

EXPLORING ENGINEERING STUDENTS' COLLEGE EXPERIENCES USING SOCIAL MEDIA MONITORING TOOL RADIAN6

¹Xin Chen, ²Mihaela Vorvoreanu, ¹Krishna Madhavan
School of Engineering Education, Computer Graphics Technology
Purdue University

Abstract

Students nowadays talk about their college life in various online social media platforms such as Twitter, Facebook, and many blogs and forums. Public discourse on these social media sites can provide insights into students' college experiences. This paper focuses on exploring the informal social media data created by students in their daily life, rather than investigating the formal usage of social media platforms in the classroom settings. Many social media monitoring tools are developed for monitoring business brands or public events. These tools may also be used to collect social media data for educational research and policy recommendation purposes. This study uses social media monitoring tool Radian6 to collect data from Twitter about engineering students' college experiences. The data are analyzed both in Radian6 and manually using qualitative content analysis. The results shed light into the academic context and social context of engineering students' learning experiences. This exploratory study also considers the potential of social media analytics tools for engineering education research. Social media analytics tools specifically for educational purposes need to be developed in the future.

Introduction

Maintaining the U.S. scientific and technical talent pool has been a top priority in higher education. In 2005, the National Academies report "*Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*" [1] proposed four recommendations to enhance American science and technology competency in 21st century. One of the four recommendations is concerned with engineering higher education, suggesting to "[m]ake United States the most *attractive* setting in which to study...so that we can develop, recruit, and retain

the best and brightest students..." (p. 162). Questions remain how *attractive* educational settings are to engineering students nowadays, how pleasant students' experiences are in their current educational settings, and how to make them more attractive in the future. Scholarly research in engineering education has been done aiming to retain engineering students and enhance their success. However, despite all the efforts during the past years, a 2010 revisit of the 2005 report entitled "*Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*" [2] stated that the U. S. scientific and technical global competitiveness position had continued to deteriorate in the previous five years (p. 2).

Meanwhile, social media sites such as Twitter and Facebook have become important venues for the young generation to communicate and exchange information. A Pew Research Center report [3] indicates 93% of young adults (age 18-29) use the Internet and 72% of them use social networking sites. The public conversation online has accumulated large amount of real-time data generated in informal settings that can bring valuable insights into students' college life. However, these data remain unexplored by engineering education researchers and educational policy makers.

Many social media analytics tools such as Radian6 (radian6.com), Visible Intelligence (visibletechnologies.com), and NodeXL (nodexl.codeplex.com), have been developed to collect social media data, and then use these data to monitor business brands and public events, or analyze the topologies of social networks. However, engineering education research has not leveraged the potential of these tools to analyze social media data relevant to engineering students.

This paper contributes to the body of literature regarding engineering students' college

experiences by utilizing social media monitoring tool Radian6 to analyze public discourse on Twitter about engineering students' college experiences. The goal of this paper is to address the following research questions: (1) What can public discourse on Twitter imply about engineering students' college experiences? (2) What patterns, common themes or issues can be identified that could inform future efforts in students' retention and academic success? In the process of answering these questions, we investigate the utility of social media monitoring tools for conducting scholarly research. We provide an initial prototype of a method to collect and analyze vaguely defined and time dispersed social media data for educational purposes.

To provide necessary background, the next section reviews literatures on studies about college students' learning experiences, characteristics of public discourse online, and usage of social media analytics tools.

Literature Review

A Campus-Wide Focus on the Student Experience

The National Association of Student Personnel Administrators report "*Learning Reconsidered: A Campus-Wide Focus on the Student Experience*" [4] proposed a learning model considering the broad context of learning environment and the interconnectedness of many factors influencing students learning. In this report, learning is defined as "a comprehensive, holistic, transformative activity that integrates academic learning and student development, processes that have often been considered separate, and even independent of each other" (p. 2). The student learning contexts considered in this framework include social context, academic context, and institutional context. Social context mainly refers to personal relationships, group memberships, and inter-group connections. The academic context primarily involves opportunities for reflective judgment and critical thinking, constructivist classroom teaching methods, interdisciplinary courses, experiential learning, and integrative conversations with faculty in all domains. The institutional context refers to rewards structure

and campus culture. The framework suggests that these three contexts, students themselves, and learning outcomes are interconnected and integrated.

