” “college,” “learning,” and “research.” The term “research,” is interesting given that the relatively minimal research conducted on the part of the CCI sector. However, this may be explained by other terms associated with the CCI sector, including “mikemccready,” “vrara_alberta,” “mergevr,” and “vrstories.” These accounts all related to Virtual Reality technology, and different organizations which are developing this technology, and representing both the interests and potential impacts of VR. Inclusion of these terms seems to have been driven by Lethbridge College opening a virtual reality book club in mid-January. Polytechnics also received messages related to their sports teams, however they also received messages related to events occurring in their cities during the collection period. As well polytechnics received messages which included terms related to events within their cities. The terms “yyc,” and “yeg,” are airport code shorthand commonly used on twitter to refer to Calgary and Edmonton respectively, while "greycupfestival" and "horsepower" referred to the 106^{th} Grey Cup, the championship game of the Canadian Football League. This game was played on November 25, 2018 in Edmonton, where NAIT is located, and was played by the Calgary Stampeders. One of the hashtags used to indicate support for the Calgary Stampeders was "#horsepower," due to the Stampeders' use of a horse as their logo. As mentioned before "calgaryartsdev," "nmc_canada," "winsportcanada," and "skinaskisa," all relate to organizations with headquarters in Calgary, and appear as being more strongly associated with the Polytechnics. The term "calgarylibrary" likely appears due to Calgary's central public library opening in the November of 2018. Both the NAIT and SAIT students' associations appear in the top 30 terms as both "naitsa," and "saitsa," as well as the SAIT Alumni association as "saitalumni." Finally, the term "innovation," appears attached to the Polytechnic sector. The difference between these institutions is not as expected, with CCI's having more academic and research-oriented terms associated while Polytechnics had more terms associated with their local communities.

Finally, institutional classifications were sorted into "big school" and "small school" categories. The CARI, BASI, and Polytechnic sectors were considered "big schools" as they represent the largest student bodies in the Albertan post-secondary system. By contrast, CCI's and IAI's were "small schools" due to their smaller student bodies and more limited program offerings. The results of this comparison are included in table 22. Though both groups continued to have references to institutional sports teams, sports bodies, and conferences these references are no longer overwhelming the full list. Individuals employed in the CARI, BASI, and Polytech sectors appeared in the large school terms, including Tim Caulfield, Ian Holloway, and Tamzin Blewett. Oddly Rachel Notley again appears to be a common topic in Twitter posts tagging institutions in these three sectors. Other topics included in the messages to these sectors included a collection of terms related to events in the cities in which the institutions were located, including “yeg,” “yyc,” “calgarylibrary,” ”nmc_canada,” “cbcnorth,” and “winsportcanada.” The large schools, whose mandates include research, were also proportionally more likely to receive messages including the terms “nserc_csrsng,” “innovation,” “researchers,” and “conference,” suggesting that received messages may have discussed research initiatives within the sector. CCI’s, by comparison, received messages discussing Augmented and Virtual Reality, AR and VR respectively. These terms included “cathyhackl” (referencing an AR speaker,) “arstories,” “mergevr,” “vrara_alberta,” and “mikemccready.” Once again Danielle Larivee is mentioned frequently enough in messages sent to smaller schools to appear in the list of the top 30 associated terms.

The log-odds calculation was repeated with tweets sorted into “primary accounts” and “secondary accounts” categories. As this creates only two groups for comparison there was only one table generated, so additional sorting was not required for ease of comparison. Just as before the 30 terms most strongly associated with primary accounts were selected, as were the 30 terms most strongly associated with secondary accounts. The results of this process are included in table 23. Terms most strongly associated with the academic mission of institutions, terms such as “research,” “learning,” “program,” “class,” “innovation,” and “education,” were most likely to be associated with primary accounts. By comparison, a full two thirds of the terms most strongly associated with secondary accounts are sports related, with additional terms having the potential to be used in a sports context depending on usage. The sole term associated with secondary accounts which is unambiguously not sports related is the twitter account of Ian Holloway.

The data from Facebook was then examined using the same process, using log-odds ratios to estimate which terms were most strongly associated with which institutional classification. As the overwhelming volume of data was generated from the CARI and CCI sectors they were compared against each other. Words which were used fewer than 5 times were filtered out. Once again, a log-odds model was run in order to establish associations, and subsequently the top 30 terms most strongly associated with each institutional classification were extracted. The results of this are included in table 24. Unlike the messages posted to Twitter, there is not a neat division between the sectors. The CARI sector's strongest associations do feature some academic oriented terms such as "semester," "courses," "study," "reading," and "professor." At the same time, the CARI sector also includes words which suggest other specific topics such as "discrimination," and "indigenous." While the CARI sector appears more focused on specific topics, the CCI sectors contains significantly more expressive language. Words such as "fantastic," "success," "loved," "super," and "wonderful," are present in the top 30 terms. Another interesting difference is that the CCI sector includes the name of specific institutions in their posts, as well as the terms "experience," and "remember." This suggests that posts made on CCI's Facebook pages may include more retrospective or nostalgic comments. Due to the minimal volume of data gathered from other institutional classifications, additional single classification vs single classification tests were ruled out.

To further investigate the topics used in Facebook posts while mirroring the examination of Twitter information institutions were divided into a "large institutions" group and a "small institutions" group. A log-odds model was again run to determine the terms which were most strong associated with each grouping of institutions. Identical to the previous log-odds analysis the 30 terms most strongly associated with each group were selected, and the results of this selection are included in table 25. Like the comparison between the CARI sector and the CCI sector, this is no clear separation between larger institutions and smaller institutions in terms of topic. There are some trends which continue from the CARI vs CCI analysis, such as the larger institutions having a stronger association with terms such as "courses," "semester," "study," "reading," "read," "professor," "assignments," and "scholarship."

Smaller institutions have some strongly associated terms which suggest that posters may be more sentimental including “story,” “remember,” “experience,” and “chose.” The smaller institutions also include some more forward-facing terms including “choose,” “future,” “continue.” Finally, smaller institutions also include more positive, descriptive terms that were strongly associated with the CCI sector, including “fantastic,” “loved,” “super,” and “beautiful.” These descriptive terms’ inclusion is likely driven by the very high proportion of posts which were on the Facebook pages of CCI institutions specifically highlighting some feature of the institution and their campuses.

Redundancy, Context, and Relation

As a check on the context of term usage, a pair-wise correlation was constructed to identify terms which were often used within the same tweet. Stop words were included in this list, allowing for the inclusion of terms such as “not,” “isn’t,” or other words which reverse the meaning of other terms. A data frame was assembled, composed of all the unique terms which had been included in the top 30 associated terms for any of the examined log-odds ratio comparisons. Effectively this illustrates which terms often referred to the same topic and had been double counted in the assembly of the log-odds ratios, as well as identifying terms whose meaning has been incorrectly interpreted by the previous analysis. The resulting listed was more than 8000 rows long. To assist with interpretation the list was filtered down to only terms with a correlation coefficient greater than 0.25, signifying that the terms appear within the same tweet in at least twenty five percent of the data. By default, a pair-wise correlation will produce duplicates as each term was included once in the “item 1” column, and once in the “item 2” column. To remove duplicates the resulting list was filtered to remove every duplicate row. The cleaned output is included in table 26. None of the items included were associated with a negative term. Many of the sports terms which dominated the LDA analysis were closely related to each other, while the term “regional,” was unexpectedly in reference to sports conferences.

The most commonly used words in Facebook replies, as well as their associated AFINN scores were ranked based on their usage, pared down to the top thirty most used words, and then matched against the 30 most often words used in Twitter messages. A marker was assembled to note if the term appeared in both lists. The results of this comparison are attached in table 27. Not only is there significant overlap in the first five terms, four of the five overlap, but a full two thirds of the words used overlap. Only four distinct words with negative scores appear across both sources, with two of them in common between the two groups. In order to compare post frequency, the time path of the two social media platform posts were overlaid upon each other. A dual y-axis was employed to assist with comparing the two platforms despite the different volume of interactions. These overlaid time paths are presented in figure 3. Except for the gap within a week before Christmas and a week after Christmas, the two platforms’ activity do not seem to line up. To confirm this suspicion a very simple linear regression model was estimated, regressing Twitter post volume onto Facebook post volume. The results of this regression are attached in table 28. Though the coefficient on Facebook posts is both positive and statistically significant, the extremely low Adjusted R-squared suggests that this model is a very poor fit for the data, and that in turn the volume of posts on one platform is not a good predictor of post volume on the other.

Chapter 5: Discussion and Conclusions

Conclusion

Institutional Classifications. Imagine a spectrum of post-secondary institutions with graduate degree program teaching, research focused institutions on one end and smaller, community driven, certificate and diploma granting institutions at the other side. As you progress across the spectrum beginning at the research-intensive side moving towards the community side you increasingly find that sports activities make up a higher proportion of topics sent to institutions through Twitter. At one end, in the CARI sector, sports make up a comparatively smaller proportion of discussed content. Instead activities unique to this sector, such as research and medical schools, make up a comparatively larger portion of the content mentioned by Twitter users. As you proceed across this spectrum, through the BASI, Polytechnic, and finally the CCI and IAI sectors, sports compose an increasingly larger proportion of the content sent to institutions on Twitter. This is driven, in part, by secondary accounts receiving an increasingly high proportion of tweets when compared against primary institutional accounts. Through each section of the hierarchy there is content unique to each institutional classification, seemingly driven by the unique academic activity of each institution, their involvement in their local communities, and events occurring near their campus, and this content is most often directed towards the primary accounts of each institution. Little difference was found between different institutional classifications in Facebook posts, save that sectors including large universities tended to have a stronger association with language that may be considered more academic. Regardless of platform language used in messages which mention institutions is moderately positive. Anticipation, trust, and joy are the most common emotions expressed towards institutional accounts, while anger, disgust, and sadness are the least common. From the perspective of the post-secondary sector, both Facebook and Twitter are predominantly positive spaces.

Primary vs Secondary Twitter accounts. There is seemingly no difference in tone between messages sent to primary institutional Twitter accounts and secondary institutional Twitter accounts. Across both account types messages tend to be moderately positive, with moderately negative sentiment being the least common. The distribution of expressed emotion is quite similar to that of the separation by institutional classification. Anticipation, joy, and trust were the most commonly expressed emotions, with anger, fear, and disgust being the least common. Regardless of the sort of Twitter account that is receiving the message, messages sent to institutions remain quite positive and upbeat. Messages sent to primary accounts mention the academic function of institutions, as well as popular figures and other institutional figures active on social media, and events on campus. Messages that mention secondary institutional accounts overwhelmingly discuss sports teams. When combined with the results of the separation along institutional classification these results suggest that in general institutions can expect their primary accounts to receive positive messages related to their academic business and their activity in the community, while secondary accounts more than anything will receive messages related to their sports teams.

Facebook vs Twitter. Facebook accounts received significantly fewer messages than Twitter accounts over the same period. This may be driven by the ability of Twitter users to tag institutions in any message, while Facebook comments are restricted to replies on posts made by the institutions themselves. An alternative explanation is not that users' ability to direct content towards institutions is limited by social media platform design, but rather that each platform is used for different purposes by users. It may be the case that users following institutions on Facebook simply derive no gratification from engaging in discussion on that platform, while they may be more comfortable mentioning institutions in Twitter posts. While there are differences in content sent to institutions that can be separated by the institutions' classification within the post-secondary sector, these differences are much weaker in the messages posted to Facebook. Posts sent to institutions through Twitter are more likely to mention institutional features, while posts made on institutional Facebook pages tend to be more oriented towards experiences. Regardless of platform the language that institutions receive tends to be positive, with strongly negative language and strongly positive language being the most uncommon sentiment expressed. Regardless of platform messages are most likely to express joy, trust, and anticipation, with anger, disgust, and sadness being the least commonly expressed emotions. Despite the tremendously lower volume of posts received, Facebook also appears to be a mostly positive platform. To reduce both platforms to an overly simplistic statement, both Twitter and Facebook appear to be vehicles for an institution's audience to relay "congratulations," and "thanks."