In this paper, we consider social, academic and institutional contexts as environmental factors influencing students' experiences and use them as guidelines to formulate the initial search terms for retrieving data relevant to engineering students' experiences. We refer to this framework as "the campus-wide student learning model" in the rest of this paper.

Theories and frameworks like "the campus-wide student learning model" provide the foundation for considering students' learning as a holistic experience. Most empirical studies focus on various aspects and address specific issues in students' experiences. We review empirical studies regarding engineering students' learning experiences in the next section.

Engineering Students' Learning Experiences

One of the major research projects regarding engineering students' experiences is the Academic Pathways Study (APS) conducted by the Center for the Advancement of Engineering Education (CAEE). APS consists a series of longitudinal and multi-institutional studies on undergraduate engineering students' learning experiences and their transition to work. It involved over 5,400 students across the country. The research team included over 130 faculty, research scientists, graduate and undergraduate research assistants, and staff representing 12 universities and six national organizations. They used various research methods including surveys, structured interviews, semi-structured interviews, ethnographic observation, engineering design task, and small focus groups [5,6]. The CAEE final report[7] provides a full list of over 150 journal and conference publications and dissertations produced from the APS study and three other smaller projects from CAEE. The CAEE website [8] provides 32 research briefs from the APS study including topics such as *developing identity as an engineer* [9], *conceptions of engineering* [10], *workload and life balance* [11], and

persistence in engineering as a college major and as a career [12].

Other smaller research projects usually focus on engineering students' experiences in particular classes. For example, Courter et al. interviewed freshman engineering students about their experiences in a freshman design class using open-ended questions and identified aspects of their experiences that could lead to improved student retention rate in engineering [13]. Using multiple survey instruments, Demetry and Groccia evaluated and compared mechanical engineering students' experiences in two introductory materials science classes with one implementing active learning and cooperative learning strategies [14]. Torres et al. presented students' experiences of learning robotics within a virtual environment and remote laboratory, where students knowledge was assessed via automatic correction tests and students opinions were collected using self-evaluation questionnaires [15]. Grimes et al. evaluated civil engineering students' experiences in a visual classroom named VisClass equipped with 2D digital whiteboards, 3D wall display and specialized software for engineering visualization combining both qualitative and quantitative data from observation, interview, survey, and examination of students' work [16].

The studies above collect data using traditional methods such as survey, interview, focus group, and observation. Traditional prompt methods such as survey and interview usually require the participants to remember and reflect on what they have experienced, which may not be accurate and fresh anymore. Even with observation, the participants might behave unnaturally knowing they are being observed. Moreover, data collected through these methods are usually confined to specific institutions or classes. This paper complements those studies by collecting data relevant to engineering students' college experiences from Twitter. Compared with data collected using traditional prompt methods, data collected from social media sites are generated spontaneously in informal and natural settings, and thus could bring new insights into students' experiences. Literature in the next section elaborates on public discourse on social media sites.

Public Discourse Online

As of March 2011, over 2 billion people (30% of the world's entire population) around the world were using the Internet [17]. In July 2009, it was estimated that 625 million of the Internet users were active users (users who use the Internet everyday or every other day) [18]. Users create a tremendous amount of content on the web, much of which is in the form of public conversations. Conversations online happen in relative informal and natural settings compared with formal classroom or interview settings.

Goffman's theory of social performance [19,20] is used to explain people's online behaviors related to their personal identity management. One of the most fundamental aspects of this theory is the *front-stage* and *back-stage* of the performances. "The front-stage is the observable space, the setting in which explicit performances are constructed and displayed where individuals 'play their parts'", while "back-stage is a more private area, where intimacy and familiarity see a relaxing of the strictures of performance" [20] (p.2). The question of whether the web is the front-stage or back-stage is a relative issue. For students, compared with formal classroom settings, social media sites are informal back-stages [21]. While there is a call to connect informal and formal settings to enhance students' classroom experiences [21], the focus of this paper is to explore data created at the informal back-stages, and is not constrained by the classroom usage of social media. When students post content online they usually post what they think and feel at that moment, while for some traditional data collection methods such as survey and interview, students have to reflect on what they thought and did before. In this sense, the data collected from online conversations is closer to what students were exactly experiencing at that moment. Conversations at the back-stage can shed light into what students really feel and think about their experiences.

Many web data mining and analytics tools have been developed to analyze online public discourse for purposes ranging from marketing to scholarly research.

Social Media Monitoring and Social Network Analysis

There are two primary categories of social analytics tools: monitoring tools (e.g., [Radian6](http://Radian6.com), radian6.com; Visible Intelligence, visibletechnologies.com) and network analysis tools (e.g. NodeXL, nodexl.codeplex.com). Social media monitoring tools mostly use histogram analysis, text-based analysis (e.g., tag-cloud based on word frequency counting), sentiment analysis and geographical distribution analysis (e.g., Visible Intelligence shows the geographical location of the data on a world map), while network analysis tools show how people or communities are connected. Since this current paper does not look at network relationships among people who talk about their college experiences, we only review relevant literature on social media monitoring below.

Monitoring business brands on social media sites has become a popular strategy on making marketing decisions. According to a report by University of Massachusetts Dartmouth Center for Marketing Research, 70% of the 2010 Inc. 500 companies (The Inc. 500 is a list of the fastest-growing private U.S. companies compiled annually by Inc. Magazine www.Inc.com) monitor their companies' names or brands on social media [22].

Social media monitoring tools are also used to analyze public events on social media. For example, Marcus et al. developed a web portal called TwitInfo to visualize and summarize events on Twitter that is more sophisticated and user-friendly than the trending topics function provided by the Twitter web UI [23]. Gaffney analyzed the Iran election event using data crawled from Twitter containing the hashtag #iranElection [24]. A group of researchers from Japan developed an earthquake detection system merely by monitoring tweets on Twitter and found that notifications delivered by this system is much faster than the announcements broadcast by the JMA (Japan Meteorological Agency) [25].

Monitoring business brands and public events is easier than the task we set to achieve in this paper in the sense that there is always a set of keywords

(the companies' names, the brands names or the names of the events) that can be searched, thus the data set is easier to locate. The concepts of the brands and events are well defined. Also, conversations about public events usually have peaks in certain time periods. However, in the case of this paper, vocabularies being used to describe engineering students' college experiences are very diverse and the time these conversations happen is dispersed. We went through the process of locating and refining relevant keywords. This paper pioneers the method of identifying and analyzing vaguely defined and time dispersed data on social media relevant to engineering students' experiences.

In sum, the informal public discourse on social media sites can provide valuable insights into engineering students' college experiences. These data remain unexplored by engineering education researchers. Analyzing these data is challenging because keywords associated with college experiences are very diverse and vaguely defined, and conversations associated with college experiences are time dispersed. We provide an initial strategy to tackle this challenge by mining data from Twitter with Radian6 using a set of keywords associated with engineering students' college experiences. During the process of retrieving data using this set of keywords, we identified a twitter hashtag #engineeringProblems that is specifically related to engineering students' academic problems, so we performed a further analysis of the tweets using this Twitter hashtag.

Methods

Data Collection

The data presented here were retrieved from Twitter using social media monitoring tool Radian6. We set the language to English and region to United States. The time period is from October 31st, 2011 to November 30th, 2011, containing 31 days of data. The educational Radian6 account we had access to could retrieve back 31 days of data since the day (in this case, November 30th, 2011) the data retrieving profile was created. This time period is close to the end of the semester at most schools in the United States, so students' conversation online might be different

from other time periods. Therefore, in order to achieve consistency at different stages of analysis, we did not use the data accumulated after November 30th.

Data source

Radian6 has the capability to mine data from many different social media types, including MicroMedia, Facebook, MySpace, blogs, blog comments, Image, Video, News, Buy/Sell forums, and forum replies. We initially included all media types, and quickly found that data from MySpace, Images, Videos, and Buy/Sell were irrelevant to the purpose of this study. Data from mainstream news were not useful here either, because they are usually official, formal and not from the students' perspectives. A large part of data posted on Facebook is private, so the data mined from the public part of Facebook were very limited. This limited data largely consist of quizzes, jokes or short stories relevant to engineering posted by students, but they do not reveal insights to students' personal experiences, so we did not include Facebook data in this analysis.