Discussion

These results present some good news for the Albertan post-secondary sector as their social media audiences are generally quite positive when mentioning institutions, at least in their public messages over the examined time period. Though the identity of their audience remains unknown there is no denying that the messages received across both platforms were generally positive. Anecdotally, the most hostile messages noticed during the data collection process were from a Facebook announcement that a campus-based restaurant would offer a plant-based burger, hardly a dire topic. If institutions are actively involved in moderation of their Facebook pages as has been suggested (Junco & Chickering, 2010), their efforts have clearly been effective. Regardless of whether it is by moderation or simply by coincidence, the tone of conversation seems to represent a generally positive environment and it may be the case that new members of the audience are following the example. To quote boyd (2007), "In the absence of certain knowledge about audience, participants take cues from the social media environment to imagine the community" (p. 131). Institutional communications at the examined institutions can rest assured that their normal state is one of general positivity and sudden surges in aggressive or negative expression are both uncommon and should not be dismissed as the normal course of business. In the case of both platforms institutions and their audiences have resisted the temptation to publish negative messages in an effort to go viral (Stieglitz & Dang-Xuan, 2013; Tsugawa & Ohsaki, 2017). Institutions seem to enjoy a positive environment amongst those who mention them, and monitoring social media for deviations from this environment may be of interest for institutions with a deep-seated interest in their reputation as suggested in Eriksson & Olsson (2016). For Canadian institutions this may suggest a baseline environment for their Facebook and Twitter accounts; A moderately positive environment, where discussion is most likely to be centered around local events, unique institutional features, and a varying proportion of content related to an institution's sports teams.

If there is a desire on the part of the institution for more audience engagement, the audience's positive tone would suggest a relatively friendly community exists. Assuming that audiences have not undergone significant changes in behaviour in the last decade, a significant assumption, previous research indicating that institutions do not encourage interaction through social media (Beverly, 2013; Davis et al., 2015; Kimmons et al., 2016; D. L. Linvill et al., 2012) should be reconsidered in the context of this positive audience. Institutions have not necessarily avoided this interaction in order to avoid conflict, but instead may have not seen any benefit from interaction for achieving their goals. In the case of Twitter, what discussion there is seems to be driven in part by the unique characteristics of each institution and events occurring in the cities around the institutions. That there is discussion at all, especially discussion which so frequently includes terms such as “congratulations,” and “thanks,” suggests that users may also be using these platforms to achieve social gratification by engaging in cursory social interaction (Gan & Wang, 2015). It is very difficult to further examine the gratifications of users without an opportunity to interview them, as many gratifications unfortunately are not observable through the data that was collected. Users could be following institutions for their own purposes without actively participating in the discussion.

Facebook’s differences from Twitter were possibly due to the examined comments being composed solely of replies to an initial institutional post, however it is also possible that users seek different gratifications from Facebook than from Twitter. The ability for users to direct a tweet to institutions as well as anyone else they so desired, without having to reply to an already existing post provides greater freedom for Twitter users to engage in discussion. Likewise, the ability of Twitter users to select a username totally divorced from their identity, unlike the many Facebook users whose username is their actual name, may have altered their behaviour. Consistent with previous findings (Beverly, 2013; Kimmons et al., 2016; D. L. Linvill et al., 2012) many of these initial Facebook posts were news items, informing the audience about events on campus. However, there were multiple posts within the data which encouraged students to comment upon the post with a personal anecdote, as well as pictures of the various campuses and alumni success stories. This supports the findings of Clark, Fine, and Scheuer (2017) and Schultz and Peltier (2013), and suggests that institutions may be using Facebook pages to facilitate engagement or conversation in an effort to strengthen the audience’s sense of attachment. As an example, there were many posts made on Facebook that asked or encouraged the audience to relay their experiences and memories of the institution, often using terms like “campus” and “community”, and these posts received numerous responses. This could suggest that Facebook commenters were achieving some form of social gratification (Gan & Wang, 2015) from engaging in these discussions. These findings may support previous findings (Nevzat et al., 2016), which suggest that social media is a medium for communities centered around institutions to interact, however without a more detailed examination of the content as well as the specifics of which users are responding to which users most frequently this is far from conclusive.

From a financial perspective, institutions may benefit from producing content which they believe is likely to be reshared within these communities (Arriaga et al., 2017; Stieglitz & Dang-Xuan, 2013; Wolny & Mueller, 2013), fostering strong feelings of belonging to the institution (Nevzat et al., 2016; Rutter et al., 2016) and potentially leading to increased donations (McAlexander & F. Koenig, 2012). Smaller institutions may find more value in developing these communities on Facebook due in large part to the higher share of messages on Facebook, however this strategy may only yield a short-term benefit as younger internet users shift away from Facebook (Gruzd et al., 2018). Of note is that, as mentioned in the descriptive statistics, the smaller institutions in the CCI sector received a proportionally smaller number of responses to Facebook posts than the larger institutions of the CARI sector. At face value this would seem to contradict previous findings (Peruta & Shields, 2017) which suggested that smaller institutions’ content may be more likely to encourage interaction, however the number of users that viewed each post was not observable. It may be the case that larger institutions simply have a larger audience and generated the higher number of replies through sheer volume of views rather than by posting more engaging content.

The number of messages received across both platforms would indicate an audience that is already engaged, or at least institutions which have found a way to promote engagement from their audiences. Across all sectors the frequent mentions of sports teams and events may suggest that institutions have heeded previous recommendations and are using sport to drive institutional attachment, in turn providing strong support for the finding of France, Finney, and Swerdzewski (2010). Given the level of engagement of the audience, the positive tone of conversation, and common topics aligning to either an institution's defining features or to their collegiate athletics, it may be the case that institutions are savvier than expected in their use of both Twitter and Facebook. An alternative, though far simpler, explanation may be that a large proportion of the audience for Albertan higher education institutions are athletics fans and in turn this large proportion biased the results towards positively. Regardless of audience composition, only minor differences in expressed sentiment exist between Facebook and Twitter, though Facebook pages receive fewer messages than Twitter accounts. This somewhat contradicts the findings of Eriksson and Olsson (2016), however none of the data gathered overlapped with any times of crisis for any of the examined institutions. Quite the opposite in fact, the data collection process overlapped with Christmas, New Years, and the start of a new semester. All three of these events are generally treated as positive occasions and may contributed to an unusually positive audience. Had the data collection period overlapped with some dour event institutions may have used different communications strategies on each platform, as recommended in Eriksson and Olsson (2016), and the resulting content directed towards each platform may have been different. A wider data collection period may contribute to more negative content by including deadlines for dropping classes, or when midterms are most common.

Taking a broader perspective on the findings of this paper, the lack homogeneity in Twitter messages is intriguing. Apart from the proportion of content dedicated to sports, different institutions are indistinguishable from one another in terms of emotion and sentiment expressed in Twitter posts. What content differences do exist are very specific to institutional features themselves. Across all examined institutions the different audiences, specifically those that mention institutions in their tweets (Smock et al., 2011), appear to have been behaving fairly consistently in terms of sentiment, emotion, and topic. This consistent use suggests that from a uses and gratifications perspective these audiences appear to be using Twitter to achieve similar gratifications, and that the real difference in use is restricted to the institutions the audience has chosen to follow and mention.

Taking a slightly broader perspective on the findings again, the comparison between Twitter and Facebook suggests that users are likely using these platforms differently and that in turn institutions may be well served by using different strategies for both Facebook and Twitter. From the perspective of sentiment and emotion there is no difference in the content posted to Twitter and to Facebook. Both platforms return the same ordered rankings of both expressed sentiment and expressed emotion. These results must be interpreted with some level of skepticism, however. Despite the very high proportion of positive language, these results do not necessarily indicate that the audience feels positively about the institutions themselves, rather that when they engage in discussion which mentions institutions, or when the audience engages in discussion on institutions' Facebook pages, the discussions tend to be positive. The topics discussed and the language employed seem to have a more nuanced difference than expected. Language sent to institutions through Twitter included a large proportion of mentions of other Twitter accounts and many terms which were likely hashtags, especially terms related to sports, while messages published through Facebook included much more descriptive language. The use of more descriptive language in Facebook posts may indicate a higher proportion of comments discussing experiences. This difference in language suggests that the different audiences for Facebook and Twitter may be interested in different sorts of discussions.

This study likely suffers from notable endogeneity challenges. Users of both social media platforms are not viewing a strictly chronological string of content, rather algorithms produced by both platforms' parent companies are modifying which content appears in which order. The resulting filtered content is tailored to encourage some particular form of user interaction, be it sharing or interaction. Likewise, both Twitter and Facebook modify these algorithms to match any changes in their business strategies. As a result, there are three possible influences which may be skewing the results observed in this paper. First, the content viewed by users on both platforms may have been prioritized by the underlying algorithms due to a high likelihood that users would find it to be agreeable. Should this be the case the resulting positive tone of discussion is partially engineered; The users who responded to institutions were the users the algorithms assessed as having a high likelihood of reacting positively. Second, depending on the goals of Facebook and Twitter, different content from different sources may have been prioritized and skewed the number of interactions observed on each platform. Twitter may have prioritized content considered to be news, while Facebook may have tailored content to encourage users to interact with other personal pages rather than an institution's business page. In this case the differences in the level of interactivity may be driven by differences in the number of users who saw content from institutions. Due to the frequent practice of higher education using Twitter as a newsfeed, it may be the case that more Twitter users saw items about institutions than Facebook users, and in turn this drove the differences in interactivity. Third, content related to sports may have been so popular, and possibly crowded out other content, simply because it was prioritized by both Facebook and Twitter's underlying algorithms. An even more speculative interpretation could be that users for whom sports related content is prioritized may be more likely to interact with that content than users for whom sports related content is deprioritized. In either case, content manipulation by the social media platforms themselves may explain why such a high volume of sports related content was gathered in this study.

To what extent these endogeneity problems should concern institutions themselves is unclear, however. From a strictly naïve perspective, institutions have no influence over how Facebook or Twitter choose to prioritize content or account. Regardless of how the content viewed by users has been tailored, the takeaways for institutions themselves are unchanged. Knowing that the evidence presented in this paper suggests that the audience tends to respond positively to institutions, institutions may be better served from adoption social media strategies which encourage a higher level of interaction amongst their audience. Should the findings in this paper be universally applicable, it is very unlikely that institutions will be on the receiving end of much negative content.

The most definite difference between Twitter and Facebook lies in the different volumes of published content which targets institutions. As this paper did not estimate the size of the audience, through recording the number of followers or otherwise, it is difficult to estimate how much of the audience has chosen to interact with institutions through these platforms. Assuming that both platforms have the same audience, a substantial assumption, Twitter audiences appear to be more interested in discussion in order to achieve a gratification whereas they do not make use this feature of Facebook (Smock et al., 2011). Alternatively, and perhaps more realistically, if we assume that the audiences sizes for institutions mirror the social media usage of online Canadians then the Facebook audiences should be twice as large as Twitter audiences (Gruzd et al., 2018) and Twitter audiences appear to be even more disposed towards interaction. Given the difference in content targeting institutions, it appears that audiences have self selected towards Twitter and Facebook in order to use them differently and to achieve different gratifications. Perhaps this should be unsurprising, given previous applications of uses and gratifications which suggested that different platforms are used for different reasons (Korhan et al., 2016; Smock et al., 2011). From the perspective of institutions this suggests that the value of Facebook does not lie in interacting with their audience, while there may be some value to be gained from making use of the higher level of interactivity of Twitter audiences. In turn institutions may be well positioned to make use of social media’s interactivity to pursue institutional goals such as promoting campus sustainability initiatives (Hamid et al., 2017), or simply improving students’ perceptions of their educational experience (Neier & Zayer, 2015; Rutherford, 2010). Given the significant differences in messages received between Facebook and Twitter, institutions may be well served by using Twitter for interactive purposes while restricting Facebook to more focused newsfeed-like uses. This approach would seem to better match the audience’s desired use of each platform.