All data entries retrieved under the "MicroMedia" category are from Twitter in the case of this study; No other MicroMedia sites contain relevant data based on the data mining ability of Radian6 and our search criteria.

We found data from Twitter, blogs and forums can be useful to address our research questions. However, these three social media sites are designed for different purposes and formatted in different ways. The posts are of different lengths and the contents are of different depths. Therefore, they call for different methods of analysis. To choose a focus for this paper, we settled on analyzing Twitter data. Analyzing Twitter data serves as a pioneering example we use to investigate the utility of social media monitoring tools in educational research. We acknowledge that data from blogs and forums are also worthwhile to look at in order to gain insight into students' experiences, and should be included in the future research.

Searching keywords

The searching terms were developed based on the definitions and explanations of social context, academic context, and institutional context in the "the campus-wide student learning model" mentioned in the literature review section. Based on this framework, we developed the initial sets of keywords related to each context. Besides these three sets of keywords, we also developed an overall-scope keyword set in order to confine the boundary of retrieved data to engineering college students. From the initial keyword sets, we did many test runs to try out different combinations of keywords to see whether they can generate useful data sets about engineering students' experiences. In every test run, we read samples of the retrieved data, examined the wording differences in the relevant posts and irrelevant posts, and then refined the keywords accordingly. For example, we used "engineering" as one of the keywords in the initial test runs, and found that in people's casual conversations, "engineer" is used more frequently compared with "engineering". The keyword "engineering" mostly showed up in news or other official posts rather than students' personal experiences. So we decided to use "engineer" in keyword set 1 below, and put "engineering" in keyword set 2 (explained below after keyword set 1).

Overall scope keywords (Keyword Set 1):

In order to retrieve tweets related to engineering students, we used a set of keywords to retrieve a big pool of tweets first: *engineer, student, college, school, homework* and *class*. Considering plural forms, synonyms, and abbreviations, the logic is as follows:

1. Choose one from each of the following three parentheses, and use "AND" among them: (engineer, engineers), (student, students), (college, school), so there are eight combinations;
2. Choose one from each of the following two parentheses, and use "AND" between them: (engineer, engineers) (homework, hw), so there are four combinations;
3. Choose one from each of the two parentheses, and use "AND" between them: (engineer,

engineers) (class, classes), so there are four combinations;

4. There are all together 16 combinations, then use “OR” in between them.

According to the logic described above, the searching terms were:

(“engineer” AND “student” AND “college”)
OR (“engineer” AND “student” AND “school”)
OR (“engineer” AND “students” AND “college”)
OR (“engineer” AND “students” AND “school”)
OR (“engineers” AND “student” AND “college”)
OR (“engineers” AND “student” AND “school”)
OR (“engineers” AND “students” AND “college”)
OR (“engineers” AND “students” AND “school”)
OR (“engineer” AND “homework”)
OR (“engineer” AND “hw”)
OR (“engineers” AND “homework”)
OR (“engineers” AND “hw”)
OR (“engineer” AND “class”)
OR (“engineer” AND “classes”)
OR (“engineers” AND “class”)
OR (“engineers” AND “classes”)

To further reduce irrelevant tweets, and to locate content related to students’ social, academic and institutional contexts, we used the following four sets of keywords to refine the big pool of tweets retrieved based on keyword set 1. The logic is that from the pool retrieved above, we aimed to retrieve tweets containing any one keyword in the following four categories.

Keywords related to social context (Keyword Set 2-1):

Students’ experiences in the social context mainly related to personal relationships, group activities and interactions with groups outside of engineering and the larger society. Keywords in this category are: *team, team member, teammate, life, club, campus lounge, student lounge, diversity, service and society.*

Keywords related to academic context (Keyword Set 2-2):

This category of keywords are related to students’ academic life: *instructor, instructors, teacher, teachers, professor, professors,*

classroom, interdisciplinary, critical, analytical, design, senior design, seniordesign, math, hands-on, project, projects, lab, laboratory, library, libraries, tests, (We were to include “test” too, but Radian6 did not allow this keyword. It alerted: “test” will produce too many results, please be more specific), exam, and exams.

Keywords related to institutional context (Keyword Set 2-3):

Keywords developed under this category are: *tuition, computer, computers, campus, and equipment.*

Note that these keywords were developed based on the frameworks and our empirical test runs. They may not cover everything related to students’ experiences. It is very hard to define the boundary of vocabulary people use to talk about college experiences. In the process of eliminating irrelevant data, we may have also ruled out many relevant data sets.

Data were retrieved under the three categories above, but the tweets retrieved using a specific keyword did not necessarily talk about content related to the category the keyword belonged to. One tweet can contain several keywords belonging to different categories. So we retrieved all the tweets using these keywords and then manually examined the content of each tweet.

Including “engineering” (Keyword Set 2-4):

As mentioned earlier, “engineering” is not as commonly used as “engineer” in people’s casual conversation. It is still a relevant keyword to the topic, so we put “engineering” in keyword set 2.

Eliminating keywords (Keyword Set 3):

During the process of refining keywords, we found many tweets retrieved are recruitment advertisements from companies and organizations recruiting engineers. So we eliminated tweets containing the following keywords: *job, jobs, in search of, and looking.*

Our final search criterion is as follows:

(Keyword Set 1) AND (Keyword Set 2) NOT (Keyword Set 3)

Keyword set 2 is a combination of every keyword in keyword set 2-1 to 2-4. We used “OR” between each keyword in keyword set 2.

Data Analysis

Based on the filtering criteria and search keywords above, we retrieved 179 tweets from the 31 day period. Then we exported all the data into a .csv file and manually examined each tweet. We found 116 out of 179 of these tweets were actually about engineering students’ college experiences. We categorized them into social context, academic context, institutional context, and miscellaneous. We then presented the findings for each category. This is the first stage of our data analysis.

During the process of the first stage data analysis, a recurring Twitter hashtag #engineeringProblems caught our attention. A Twitter hashtag is a keyword starting with a # sign that users use to identify posts around certain topics. In this case, we found that #engineeringProblems is the hashtag that engineering students use to post problems about their learning experiences. We identified 667 tweets in the same 31-day period using this hashtag. The content of these tweets supports and details the main findings from the first stage analysis, and further reveals issues in engineering students’ learning experiences. So we did a further analysis of these tweets, which is the second stage of our data analysis.

Results of these two stages of data analysis are presented in the next section.

Results

First Stage Data Analysis

Conversation cloud

For the 179 tweets retrieved, we did a conversation cloud based on word frequency count (see Figure 1). Filler words such as “is, of, or, an, a, don’t, she, he ...” are automatically

eliminated in Radian6 when generating the conversation cloud. The conversation cloud reveals some information from the tweets. We can see that keywords such as *studying*, *homework*, *classes*, and *teachers* rank very high. However, many important keywords in the conversation cloud cannot be distinguished and compared clearly since their frequency is very close to each other compared with the high frequencies of “engineer” and “engineers”. So we performed a manual content analysis of these tweets.

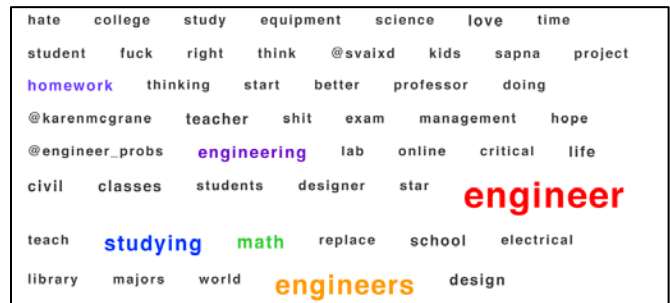


Figure 1. Conversation cloud from the first stage data.

Qualitative content analysis

We exported the 179 tweets into a .csv file, and manually filtered them. We found 116 out of 179 tweets were actually about engineering students’ college life. We did a content analysis [26-28] of the 116 tweets. We categorized them into four categories: social context, academic context, institutional context, and miscellaneous. There were 74 tweets under academic context; 27 tweets under social context; 3 about institutional context and 14 in miscellaneous. The sum is bigger than 116 here, because some tweets belong to more than one categories. We then identified sub-categories under each major category. The following are the main findings from each category.