Institutions’ use of Facebook posts encouraging the audience to post a fond memory or offer an opinion, sometimes to be entered in a draw for a prize, contradict previous findings which have suggested that institutions do not encourage interactivity (Beverly, 2013; Davis et al., 2015; Kimmons et al., 2016; D. L. Linvill et al., 2012). This may be explained by these previous papers’ (Beverly, 2013; Davis et al., 2015; Kimmons et al., 2016; D. L. Linvill et al., 2012) focus on Twitter, and lack of examination of Facebook. A possible explanation for this unexpected difference between Facebook and Twitter is that audiences may be using different platforms to achieve different gratifications. In terms of application, this suggests that institutions may be best served by using Twitter to pursue more interaction with their audience and using Facebook in circumstances where interaction is less important. Previous works have focused on how institutions can increase engagement on their Facebook pages (Arriaga et al., 2017; Brech et al., 2017), while other works have examined the lack of interactivity of institutions through Twitter (Beverly, 2013; Kimmons et al., 2016; D. Linvill et al., 2012). The results in this paper suggest that institutions may be using these tools opposite of what would be most effective and would be better served by pursuing more engagement with their Twitter audiences while accepting the limited engagement of their Facebook audiences.

Study Limitations

One of the primary flaws with the approach taken in this paper is that there were not any controls in place to filter or identify sports related content. Adding a third sort of account, sports related accounts, would have significantly reduced the difficulty of parsing out sports related content. Unfortunately, the approach used in this paper did not account for the popularity of collegiate athletics amongst Twitter followers of post-secondary institutions. As a result, the volume of sports content simply crowds out any other topics that may have generated discussion. This problem is highlighted by the declining ratio of non-sports content to sports content as you move down the institutional hierarchy. It may be the case that there are topics of interest to the Twitter audiences of CCI’s, however those discussions are currently drowned out. While the top of the institutional hierarchy seems to receive sufficient tweet volume across a breadth of topics to offset the sports discussions, driven in large part by the relatively high proportion of messages sent to primary accounts over secondary accounts, CCI’s are instead overpowered by sports mentions. This problem exists partially by construction, as ten sports related accounts were included in the list of secondary accounts. Had athletics accounts been separated into their own category this problem could have been mitigated. Similar issues may exist for institutional Facebook pages across the broader post-secondary sector, however insufficient data was available in order to identify this effect. Facebook accounts generally may receive more interaction on sports-related posts than on non-sports-related posts, however the sample of Facebook data available, as well as the lack of simple identifiers like hashtags, was insufficient to examine this in more depth.

The data cleaning process could have been improved, and valuable information including punctuation strings could have been preserved, through better application of the tidytext package. By default, the tokenize function automatically deletes punctuation including hashtags, emoji, and "@" signs. During the data cleaning process this shortfall was noted, however simply disabling this option within the tokenize function produced long strings of seemingly random punctuation, while using other options to preserve punctuation split each piece of punctuation as its own token. As an example, a simple emoticon such as ":)" was divided into all three distinct pieces of text: a colon, a space, and a closing bracket. When items were grouped or the data frame manipulated, R will automatically reorder the data frame according to the values present in the left-most column, leading to undecipherable strings of punctuation and random pound signs without context. Preservation of "@" signs and "#" signs with their associated text strings is possible using regular expressions, however this code was not developed until far too late in the writing process for integration into the analysis. An additional limitation was the lack of lexicons which could attach to phrase composed of multiple words. Phrases such as "way to go" convey a moderately positive sentiment, however, were they were excluded due to none of the three composite words having a uniquely positive effect.

Just as the data cleaning could have been improved with foresight, so too could the content analysis have been improved through superior planning. Rather than finding unique terms in the content analysis results and subsequently trying to determine why they had appeared; a better approach would have been to assemble a rough timeline of events or newsworthy items from each institution in advance. This would have simplified some portions of the analysis. Though this would have expanded the data collection phase even more, it would have provided useful added context to understand why different terms were frequently mentioned. Had some sort of temporal component been attached to terms which had been frequently mentioned, it would also have explained why specific terms clustered around specific points in time.

This study has highlighted the importance of timing in conducting research. Following Facebook's API access restrictions, it has become exceedingly difficult to gather data from Facebook at a large scale for research purposes. Had this study been undertaken four years ago there would have been significantly more Facebook information available, and this may have provided an approximately equivalent amount of data to that gathered from the included Twitter accounts by including secondary Facebook accounts. Having a similar volume of data from each platform, or at least a similar number of accounts, would have allowed for a more proportional comparison. Similarly had this research been undertaken years ago both platforms may have been proportionally more popular than some newer competitors. Complicating this paper's relevance further is that Twitter remains continuously embroiled in fresh controversies, while recent news suggests that the number real individual users on the platform has begun to decrease (Wang, 2018). If Twitter truly is in decline, and is becoming a decreasingly important communications platform for institutions, then research suggesting that it is a tool generally used for positive discussion of the unique characteristics of institutions has no real application. From a strictly academic perspective, research into a potentially soon-to-be abandoned platform has a time limit on value, and likely will not be of use into the future.

Given that institutions received predominantly positive messages their Facebook and Twitter usage have little reason to change, absent a deliberate decision on the part of institutional leadership to revamp their social media goals. The Facebook and Twitter audiences for Albertan post secondary education seem to be remarkably positive, fans of sport, and often discuss the unique functions of each institution. Given the previous findings that institutions are most likely to use their Twitter pages a newsfeeds (Beverly, 2013; Kimmons et al., 2016; D. L. Linvill et al., 2012) there is no reason to believe that this is not a satisfactory outcome for institutions. Though they may not intend for their social media to be highly interactive, the audience appears to still be quite pleasant. If the sole motivation for institutions is information sharing, then the state of audience interactions is already an ideal outcome for institutions as it does not appear to interfere with that purpose. They have accomplished their goals and their audience is not displeased, providing little reason for institutions to alter their social media behaviour. However, as this study did not include an examination of institutions' social media policies, it is currently unknown if institutions have successfully achieved the goals which motivated their social media usage. It may be the case that a certificate and diploma granting institution had aimed to use their social media accounts to promote the academics of the institution, and instead have found their Twitter and Facebook pages primarily used to discuss their sports teams. In this instance, despite the generally positive reception of their audience, the institution may still seek to change their published content to guide discussion more towards their academic programming. Also, of note is that this study did not examine the number of retweets or shares of institutional posts. If the sole purpose for an institution's Twitter and Facebook use is for information dissemination, this paper's approach lacks any way to measure if this goal has been achieved.

Future Research

Sample size was not a limitation of this study, as it captured all the Twitter and Facebook traffic directed towards institutions within the examined time frame. That there was much less traffic than expected is more due to a coding error at the outset of the research project than anything other cause. Regardless, it is currently unknown whether there is a constant stream of Twitter and Facebook content sent to institutions throughout the whole year, or whether there are ebbs and flows throughout the course of the year. Even within the limited window examined there was a clear dip centered around Christmas and the starting of the new year. There may be other calendar dates which cause similar peaks and valleys. As a component of this, it is unknown whether the unique topics directed towards each institution are the same throughout the course of the year, or if the results observed in this study were solely the product of significant news items the occurred during observation. In order to control for the possibility that timing is important, or that the content most discussed by the audience is a product of either recent events or the time of year, a wider time frame should be collected. Ideally a full year worth of data would be gathered. Not only would this provide a more comprehensive view of the content directed towards institutions, it would also allow for data to be split into different time periods. Splitting according to time periods would allow for easy examination of whether the most common topics are constant through the year or if they change periodically.

As per the failing noted in the study flaws, any additional data gathering that takes a similar approach would be well advised to include all possible sports accounts in the sample. While every institution was limited to a maximum of five accounts in order to prevent institutions with many accounts from overwhelming the analysis, future research may be well served by including all possible accounts including student unions and other student associated organizations. This would capture tweets which were not captured in this analysis and would provide a fuller understanding of all the content directed to institutions, not just to their primary accounts and a handful of secondary accounts. An additional avenue which was not explored was the examination of the identity of the posters. In the case of Facebook this was specifically omitted due to privacy concerns. In the case of Twitter, however, there may be some value in scraping the public profile information of posters as a quick check to determine if they are affiliated with or employed by any of the institutions under examination. Though there are many very real privacy concerns with this sort of analysis, a rapid and cursory examination to see if the poster's public profile mentions any of the examined accounts or institutions may provide some value. Thought it would not be a perfect classification, it would allow a researcher to examine if comments sent from accounts affiliated with the institution were more positive than accounts that were unaffiliated and would provide an avenue to examine whether the content directed at institutions was positive because it was coming from within.

Instagram is another platform with a large user base which was not included. Given the high rates of Instagram adoption amongst younger Canadians (Gruzd et al., 2018) it is an important social media platform worthy of examination. However, as it is owned by Facebook, the Instagram developer API faces similar restrictions as the Facebook API and there is not currently publicly available access to download large quantities of Instagram posts. Were a researcher to receive permission to gather information a very similar analysis could be performed on the comments made on institutions' posts. Pursuing this line of research would likely require developing some sort of image classification standard, as Instagram posts made by institutions will always be pictures with attached text.

Finally, additional research into this topic would be well served by requesting and collecting policy documents from institutions in order to establish their goals for social media use. A better understanding of the institutions' desired outcomes would assist with determining whether their goals are in line with their audiences' behaviour, and whether social media as a tool is being used effectively. To date there has only been one paper (Pasquini & Evangelopoulos, 2017) which has examined the motivations for institutions' social media use. This is essentially the first half of a broader question; "What are you trying to accomplish with social media?" By comparison the research present in this paper is the part of the second half of a broader question; “What is happening on your social media pages?” Whether those two results are aligned is the real topic of most value to institutions.

Regardless of how the direction taken by future research into this topic, social media remains a common tool for higher education institutions (Tess, 2013) and it is unlikely that the diversity in social media usage across institutions (Pasquini & Evangelopoulos, 2017) will soon disappear. Likewise, given the ongoing popularity of different social media platforms (Gruzd et al., 2018) it is also tremendously unlikely that social media will soon vanish. Institutions, and the broader academic field which studies them, may derive benefit from examining how the use of social media can be improved and how social media can provide more benefit to institutions. This paper’s findings suggest that a better way forward could involve an understanding that different social media platforms’ users may be using each platform for different reasons, that different strategies may be required for different social media platforms, and that there is some reason to believe that institutions are currently operating with a moderately positive audience.

References

Adams, B., Raes, A., Montrieux, H., & Schellens, T. (2018). "Pedagogical tweeting" in higher education: boon or bane? *International Journal of Educational Technology in Higher Education*, *15*(19), 1–16. https://doi.org/10.1186/s41239-018-0102-5

Agostino, D., & Arnaboldi, M. (2017). Social media data used in the measurement of public services effectiveness: Empirical evidence from Twitter in higher education institutions. *Public Policy and Administration*, *32*(4), 296–322. https://doi.org/10.1177/0952076716682369

Anderson, T. (2015). *Seeking Internationalization: The State of Canadian Higher Education* (Vol. 45).