Academic context

A large portion of tweets (74 out of 116) fell into this category. A large number of tweets (56 out of 74) in this category are complaining about heavy study load, classes and professors. They express a stressful experience and lack of social life. This echoes a previous study on engineering students’

Category	Number of tweets	Sub-category	Number of tweets	Examples
Academic Context	74	Complaints about heavy study load and exams	33	“still feel sick as life&eye still irritated...just woke up bouta go2the engineer lab do my hw” “if u wanna b an engineer, u must act like one...so goodbye social life and hello 2 countless hours of studying and hw”
		Complaints about classes	16 (5 among the 16 talk about useless classes: history, sociology, and chemistry for computer engineers)	“Man I hate this engineer design class” “All this studying and shit better pay off...the fuck I need chemistry n sociology for n I’m tryna b a computer engineer lol”
		Complaints about professors	7	“my professor who sits in a chair all class just tried to call all us engineers lazy...get the f outta here lady” “My teacher’s voice hits the resonant frequency of my eyelids.”
		Enjoy what they do	4	“math homework is so easy now that i actually take time to do it, becoming a civil engineer might not be so hard after all”
		Other	15	“Learning in my audio engineer class. my professor is #Mr.youtube”
Social Context	27	Non-engineer students’ opinions on engineering students	12 (negative: 5; positive:3; netural:4)	“I hate having class in the Engineering building. Engineers don’t shower” “i remember why i hate engineers, all they talk about is girls and booze, o and making fun of teachers while in class” “Engineers are so interesting”
		Lack of women and disrespectful to other under-represented groups	5	“one shot to all the black engineers out there...in class” “i get a B- on my public speech because i talked about The National Society Of Black Engineers” “eighty five kids leaving the classroom before mine.. of those 85 four are girls. engineers math class #Stereotypical”
		Help others do homework	3	“What am I going to tell our kids about mommy and daddy in school? Daddy did all of mommy’s math homework. Yay engineers” “just got some engineers to help me with my math homework”
		Other	7	“Studying during my manicure. Life of a blonde engineer” “for the first time we having a group effort” “the teacher is really hot”
Institutional Context	3	N/A	N/A	“Engineer student needs new school equipment and cant afford new stuf.”
Misc.	14	Self-identity	3	“I am fucking studying engineering but who am I really? A nerd!!! Dawg a nerd! No one looks at engineers as special alight! Read that!” “what kind of architect or engineer am I planning to be?”
		Other	11	“no one understands us” “I just want a better life...” “math class is like philosophy class for engineers”

Table 1. Main findings and example tweets from each category from the first stage data analysis.

life balance from the APS (Academic Pathways Study) project by Loshbaugh et al. [11], which indicates engineering students desire a more balanced life than their academic environment allows. Many engineering students also complain that classes such as sociology and history are useless to them. Only a very small number of tweets (4 out of 74) indicate that the students are enjoying what they do. Examples are provided in Table 1. We kept the typos and acronyms from the original tweets.

Social context

Interestingly, a fair amount of tweets (12 out of 27) in this category are from non-engineering students. These tweets reveal the stereotypical image of engineers. For example, non-engineering students tend to think that engineering students are geeky, good at math (so non-engineering students tend to ask engineering students to help with math homework), and they do not shower that often, which makes the engineering building smelly. There are also very positive tweets saying engineers are very interesting, or they intend to date engineering guys. Another sub-category indicates the lack of representation and respect of women and other minorities.

Institutional context

Very few tweets fell into this category. This may be due to the search criteria we used. We searched for general experiences related to all engineering students thus we did not use any university names in the search, and institutional contexts vary from institution to institution. It may be also because students do not like to talk about institutional affairs on Twitter.

Miscellaneous

We mainly focused on students' experiences directly influenced by the environment, and did not intend to dig deep into student self-related issues. However, we found a recurring theme about self-identity. Several tweets in this category express students' confusion about formulating the identity of being an engineer. This confirms previous studies regarding identity formulation of

being an engineer [9,10], which indicate engineering students usually have difficulty formulating the identity of being an engineer and what it means to be an engineer even at the later stage of their engineering degrees.