Arceneaux, P. C., & Dinu, L. F. (2018). The social mediated age of information: Twitter and Instagram as tools for information dissemination in higher education. *New Media & Society*, 1–22. https://doi.org/10.1177/1461444818768259

Arnold, R. D., & Wade, J. P. (2015). A Definition of Systems Thinking: A Systems Approach. *Procedia - Procedia Computer Science*, *44*, 669–678. https://doi.org/10.1016/j.procs.2015.03.050

Arriaga, J. L. D. O., Andreu Domingo, D., & Berlanga Silvente, V. (2017). Facebook in the low-cost fashion sector: the case of Primark. *Journal of Fashion Marketing and Management*, *21*(4), 512–522. https://doi.org/10.1108/JFMM-08-2016-0069

Barnes, N. G., & Lescault, A. M. (2013). College presidents out blog and out-tweet corporate CEO's. Retrieved October 8, 2018, from https://www.umassd.edu/cmr/socialmediaresearch/collegepresidentsoutblog/

Bélanger, C. H., Bali, S., & Longden, B. (2014). How Canadian universities use social media to brand themselves. *Tertiary Education and Management*, *20*(1), 14–29.

https://doi.org/10.1080/13583883.2013.852237

Beverly, J. A. (2013). *Public Relations Models and Dialogic Communication in the Twitterverse: An Analysis of How Colleges and Universities are Engaging Their Publics Through Twitter*. University of Southern Mississippi. Retrieved from http://aquila.usm.edu/dissertations/159

Brech, F. M., Messer, U., Vander Schee, B. A., Rauschnabel, P. A., & Ivens, B. S. (2017). Engaging fans and the community in social media: interaction with institutions of higher education on Facebook. *Journal of Marketing for Higher Education*, *27*(1), 112–130. https://doi.org/10.1080/08841241.2016.1219803

Chugh, R., & Ruhi, U. (2018, March 1). Social media in higher education: A literature review of Facebook. *Education and Information Technologies*. Springer New York LLC. https://doi.org/10.1007/s10639-017-9621-2

Clark, I. D., & Norrie, K. (2013). Research and reluctance in improving Canadian higher education. In P. Axelrod, R. D. Trilokekar, T. Shanahan, & R. Wellen (Eds.), *Making Policy in Turbulent Times: Challenges and Prospects for Higher Education* (1st ed., pp. 189–220). Montreal and Kingston: McGill-Queen’s University Press.

Clark, M., Fine, M. B., & Scheuer, C.-L. (2017). Relationship quality in higher education marketing: the role of social media engagement. *Journal of Marketing for Higher Education*, *27*(1), 40–58. https://doi.org/10.1080/08841241.2016.1269036

Constantinides, E., & Stagno, M. C. Z. (2011). Potential of the social media as instruments of higher education marketing: a segmentation study. *Journal of Marketing for Higher Education*, *21*(1), 7–24. https://doi.org/10.1080/08841241.2011.573593org/10.1080/08841241.2011.573593

Copestake, A. (2004). Natural language processing course guide. Retrieved July 6, 2018, from https://www.cl.cam.ac.uk/teaching/2002/NatLangProc/revised.pdf

Correa, T., Hinsley, A. W., & de Zúñiga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2), 247–253. https://doi.org/10.1016/j.chb.2009.09.003

Davis, C. H. F., Deil-Amen, R., Rios-Aguilar, C., & González Canché, M. S. (2015). Social media, higher education, and community colleges: A research synthesis and implications for the study of two-year institutions. *Community College Journal of Research and Practice*, *39*(5), 409–422. https://doi.org/10.1080/10668926.2013.828665

Dolan, R., Conduit, J., Fahy, J., & Goodman, S. (2016). Social media engagement behaviour: a uses and gratifications perspective. *Journal of Strategic Marketing*, *24*(3–4), 261–277. https://doi.org/10.1080/0965254X.2015.1095222

Dumitrica, D. (2017). Fixing higher education through technology: Canadian media coverage of massive open online courses. *Learning, Media and Technology*, *42*(4), 454–467. https://doi.org/10.1080/17439884.2017.1278021

Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education*, *20*(2).

Eriksson, M., & Olsson, E. K. (2016). Facebook and Twitter in crisis communication: A comparative study of crisis communication professionals and citizens. *Journal of Contingencies and Crisis Management*, *24*(4), 198–208. https://doi.org/10.1111/1468-5973.12116

France, M. K., Finney, S. J., & Swerdzewski, P. (2010). Students' group and member attachment to their university: A construct validity study of the university attachment scale. *Educational and Psychological Measurement*, *70*(3), 440–458. https://doi.org/10.1177/0013164409344510

Gan, C., & Wang, W. (2015). Uses and gratifications of social media: a comparison of microblog and WeChat Article information. *Journal of Systems and Information Technology*, *17*(4), 351–363. https://doi.org/10.1108/JSIT-06-2015-0052

Gao, F., Luo, T., & Zhang, K. (2012). Tweeting for learning : A critical analysis of research on microblogging in education published in 2008-2011. *British Journal of Educational Technology*. https://doi.org/10.1111/j.1467-8535.2012.01357.x

Grimmer, J., Stewart, B. M., & Michael Alvarez, R. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028

Gruzd, A., Jacobson, J., Mai, P., & Dubois, E. (2018). *The State of Social Media in Canada 2017* (Vol. Version 1). https://doi.org/10.5683/SP/AL8Z6R

Guo, S., & Chase, M. (2011). Internationalisation of higher education: Integrating international students into Canadian academic environment. *Teaching in Higher Education*, *16*(3), 305–318. https://doi.org/10.1080/13562517.2010.546524

Hamid, S., Ijab, M. T., Sulaiman, H., Md. Anwar, R., & Norman, A. A. (2017). Social media for environmental sustainability awareness in higher education. *International Journal of Sustainability in Higher Education*. Emerald Group Publishing Ltd. https://doi.org/10.1108/IJSHE-01-2015-0010

Jones, G. (2009). Sectors, institutional types and the challenges of shifting categories: A Canadian commentary. *Higher Education Quarterly*. https://doi.org/10.1111/j.1468-2273.2009.00439.x

Jones, G. A. (2014). An introduction to higher education in Canada. In K. M. Joshi & S. Paivandi (Eds.), *Higher education across nations* (Vol. 1, pp. 1–38). Delhi: B.R. Publishing Corporation.

Junco, R., & Chickering, A. W. (2010). Civil discourse in the age of social media. *About Campus*, *15*(4), 12–18. https://doi.org/10.1002/abc.20030

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*, 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in instructional design, educational technology, and online learning research. *TechTrends*, *62*(5), 492–500. https://doi.org/10.1007/s11528-018-0307-4

Kimmons, R., Veletsianos, G., & Woodward, S. (2016). Institutional uses of Twitter in U.S. higher education, *42*, 97–111. https://doi.org/10.1007/s10755-016-9375-6

Korhan, O., Metin Ersoy, •, & Ersoy, M. (2016). Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. *Qual Quant*, *50*, 1799–1816. https://doi.org/10.1007/s11135-015-0236-7

Langone, A. (2018). Facebook’s Cambridge Analytica controversy could be big trouble for the social network. Here’s what to know. Retrieved October 5, 2018, from http://time.com/5205314/facebook-cambridge-analytica-breach/

Linvill, D. L., McGee, S. E., & Hicks, L. K. (2012). Colleges’ and universities’ use of Twitter: A content analysis. *Public Relations Review*, *38*(4), 636–638. https://doi.org/10.1016/j.pubrev.2012.05.010

Linvill, D., Mcgee, S. E., & Hicks, L. K. (2012). Colleges’ and universities’ use of Twitter: A content analysis. *Public Relations Review*, *38*(4), 636–638. Retrieved from http://www.sciencedirect.com/science/article/pii/S0363811112000811

Litwin, J. M. (2010). The secondary markets of higher education. *Higher Education Management and Policy*, *21*(3), 1–18. https://doi.org/10.1787/hemp-21-5ksb3fn6x1lt

Macnamara, J., & Zerfass, A. (2012). Social media communication in organizations: The challenges of balancing openness, strategy, and management. *International Journal of Strategic Communication*, *6*(4), 287–308. https://doi.org/10.1080/1553118X.2012.711402

Marwick, A. E., & boyd, danah. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media and Society*. https://doi.org/10.1177/1461444810365313

McAlexander, J. H., & F. Koenig, H. (2012). Building communities of philanthropy in higher education: contextual influences. *International Journal of Nonprofit and Voluntary Sector Marketing*, *17*(2), 122–131. https://doi.org/10.1002/nvsm.1415

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013, April). A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication. *Journal of Medical Internet Research*. https://doi.org/10.2196/jmir.1933

Neier, S., & Zayer, L. T. (2015). Students' perceptions and experiences of social media in higher education. *Journal of Marketing Education*, *37*(3), 133–143. https://doi.org/10.1177/0273475315583748

Nevzat, R., Amca, Y., Tanova, C., & Amca, H. (2016). Role of social media community in strengthening trust and loyalty for a university. *Computers in Human Behavior*, *65*, 550–559. https://doi.org/10.1016/j.chb.2016.09.018

Nielsen, F. A. (2011). AFINN. Kongens Lyngby: Informatics and Mathematical Modeling, Technical University of Denmark. Retrieved from http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

Oz, M., Zheng, P., Chen, G. M., & Park, R. H. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, *20*(9), 3400–3419. https://doi.org/10.1177/1461444817749516

Pasquini, L. A., & Evangelopoulos, N. (2017). Sociotechnical stewardship in higher education: a field study of social media policy documents. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-016-9130-0

Peruta, A., & Shields, A. B. (2017). Social media in higher education: understanding how colleges and universities use Facebook. *Journal of Marketing for Higher Education*, *27*(1), 131–143. https://doi.org/10.1080/08841241.2016.1212451

Quan-Haase, A., Martin, K., & Mccay-Peet, L. (2015). Networks of digital humanities scholars: The informational and social uses and gratifications of Twitter. https://doi.org/10.1177/2053951715589417

Radovanović, D., Hogan, B., & Lalić, D. (2015). Overcoming digital divides in higher education: Digital literacy beyond Facebook. *New Media and Society*, *17*(10), 1733–1749. https://doi.org/10.1177/1461444815588323

Rowe, I. (2015). Civility 2.0: a comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121–138. https://doi.org/10.1080/1369118X.2014.940365

Rutherford, C. (2010). Using online social media to support preservice student engagement. *MERLOT*

Journal of Online Learning and Teaching, *6*(4), 703–711.

Rutter, R., Roper, S., & Lettice, F. (2016). Social media interaction, the university brand and recruitment performance. *Journal of Business Research*, *69*(8), 3096–3104. https://doi.org/10.1016/j.jbusres.2016.01.025

Safko, L. (2012). Thumbs up for microblogging. What's in It for you? In *The social media bible: tactics, tools, and strategies for business success* (pp. 289–316). Hoboken, New Jersey: John Wiley & Sons, Incorporated. Retrieved from https://ebookcentral-proquest-com.ezproxy.royalroads.ca

Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment*, *84*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schroeder, A., Minocha, S., & Schneider, C. (2010). The strengths, weaknesses, opportunities and threats of using social software in higher and further education teaching and learning. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/j.1365-2729.2010.00347.x

Schultz, D. E., & Peltier, J. J. (2013). Social media's slippery slope: challenges, opportunities and future research directions. *Journal of Research in Interactive Marketing*, *7*(2), 86–99. https://doi.org/10.1108/JRIM-12-2012-0054

Shields, R. (2016). Following the leader? Network models of "world-class" universities on Twitter. *Higher Education*. https://doi.org/10.1007/s10734-015-9900-z

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach* (1st ed.). O'Reilly Media. Retrieved from https://www.tidytextmining.com/

Smith, A., & Anderson, M. (2018). *Social Media Use in 2018*. Pew Research Center. Retrieved from www.pewresearch.org.

Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, *27*(6), 2322–2329. https://doi.org/10.1016/j.chb.2011.07.011

Snelson, C. L. (2016). Special issue qualitative and mixed methods social media research: A review of the literature. *International Journal of Qualitative Methods*, *January-December*, 1–15. https://doi.org/10.1177/1609406915624574

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

Tan, A.-H. (1999). Text Mining: The state of the art and the challenges. Retrieved July 7, 2018, from https://pdfs.semanticscholar.org/9a80/ec16880ae43dc20c792ea3734862d85ba4d7.pdf

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual)-A literature review. *Computers in Human Behavior*, *29*(5). https://doi.org/10.1016/j.chb.2012.12.032

Tsugawa, S., & Ohsaki, H. (2017). On the relation between message sentiment and its virality on social media. *Social Network Analysis and Mining*, *7*(1). https://doi.org/10.1007/s13278-017-0439-0

Van Der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly*, *28*(4), 295–704. Retrieved from https://www.jstor.org/stable/25148660

Veletsianos, G., Kimmons, R., Shaw, A., Pasquini, L., & Woodward, S. (2017). Selective openness, branding, broadcasting, and promotion: Twitter use in Canada's public universities. *Educational*

Media International, *54*(1), 1–19. https://doi.org/10.1080/09523987.2017.1324363

Wang, S. (2018, July 27). Twitter plunges most in four years on drop in monthly users. *Bloomberg*. Retrieved from https://www.bloomberg.com/news/articles/2018-07-27/twitter-projects-users-to-decline-profit-short-of-estimates

Wells, B., & Spinks, N. (1999). Communicating with the community. *Career Development International*, *4*(2), 108–116. https://doi.org/10.1108/13620439910254759

Wheeler, S. (2009). Learning space mashups: Combining web 2.0 tools to create collaborative and reflective learning spaces. *Future Internet*. https://doi.org/10.3390/fi1010003

Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative Market Research: An International Journal*, *16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041

Wolny, J., & Mueller, C. (2013). Analysis of fashion consumers' motives to engage in electronic word-of-mouth communication through social media platforms. *Journal of Marketing Management*, *29*(5–6), 562–583. https://doi.org/10.1080/0267257X.2013.778324

Appendix A

University of Calgary

University of Alberta

University of Lethbridge

Athabasca University

MacEwan University

Mount Royal University

Northern Alberta Institute of Technology

Southern Alberta Institute of Technology

Bow Valley College

Grande Prairie Regional College

Keyano College

Lakeland College

Lethbridge College

Medicine Hat College

NorQuest College

Northern Lakes College

Olds College

Portage College

Red Deer College

Alberta Academy of Art and Design

Ambrose University

Burman University

Concordia University of Edmonton

The King’s University

St Mary’s University

Figures

Figure 1

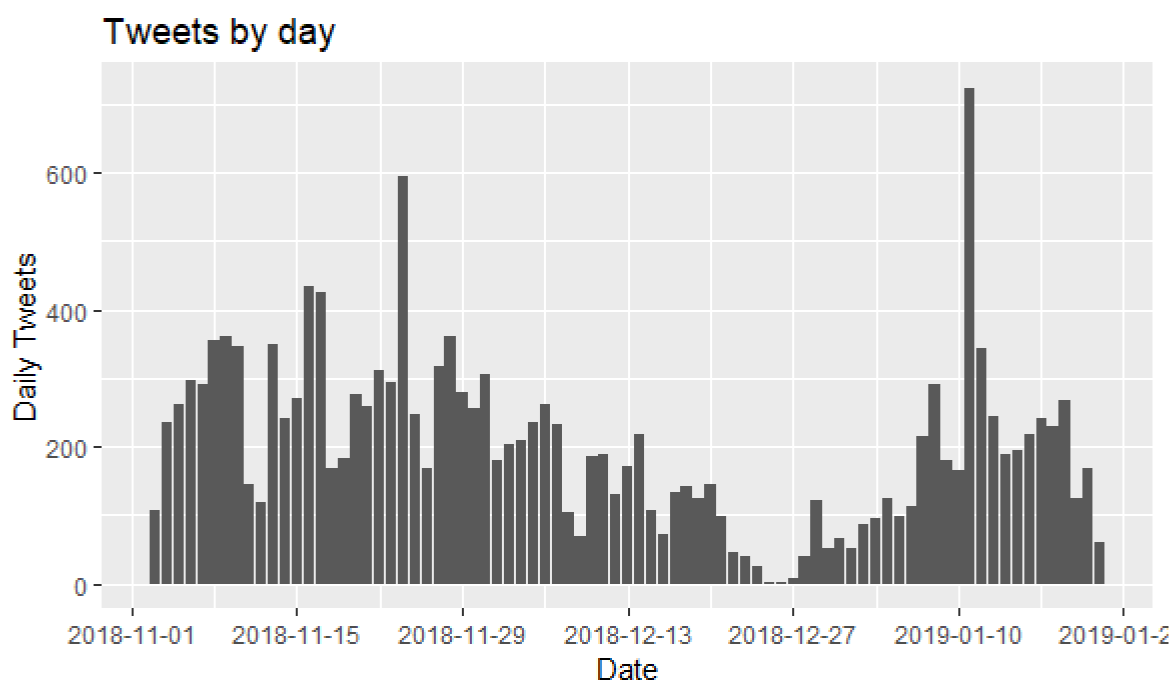


Figure 2

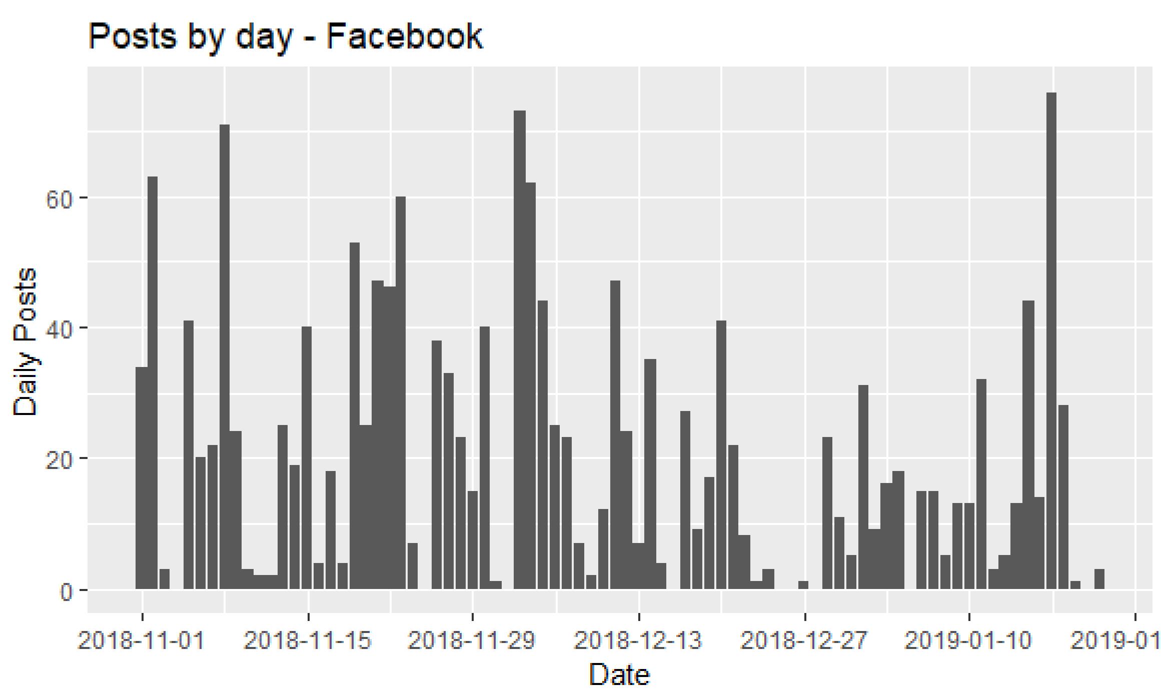
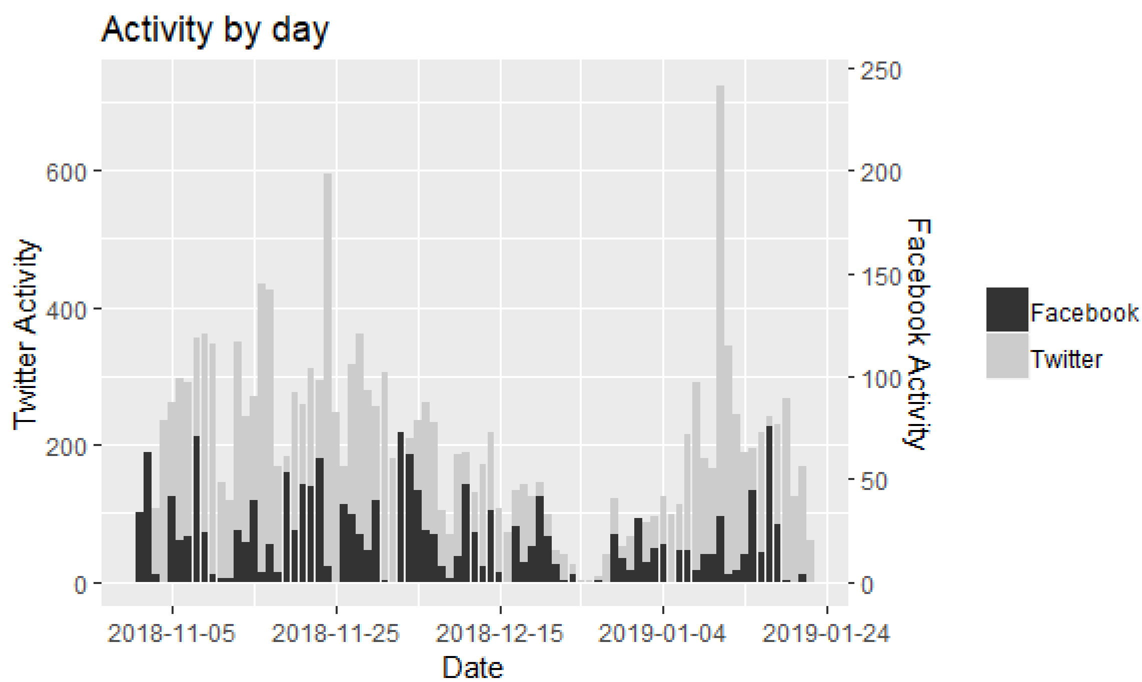


Figure 3


Activity by day
Twitter Activity
0
200
400
600
Facebook Activity
0
50
100
150
200
250
2018-11-05
2018-11-25
2018-12-15
2019-01-04
2019-01-24
Date
Facebook
Twitter

Tables

Table 1-Primary vs Secondary Account Twitter Activity

Institution Type	Primary Account	Sub Account	Ratio
BASI	1,508	828	1.82
CARI	4,834	2,034	2.38
CCI	2,781	1,758	1.58
IAI	502	265	1.89
Polytech	1,491	1,338	1.11
SACI	101	-	-

Table 2-Unique words sent to each institutional classification

Account	Unique Words
BASI	4,757
CARI	12,076
CCI	5,859
IAI	1,259
Multiple Sectors	711
Polytech	4,513
SACI	424
Primary Account	16,601
Sub Account	6,826

Table 3-Breakdown of Facebook posts by institutional classification

Institution Type	Posts
BASI	83
CARI	623
CCI	681
IAI	119
Polytechnic	115
SACI	49

Table 4-Common discussion drivers, Facebook

Institution Type	Word	Institution Posts	Replies
BASI	times	3	36
BASI	time	3	34
BASI	wasnt	3	32
BASI	people	3	30
CARI	campus	5	133
CARI	day	4	128
CARI	support	5	118
CARI	student	5	113
CARI	centre	4	106
CARI	community	4	106
CARI	university	5	100
CARI	holidays	3	65
CARI	family	3	64
CARI	health	3	63
CARI	alberta	3	53
CCI	students	4	51
CCI	canada	4	42
CCI	squash	4	40
CCI	centre	3	35
CCI	time	3	34
CCI	courts	3	30
IAI	justice	3	48
IAI	social	3	48

Table 5-Proportion of emotion expressed by institutional classification, Twitter (NRC)

Institution Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
BASI	0.071	0.252	0.023	0.078	0.187	0.075	0.091	0.222
CARI	0.061	0.232	0.031	0.073	0.182	0.059	0.088	0.275
CCI	0.059	0.25	0.025	0.063	0.213	0.069	0.091	0.229
IAI	0.088	0.236	0.053	0.094	0.159	0.067	0.109	0.193
Multiple Sectors	0.111	0.205	0.057	0.12	0.155	0.105	0.068	0.181
Polytech	0.066	0.238	0.029	0.066	0.195	0.073	0.089	0.243
SACI	0.037	0.26	0.018	0.027	0.196	0.096	0.137	0.228

Table 6-Ranking of emotion expressed by institutional classification, Twitter (NRC)

Institution Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
BASI	7	1	8	5	3	6	4	2
CARI	6	2	8	5	3	7	4	1
CCI	7	1	8	6	3	5	4	2
IAI	6	1	8	5	3	7	4	2
Multiple Sectors	5	1	8	4	3	6	7	2
Polytech	6	2	8	6	3	5	4	1
SACI	6	1	8	7	3	5	4	2

Table 7-Proportion of emotion expressed by account type, Twitter (NRC)

Account Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
Primary Account	0.056	0.242	0.027	0.063	0.204	0.057	0.093	0.257
Sub Account	0.076	0.24	0.031	0.082	0.169	0.085	0.087	0.23

Table 8-Ranking of emotion expressed by account type, Twitter (NRC)

Account Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
Primary Account	7	2	8	5	3	6	4	1
Sub Account	7	1	8	6	3	5	4	2

Table 9-Proportion of sentiment expressed by Twitter account type, Twitter (AFINN)

Institution Type	-5	-4	-3	-2	-1	1	2	3	4	5
BASI	-	0.005	0.052	0.074	0.07	0.146	0.3	0.176	0.172	0.005
CARI	-	0.005	0.033	0.08	0.055	0.167	0.352	0.19	0.107	0.01
CCI	-	0.002	0.043	0.054	0.068	0.14	0.29	0.221	0.177	0.004
IAI	-	0.003	0.076	0.035	0.06	0.164	0.271	0.186	0.202	0.003
Multiple Sectors	-	-	0.078	0.015	0.098	0.185	0.259	0.102	0.263	-
Polytech	0.001	0.005	0.038	0.052	0.088	0.155	0.265	0.209	0.184	0.003
SACI	-	-	0.048	0.016	0.032	0.365	0.27	0.143	0.127	-

Table 10-Ranking of sentiment expressed by account type, Twitter (AFINN)

Institution Type	-5	-4	-3	-2	-1	1	2	3	4	5
BASI	10	8	7	5	6	4	1	2	3	8
CARI	10	9	7	5	6	3	1	2	4	8
CCI	10	9	7	6	5	4	1	2	3	8
IAI	10	8	5	7	6	4	1	3	2	8
Multiple Sectors	8	9	6	7	5	3	2	4	1	10
Polytech	10	8	7	6	5	4	1	2	3	9
SACI	8	9	5	7	6	1	2	3	4	10

Table 11-Thirty most common terms with sentiment expressed

Word Rank	Word	Count	Sentiment
1	congratulations	397	2
2	win	331	4
3	congrats	313	2
4	amazing	270	4
5	love	219	3
6	top	214	2
7	awesome	202	4
8	excited	192	3
9	join	173	1
10	support	162	2
11	proud	158	2
12	happy	139	3
13	fun	124	4
14	exciting	117	3
15	innovation	112	1
16	opportunity	110	2
17	nice	106	3
18	hope	103	2
19	cool	101	1
20	free	97	1
21	award	91	3
22	wonderful	91	4
23	fantastic	87	4
24	glad	74	3
25	hard	73	-1

26	wow	73	4
27	share	63	1
28	drop	61	-1
29	miss	60	-2
30	beautiful	56	3

Table 12-Ranking of expressed sentiment by Twitter account type (AFINN)

Account Type	-5	-4	-3	-2	-1	1	2	3	4	5
Primary Account	10	9	7	5	6	3	1	2	4	8
Sub Account	10	9	7	6	5	4	1	3	2	8

Table 13- Proportion of emotion expressed by institutional classification, Facebook (NRC)

Institution Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
BASI	0.025	0.176	0.017	0.025	0.345	0.008	0.084	0.319
CARI	0.055	0.233	0.035	0.074	0.211	0.074	0.073	0.244
CCI	0.033	0.25	0.024	0.044	0.279	0.042	0.077	0.251
IAI	0.014	0.221	0.014	0.056	0.254	0.052	0.08	0.31
Polytechnic	0.059	0.191	0.046	0.053	0.217	0.092	0.099	0.243
SACI	0.036	0.273	0.036	0.109	0.182	0.073	0.091	0.2

Table 14-Ranking of emotion expressed by institutional classification, Facebook (NRC)

Institution Type	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
BASI	5	3	7	5	1	8	4	2
CARI	7	2	8	4	3	4	6	1
CCI	7	3	8	5	1	6	4	2
IAI	7	3	7	5	2	6	4	1
Polytechnic	6	3	8	7	2	5	4	1
SACI	7	1	7	4	3	6	5	2

Table 15-Proportion of sentiment expressed by Facebook account type (AFINN)

Institution Type	-4	-3	-2	-1	1	2	3	4	5
BASI	-	-	-	0.012	0.176	0.341	0.294	0.176	-
CARI	0.005	0.034	0.129	0.074	0.122	0.33	0.185	0.122	-
CCI	0.002	0.017	0.065	0.044	0.112	0.321	0.247	0.19	0.002

IAI	-	0.022	0.033	0.022	0.143	0.462	0.242	0.077	-
Polytechnic	0.027	0.027	0.135	0.068	0.189	0.216	0.176	0.162	-
SACI	0.029	0.029	0.029	0.029	0.176	0.353	0.206	0.147	-

Table 16-Ranking of expressed sentiment by Facebook account type (AFINN)

Institution Type	-4	-3	-2	-1	1	2	3	4	5
BASI	6	7	8	5	3	1	2	3	9
CARI	8	7	3	6	4	1	2	4	9
CCI	8	7	5	6	4	1	2	3	8
IAI	8	6	5	6	3	1	2	4	9
Polytechnic	7	7	5	6	2	1	3	4	9
SACI	5	5	5	5	3	1	2	4	9

Table 17-Expected 15 terms composing each topic (6 topics)

Estimated Topic	Term	Beta
1	acac_sport	0.012322
1	set	0.010606
1	students	0.009592
1	game	0.009163
1	wvb	0.007331
1	team	0.007253
1	mvb	0.006551
1	daniellelarivee	0.005888
1	amp	0.005732
1	win	0.005654
1	wbb	0.005498
1	day	0.005342
1	tonight	0.005303
1	wearerustlers	0.005108
1	mbb	0.004796
2	dr	0.015592
2	congrats	0.011597
2	students	0.00948
2	alberta	0.007263
2	university	0.006967
2	amp	0.006291
2	lawdeanholloway	0.004767
2	nserc_crsng	0.004732
2	student	0.004447
2	congratulations	0.004238

2	check	0.004046
2	join	0.00392
2	free	0.003836
2	day	0.003733
2	canadian	0.003693
3	students	0.013262
3	set	0.006322
3	amp	0.005703
3	whky	0.005016
3	wearethecougars	0.004604
3	tonight	0.004329
3	mru	0.004192
3	calgary	0.004054
3	day	0.003985
3	samrubuzz	0.003848
3	win	0.003848
3	student	0.003779
3	mvb	0.003436
3	duanebratt	0.003367
3	cougars	0.00323
3	drfinchy	0.00323
3	wvb	0.00323
4	acac_sport	0.02012
4	set	0.011299
4	game	0.009858
4	acacsport	0.008648
4	students	0.008013
4	broncos	0.005938
4	team	0.005823
4	ooks	0.005765
4	lead	0.005592
4	horsepower	0.005189
4	mvb	0.005131
4	huskies	0.005073
4	ookswhky	0.0049
4	day	0.004785
4	win	0.004727
5	acac_sport	0.025459
5	game	0.012654
5	students	0.011901
5	win	0.011901
5	rdcathletics	0.010395
5	mhc_rattlers	0.008888
5	briercrestwvb	0.008135
5	tonight	0.007382
5	wearerustlers	0.007231

5	mbb	0.00708
5	acac	0.005423
5	calgaryartsdev	0.005423
5	calgarylibrary	0.005423
5	nmc_canada	0.005423
5	skinakiska	0.005423
5	winsportcanada	0.005423
6	congratulations	0.006281
6	cbcnorth	0.005422
6	research	0.005394
6	students	0.00518
6	amp	0.005013
6	top	0.004405
6	caulfieldtim	0.004255
6	rachelnotley	0.004055
6	nserc_crsng	0.003837
6	de	0.003424
6	time	0.003217
6	lawdeanholloway	0.002679
6	day	0.002643
6	canada	0.002488
6	ualberta_fomd	0.002392

Table 18-Expected 15 terms composing each topic (2 topics)

Estimated Topic	Term	Beta
1	students	0.011034
1	amp	0.006262
1	day	0.00423
1	congratulations	0.004049
1	dr	0.004019
1	research	0.004019
1	student	0.003838
1	congrats	0.003327
1	amazing	0.003146
1	alberta	0.003041
1	time	0.003026
1	nserc_crsng	0.002755
1	love	0.002649
1	team	0.002499
1	cbcnorth	0.002439

1	excited	0.002423
1	daniellelarivee	0.002303
1	rachelnotley	0.002243
1	awesome	0.002228
1	week	0.002228
2	set	0.016902
2	acac_sport	0.015044
2	game	0.013546
2	wvb	0.00959
2	mvb	0.00959
2	win	0.008361
2	mbb	0.007252
2	wbb	0.007222
2	lead	0.006983
2	team	0.006773
2	acacsport	0.006473
2	tonight	0.006174
2	match	0.005364
2	huskies	0.004945
2	top	0.004915
2	whky	0.004495
2	volleyball	0.004196
2	lawdeanholloway	0.004166
2	fall	0.004106
2	quarter	0.004046

Table 19-Content differences by log-odds ratio, Twitter (CARI vs BASI)

Word	CARI Ratio	BASI Ratio	Log-Odds Ratio
congratulations	0.0218860	0.0003748	5.86769
dr	0.0211655	0.0003748	5.81940
nserc_crsng	0.0161218	0.0003748	5.42670
cbcnorth	0.0146807	0.0003748	5.29161
caulfieldtim	0.0135999	0.0003748	5.18129
lawdeanholloway	0.0130595	0.0003748	5.12279
rachelnotley	0.0119787	0.0003748	4.99816
de	0.0117085	0.0003748	4.96525
alberta	0.0114383	0.0003748	4.93157
university	0.0086463	0.0003748	4.52784

news	0.0082860	0.0003748	4.46644
canada	0.0081059	0.0003748	4.43473
researchers	0.0072053	0.0003748	4.26481
ualberta_fomd	0.0066649	0.0003748	4.15233
drtblew	0.0064847	0.0003748	4.11281
join	0.0062145	0.0003748	4.05141
professor	0.0060344	0.0003748	4.00897
science	0.0060344	0.0003748	4.00897
president	0.0058543	0.0003748	3.96525
ucalgarymed	0.0058543	0.0003748	3.96525
check	0.0056741	0.0003748	3.92016
school	0.0054940	0.0003748	3.87362
talk	0.0054039	0.0003748	3.84977
hear	0.0053139	0.0003748	3.82552
its	0.0053139	0.0003748	3.82552
sustainability	0.0053139	0.0003748	3.82552
future	0.0052238	0.0003748	3.80086
ubc	0.0052238	0.0003748	3.80086
award	0.0051337	0.0003748	3.77577
read	0.0051337	0.0003748	3.77577
tonight	0.0051337	0.0239880	-2.22423
final	0.0024318	0.0127436	-2.38969
weekend	0.0024318	0.0131184	-2.43151
win	0.0037828	0.0213643	-2.49769
set	0.0030622	0.0348576	-3.50881
loss	0.0000901	0.0101199	-6.81201
performance	0.0000901	0.0101199	-6.81201
home	0.0000901	0.0104948	-6.86447
rebs	0.0000901	0.0104948	-6.86447
volleyball	0.0000901	0.0104948	-6.86447
match	0.0000901	0.0108696	-6.91510
lead	0.0000901	0.0112444	-6.96401
mins	0.0000901	0.0112444	-6.96401
canball	0.0000901	0.0116192	-7.01132
mrucougars's	0.0000901	0.0119940	-7.05712
pts	0.0000901	0.0119940	-7.05712
wolfpack	0.0000901	0.0127436	-7.14458
lesleyabrown	0.0000901	0.0134933	-7.22704
class	0.0000901	0.0138681	-7.26657
lukaszukab	0.0000901	0.0146177	-7.34252
goal	0.0000901	0.0153673	-7.41467

cougars	0.0000901	0.0179910	-7.64208
drfinchy	0.0000901	0.0179910	-7.64208
wvb	0.0000901	0.0179910	-7.64208
duanebratt	0.0000901	0.0187406	-7.70098
mvb	0.0000901	0.0191154	-7.72954
samrubuzz	0.0000901	0.0213643	-7.89001
mru	0.0000901	0.0232384	-8.01132
wearethecougars	0.0000901	0.0254873	-8.14458
whky	0.0000901	0.0277361	-8.26657

Table 20-Content Differences by log-odds ratio, Twitter (CARI vs CCI)

Word	CARI Ratio	CCI Ratio	Log-Odds Ratio
dr	0.0211655	0.0001032	7.68073
nserc_crsng	0.0161218	0.0001032	7.28803
cbcnorth	0.0146807	0.0001032	7.15294
caulfieldtim	0.0135999	0.0001032	7.04262
lawdeanholloway	0.0130595	0.0001032	6.98412
top	0.0127893	0.0001032	6.95396
rachelnotley	0.0119787	0.0001032	6.85950
de	0.0117085	0.0001032	6.82658
university	0.0086463	0.0001032	6.38918
news	0.0082860	0.0001032	6.32778
researchers	0.0072053	0.0001032	6.12614
ualberta_fomd	0.0066649	0.0001032	6.01367
drtblew	0.0064847	0.0001032	5.97414
calgary	0.0063947	0.0001032	5.95396
people	0.0063947	0.0001032	5.95396
professor	0.0060344	0.0001032	5.87030
science	0.0060344	0.0001032	5.87030
president	0.0058543	0.0001032	5.82658
ucalgarymed	0.0058543	0.0001032	5.82658
talk	0.0054039	0.0001032	5.71111
hear	0.0053139	0.0001032	5.68686
sustainability	0.0053139	0.0001032	5.68686
ubc	0.0052238	0.0001032	5.66220
award	0.0051337	0.0001032	5.63710
read	0.0051337	0.0001032	5.63710
yeg	0.0050437	0.0001032	5.61157
canadawest	0.0049536	0.0001032	5.58557
hope	0.0049536	0.0001032	5.58557

innovation	0.0049536	0.0001032	5.58557
health	0.0048636	0.0001032	5.55910
score	0.0000901	0.0066020	-6.19579
games	0.0000901	0.0068083	-6.24018
women's	0.0000901	0.0068083	-6.24018
division	0.0000901	0.0069115	-6.26187
men's	0.0000901	0.0071178	-6.30431
quarter	0.0000901	0.0072210	-6.32507
half	0.0000901	0.0073241	-6.34553
kodiaks	0.0000901	0.0073241	-6.34553
players	0.0000901	0.0076336	-6.40524
acacsport	0.0000901	0.0078399	-6.44371
mikemccready	0.0000901	0.0080462	-6.48119
regional	0.0000901	0.0080462	-6.48119
college	0.0000901	0.0084588	-6.55334
huskies	0.0000901	0.0087683	-6.60518
basketball	0.0000901	0.0089746	-6.63873
curling	0.0000901	0.0095936	-6.73494
rdcathletics	0.0000901	0.0095936	-6.73494
stmarysu	0.0000901	0.0096967	-6.75037
volleyball	0.0000901	0.0097999	-6.76564
lead	0.0000901	0.0102125	-6.82514
match	0.0000901	0.0106251	-6.88229
fall	0.0000901	0.0110378	-6.93725
mhc_rattlers	0.0000901	0.0110378	-6.93725
acac	0.0000901	0.0120693	-7.06615
mbb	0.0000901	0.0127914	-7.14998
wearerustlers	0.0000901	0.0136167	-7.24018
daniellelarivee	0.0000901	0.0156798	-7.44371
mvb	0.0000901	0.0174335	-7.59666
wvb	0.0000901	0.0194966	-7.75803
acac_sport	0.0000901	0.0327006	-8.50412

Table 21-Content differences by log-odds ratio, Twitter (CCI vs Polytechnic)

Word	CCI Ratio	Polytech Ratio	Log-Odds Ratio
daniellelarivee	0.0156798	0.0001881	6.38091
wearerustlers	0.0136167	0.0001881	6.17737
acac	0.0120693	0.0001881	6.00334
stmarysu	0.0096967	0.0001881	5.68757

basketball	0.0089746	0.0001881	5.57592
college	0.0084588	0.0001881	5.49053
mikemccready	0.0080462	0.0001881	5.41838
players	0.0076336	0.0001881	5.34243
kodiaks	0.0073241	0.0001881	5.28272
games	0.0068083	0.0001881	5.17737
alberta	0.0066020	0.0001881	5.13298
ccaa	0.0064989	0.0001881	5.11026
learning	0.0064989	0.0001881	5.11026
vrara_alberta	0.0061894	0.0001881	5.03987
gprcwolves	0.0058799	0.0001881	4.96587
join	0.0057768	0.0001881	4.94033
watch	0.0057768	0.0001881	4.94033
mergevr	0.0056736	0.0001881	4.91434
research	0.0056736	0.0001881	4.91434
action	0.0053641	0.0001881	4.83342
underway	0.0053641	0.0001881	4.83342
women	0.0052610	0.0001881	4.80540
canada	0.0051578	0.0001881	4.77683
sets	0.0050547	0.0001881	4.74769
fourth	0.0049515	0.0001881	4.71794
arstories	0.0048484	0.0001881	4.68757
check	0.0046420	0.0001881	4.62483
mixed	0.0046420	0.0001881	4.62483
takes	0.0046420	0.0001881	4.62483
proud	0.0045389	0.0001881	4.59241
ooks	0.0046420	0.0190028	-2.03338
broncos	0.0037136	0.0195673	-2.39754
ookswhky	0.0028884	0.0161806	-2.48593
marsden	0.0001032	0.0050800	-5.62191
ocbroncoswhky	0.0001032	0.0050800	-5.62191
rdcqueenshockey	0.0001032	0.0050800	-5.62191
greycupfestival	0.0001032	0.0052681	-5.67438
hockey	0.0001032	0.0052681	-5.67438
ookswsoc	0.0001032	0.0054563	-5.72500
nait_ohaa	0.0001032	0.0056444	-5.77391
yeg	0.0001032	0.0056444	-5.77391
macewangriffins	0.0001032	0.0058325	-5.82122
yyc	0.0001032	0.0058325	-5.82122
left	0.0001032	0.0063970	-5.95449
goal	0.0001032	0.0065851	-5.99631

innovation	0.0001032	0.0067733	-6.03695
calgaryartsdev	0.0001032	0.0069614	-6.07648
nmc_canada	0.0001032	0.0069614	-6.07648
skinakiska	0.0001032	0.0069614	-6.07648
winsportcanada	0.0001032	0.0069614	-6.07648
calgarylibrary	0.0001032	0.0073377	-6.15242
calls	0.0001032	0.0075259	-6.18895
saitsa	0.0001032	0.0075259	-6.18895
saitalumni	0.0001032	0.0077140	-6.22457
naitsa	0.0001032	0.0079022	-6.25934
naitjrshaw	0.0001032	0.0084666	-6.35888
coach	0.0001032	0.0092192	-6.48173
timeout	0.0001032	0.0094073	-6.51088
naitmenshockey	0.0001032	0.0122295	-6.88939
horsepower	0.0001032	0.0171214	-7.37482

Table 22-Content differences by log-odds ratio, Twitter (big schools vs small schools)

Word	Large Schools Ratio	Small Schools Ratio	Log-Odds Ratio
dr	0.0123140	0.0001003	6.94054
top	0.0097988	0.0001003	6.61092
nserc_crsng	0.0093796	0.0001003	6.54784
cbcnorth	0.0085412	0.0001003	6.41275
caulfieldtim	0.0079124	0.0001003	6.30243
lawdeanholloway	0.0075980	0.0001003	6.24393
rachelnotley	0.0069692	0.0001003	6.11931
calgary	0.0068120	0.0001003	6.08639
de	0.0068120	0.0001003	6.08639
calgarylibrary	0.0067072	0.0001003	6.06403
nmc_canada	0.0062880	0.0001003	5.97092
winsportcanada	0.0061832	0.0001003	5.94667
calgaryartsdev	0.0059736	0.0001003	5.89692
skinakiska	0.0058688	0.0001003	5.87138
people	0.0052924	0.0001003	5.72224
yyc	0.0052400	0.0001003	5.70788
canadawest	0.0050304	0.0001003	5.64899

university	0.0050304	0.0001003	5.64899
news	0.0048208	0.0001003	5.58759
horsepower	0.0047684	0.0001003	5.57182
innovation	0.0047160	0.0001003	5.55588
yeg	0.0044540	0.0001003	5.47342
ucdinos	0.0042968	0.0001003	5.42158
researchers	0.0041920	0.0001003	5.38595
conference	0.0040872	0.0001003	5.34943
goal	0.0039300	0.0001003	5.29284
ualberta_fomd	0.0038776	0.0001003	5.27348
drtblew	0.0037728	0.0001003	5.23395
wearethecougars	0.0035632	0.0001003	5.15149
professor	0.0035108	0.0001003	5.13011
acacvolleyball	0.0000524	0.0037093	-6.14543
gary	0.0000524	0.0037093	-6.14543
player	0.0000524	0.0037093	-6.14543
kings	0.0000524	0.0038095	-6.18390
start	0.0000524	0.0038095	-6.18390
cathyhackl	0.0000524	0.0042105	-6.32829
rdckingsvb	0.0000524	0.0043108	-6.36224
mixed	0.0000524	0.0045113	-6.42783
takes	0.0000524	0.0045113	-6.42783
arstories	0.0000524	0.0047118	-6.49056
fourth	0.0000524	0.0048120	-6.52094
sets	0.0000524	0.0049123	-6.55068
action	0.0000524	0.0052130	-6.63641
underway	0.0000524	0.0052130	-6.63641
mergevr	0.0000524	0.0055138	-6.71733
gprcwolves	0.0000524	0.0057143	-6.76886
vrara_alberta	0.0000524	0.0060150	-6.84287
briercrestwvb	0.0000524	0.0062155	-6.89017
ccaa	0.0000524	0.0063158	-6.91325
games	0.0000524	0.0066165	-6.98037
cue_athletics	0.0000524	0.0068170	-7.02344
kodiaks	0.0000524	0.0071178	-7.08572
players	0.0000524	0.0074185	-7.14543
mikemccready	0.0000524	0.0078195	-7.22138
college	0.0000524	0.0082206	-7.29353
basketball	0.0000524	0.0087218	-7.37892
stmarysu	0.0000524	0.0094236	-7.49056
acac	0.0000524	0.0117293	-7.80634

wearerustlers	0.0000524	0.0132331	-7.98037
daniellelarivee	0.0000524	0.0152381	-8.18390

Figure 23-Content differences by log-odds ratio, Twitter (primary vs secondary accounts)

Word	Primary Account Ratio	Secondary Account Ratio	Log-Odds Ratio
dr	0.0135237	0.0000726	7.54056
research	0.0135237	0.0000726	7.54056
alberta	0.0102437	0.0000726	7.13981
nserc_crsng	0.0092850	0.0000726	6.99804
cbcnorth	0.0082253	0.0000726	6.82320
daniellelarivee	0.0077711	0.0000726	6.74126
rachelnotley	0.0075693	0.0000726	6.70329
join	0.0074179	0.0000726	6.67415
campus	0.0068628	0.0000726	6.56194
event	0.0068123	0.0000726	6.55129
proud	0.0066105	0.0000726	6.50790
university	0.0065096	0.0000726	6.48570
program	0.0064086	0.0000726	6.46316
school	0.0063582	0.0000726	6.45176
learning	0.0063077	0.0000726	6.44026
news	0.0061059	0.0000726	6.39334
forward	0.0057526	0.0000726	6.30737
people	0.0057526	0.0000726	6.30737
morning	0.0056517	0.0000726	6.28183
happy	0.0056013	0.0000726	6.26889
yeg	0.0055003	0.0000726	6.24266
class	0.0054499	0.0000726	6.22936
learn	0.0052985	0.0000726	6.18872
faculty	0.0051976	0.0000726	6.16098
innovation	0.0049957	0.0000726	6.10383
community	0.0048443	0.0000726	6.05944
opportunity	0.0047939	0.0000726	6.04433
sharing	0.0047434	0.0000726	6.02906
stmarysu	0.0047434	0.0000726	6.02906
education	0.0046929	0.0000726	6.01363
trail	0.0000505	0.0063195	-6.96847
women's	0.0000505	0.0063195	-6.96847
ookswhky	0.0000505	0.0066100	-7.03332
horsepower	0.0000505	0.0069732	-7.11049

play	0.0000505	0.0069732	-7.11049
players	0.0000505	0.0069732	-7.11049
regional	0.0000505	0.0072637	-7.16938
underway	0.0000505	0.0072637	-7.16938
basketball	0.0000505	0.0077722	-7.26699
mhc_rattlers	0.0000505	0.0082080	-7.34570
score	0.0000505	0.0084259	-7.38351
ooks	0.0000505	0.0085712	-7.40817
half	0.0000505	0.0087165	-7.43241
men's	0.0000505	0.0088618	-7.45626
acac	0.0000505	0.0089344	-7.46804
curling	0.0000505	0.0090070	-7.47972
broncos	0.0000505	0.0090797	-7.49131
weekend	0.0000505	0.0094429	-7.54789
wearerustlers	0.0000505	0.0095881	-7.56992
quarter	0.0000505	0.0098787	-7.61299
lawdeanholloway	0.0000505	0.0101692	-7.65481
whky	0.0000505	0.0109683	-7.76393
huskies	0.0000505	0.0120578	-7.90056
match	0.0000505	0.0130747	-8.01738
acacsport	0.0000505	0.0157623	-8.28708
lead	0.0000505	0.0169972	-8.39589
wbb	0.0000505	0.0175783	-8.44439
mbb	0.0000505	0.0176509	-8.45034
mvb	0.0000505	0.0233166	-8.85195
wvb	0.0000505	0.0233166	-8.85195

Table 24-Content differences by log-odds ratio, Facebook (CARI vs CCI)

Word	CARI Ratio	CCI Ratio	Log-Odds Ratio
university	0.0364299	0.0013298	4.77585
courses	0.0273224	0.0013298	4.36082
semester	0.0273224	0.0013298	4.36082
holidays	0.0236794	0.0013298	4.15437
sad	0.0236794	0.0013298	4.15437
study	0.0218579	0.0013298	4.03889
start	0.0200364	0.0013298	3.91336
2	0.0182149	0.0013298	3.77585
article	0.0182149	0.0013298	3.77585
favourite	0.0182149	0.0013298	3.77585
qualified	0.0182149	0.0013298	3.77585

reading	0.0182149	0.0013298	3.77585
hours	0.0163934	0.0013298	3.62385
omnivore	0.0163934	0.0013298	3.62385
read	0.0163934	0.0013298	3.62385
true	0.0163934	0.0013298	3.62385
friends	0.0145719	0.0013298	3.45393
helped	0.0145719	0.0013298	3.45393
professor	0.0145719	0.0013298	3.45393
scholarship	0.0145719	0.0013298	3.45393
assignments	0.0127505	0.0013298	3.26128
change	0.0127505	0.0013298	3.26128
cool	0.0127505	0.0013298	3.26128
discrimination	0.0127505	0.0013298	3.26128
eat	0.0127505	0.0013298	3.26128
found	0.0127505	0.0013298	3.26128
friend	0.0127505	0.0013298	3.26128
guba	0.0127505	0.0013298	3.26128
happened	0.0127505	0.0013298	3.26128
indigenous	0.0127505	0.0013298	3.26128
bow	0.0018215	0.0093085	-2.35343
christmas	0.0018215	0.0093085	-2.35343
dorm	0.0018215	0.0093085	-2.35343
fantastic	0.0018215	0.0093085	-2.35343
norquest	0.0018215	0.0093085	-2.35343
training	0.0018215	0.0093085	-2.35343
valley	0.0018215	0.0093085	-2.35343
choose	0.0018215	0.0106383	-2.54607
future	0.0018215	0.0106383	-2.54607
luck	0.0018215	0.0106383	-2.54607
story	0.0018215	0.0106383	-2.54607
success	0.0018215	0.0106383	-2.54607
continue	0.0018215	0.0119681	-2.71600
home	0.0018215	0.0119681	-2.71600
life	0.0018215	0.0119681	-2.71600
loved	0.0018215	0.0119681	-2.71600
super	0.0018215	0.0119681	-2.71600
campus	0.0018215	0.0132979	-2.86800
remember	0.0018215	0.0132979	-2.86800
experience	0.0018215	0.0146277	-3.00551
post	0.0018215	0.0146277	-3.00551
wonderful	0.0018215	0.0146277	-3.00551

chose	0.0018215	0.0159574	-3.13104
olds	0.0018215	0.0159574	-3.13104
beautiful	0.0018215	0.0186170	-3.35343
education	0.0018215	0.0186170	-3.35343
job	0.0018215	0.0345745	-4.24651
rdc	0.0018215	0.0452128	-4.63354
congrats	0.0018215	0.0465426	-4.67536
college	0.0018215	0.0518617	-4.83148

Table 25-Content differences by log-odds ratio, Facebook (Big schools vs small schools)

Word	Large Schools Ratio	Small Schools Ratio	Log-Odds Ratio
courses	0.0247117	0.0012903	4.25939
semester	0.0247117	0.0012903	4.25939
holidays	0.0214168	0.0012903	4.05294
sad	0.0214168	0.0012903	4.05294
study	0.0197694	0.0012903	3.93746
nait	0.0181219	0.0012903	3.81193
start	0.0181219	0.0012903	3.81193
2	0.0164745	0.0012903	3.67443
article	0.0164745	0.0012903	3.67443
favourite	0.0164745	0.0012903	3.67443
qualified	0.0164745	0.0012903	3.67443
reading	0.0164745	0.0012903	3.67443
hours	0.0148270	0.0012903	3.52242
omnivore	0.0148270	0.0012903	3.52242
read	0.0148270	0.0012903	3.52242
true	0.0148270	0.0012903	3.52242
bacon	0.0131796	0.0012903	3.35250
friends	0.0131796	0.0012903	3.35250
helped	0.0131796	0.0012903	3.35250
professor	0.0131796	0.0012903	3.35250
scholarship	0.0131796	0.0012903	3.35250
assignments	0.0115321	0.0012903	3.15985
change	0.0115321	0.0012903	3.15985
cool	0.0115321	0.0012903	3.15985
discrimination	0.0115321	0.0012903	3.15985
eat	0.0115321	0.0012903	3.15985
found	0.0115321	0.0012903	3.15985
friend	0.0115321	0.0012903	3.15985
guba	0.0115321	0.0012903	3.15985
happened	0.0115321	0.0012903	3.15985
bow	0.0016474	0.0090323	-2.45486
christmas	0.0016474	0.0090323	-2.45486
dorm	0.0016474	0.0090323	-2.45486
fantastic	0.0016474	0.0090323	-2.45486
norquest	0.0016474	0.0090323	-2.45486
training	0.0016474	0.0090323	-2.45486
valley	0.0016474	0.0090323	-2.45486

choose	0.0016474	0.0103226	-2.64750
future	0.0016474	0.0103226	-2.64750
luck	0.0016474	0.0103226	-2.64750
story	0.0016474	0.0103226	-2.64750
success	0.0016474	0.0103226	-2.64750
continue	0.0016474	0.0116129	-2.81743
home	0.0016474	0.0116129	-2.81743
life	0.0016474	0.0116129	-2.81743
loved	0.0016474	0.0116129	-2.81743
super	0.0016474	0.0116129	-2.81743
campus	0.0016474	0.0129032	-2.96943
remember	0.0016474	0.0129032	-2.96943
experience	0.0016474	0.0141935	-3.10693
post	0.0016474	0.0141935	-3.10693
wonderful	0.0016474	0.0141935	-3.10693
chose	0.0016474	0.0154839	-3.23246
olds	0.0016474	0.0154839	-3.23246
beautiful	0.0016474	0.0180645	-3.45486
education	0.0016474	0.0180645	-3.45486
job	0.0016474	0.0335484	-4.34794
rdc	0.0016474	0.0438710	-4.73496
congrats	0.0016474	0.0451613	-4.77678
college	0.0016474	0.0503226	-4.93290

Figure 26-Pair-wise correlations: context and redundancy check

Item 1	Item 2	Correlation
cbcnorth	nserc_crsng	0.9398
curling	regional	0.9277
curling	division	0.8161
acac	wearerustlers	0.7976
division	regional	0.7966
curling	fall	0.7183
fall	regional	0.6934
acacsport	horsepower	0.6539
broncos	horsepower	0.6271
wearethecougars	whky	0.6149
division	fall	0.6080
acac	ccaa	0.5171
set	wvb	0.4644

ccaa	wearerustlers	0.4632
acacsport	broncos	0.4097
mhc_rattlers	rdcathletics	0.3971
acac_sport	horsepower	0.3955
acac_sport	curling	0.3912
mvb	set	0.3880
acac_sport	regional	0.3850
de	top	0.3663
players	game	0.3542
acac_sport	division	0.3358
final	score	0.3229
naitmenshockey	ooks	0.3174
players	wearerustlers	0.3117
acac_sport	fall	0.3096
curling	men's	0.2914
men's	regional	0.2840
division	women's	0.2729
acac_sport	broncos	0.2718
lead	wbb	0.2715
match	set	0.2590
acac_sport	acacsport	0.2545

Table 27-Overlapping common AFINN terms, Twitter vs Facebook

Twitter Word	Word Rank	Facebook Word
congratulations	1	congratulations
win	2	awesome
congrats	3	love
amazing	4	congrats
love	5	amazing
top	6	proud
awesome	7	wow
excited	8	lol
join	9	fun
support	10	cool
proud	11	beautiful
happy	12	happy
fun	13	free
exciting	14	nice
innovation	15	miss
opportunity	16	hard
nice	17	support
hope	18	wonderful
cool	19	sad
free	20	hope
award	21	true
wonderful	22	super
fantastic	23	god
glad	24	fantastic
hard	25	pretty
wow	26	loved
share	27	lovely
drop	28	enjoy
miss	29	excellent
beautiful	30	glad

Table 28-Simple regression, Twitter post volume vs Facebook post volume

	Dependent Variable:		
		Twitter	
Facebook		1.433**	
		(0.705)	
Constant		167.411***	
		(19.905)	
Observations:		83	
R Squared		0.049	
Adjusted R Squared		0.037	
Residual Std. Error		125.902	(df = 81)
F Statistic		4.133**	(df = 1; 81)
Note:	*p<0.1	**p<0.05	***p<0.01