Second Stage Data Analysis

From the first stage of data analysis, we found that the major theme is students' complaining about classes, studying, professors and exams. Students posted about how difficult it is for engineering majors to achieve balance between studying and life. The second stage analysis reconfirms and provides more detailed insights into this theme. At the second stage of data analysis, we retrieved 667 tweets using the Twitter hashtag #engineeringProblems. Before diving into manual content analysis, we created a conversation cloud and ranked keywords in the conversation cloud.

Conversation cloud and keyword rankings

The conversation cloud based on tweets using the #engineeringProblems hashtag is shown in Figure 2. It is obvious that the frequency of the keyword #engineeringProblems is too high, which makes all other keywords undistinguishable among each other in the Radian6 conversation cloud. So we ranked the keywords according to the number of tweets they appeared in (Table 2). We performed manual content analysis of the tweets using the most frequently occurring keywords as counted by radian6 (keywords highlighted in Table 2). Keywords @engineering Problem and @engineer_probs are ranked very high, but @ is a sign to indicate a user account on Twitter. When people tweet using @, this means they are addressing a specific user account, rather than posting about a certain topic. So we did not analyze tweets using these two keywords. The keyword *engineering* also ranks very high, but this word conveys very broad meaning, while keywords such as *test*, *studying*, and *homework* represent specific contexts. Therefore, we present the analysis results of tweets using keywords *test*, *homework*, *work*, *lab*, *study*, and *studying* in the next section.

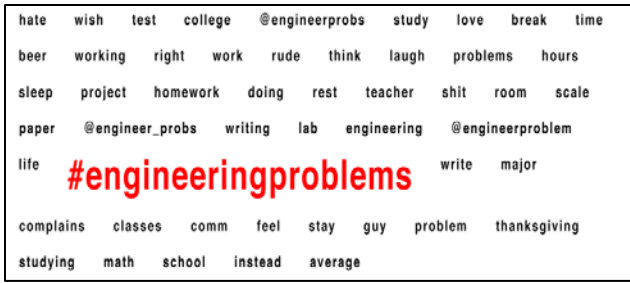


Figure 2. Conversation cloud from the second stage data.

Keyword	Number of tweets containing the keyword	Ranking
test	52	1
homework	51	2
@engineeringproblem	45	3
@engineer_probs	39	4
work	36	5
lab	32	6
studying	31	7
engineering	30	8
study	28	9

Table 2. Keywords ranking (only keywords appearing in at least 25 tweets are presented; tweets using highlighted keywords were analyzed).

Qualitative content analysis of tweets using top keywords

The sum of the number of tweets using the keywords *test*, *homework*, *work*, *lab*, *study*, and *studying* was 230. There were some tweets that were counted under multiple keywords, so the net number of total tweets was 213. They took about a third of the total 667 tweets. Because *studying* and *study* are two different forms of one original word, we combined tweets using these keywords when doing the analysis. There were 58 tweets under *study* & *studying*. The main themes under each keyword are presented in Table 3. Although we analyzed tweets under each keyword separately, we found the same repeating themes under each keyword.

Heavy workload

A predominant number of tweets under every keyword expresses that students have heavy workload being engineering majors—having a lot of studying, a lot of tests and a lot of homework.

Sacrifice

As a result of the heavy workload, students need to sacrifice a lot of things in order to study, to do homework, and to prepare for exams. Many tweets mention that they have to sacrifice enjoying Thanksgiving time with family, Halloween, birthdays, and basketball games. They also often have to sacrifice on sleep. Another practice expressed in the tweets is that students tend to skip one class or do frivolously in one class in order to do homework for another class. This theme reconfirms the study mentioned previously which indicates engineering students wish to achieve a more balanced life than their academic environment allows [11].

Negative feelings

Because of all the sacrifice they have to make, students tend to have negative feelings towards studying and exams. For example, one tweet mentions that he/she wants to step in front of a moving car before going to study.

Difficulties

In general, there are more complaints about the quantity of the work than the difficulty of the work. This may be because students tend to complain about quantity rather than difficulty on Twitter, but it may also be because professors tend to give large amounts of homework that students do not perceive as meaningful, which can be an important issue in education that needs to be addressed.

Many tweets under the keyword *test* mention that the average test scores were very low, so although students earned very low scores they still got good grades. While students may sound happy when they get a low score but still a good grade, this reveals an issue of improper assessment. The

Keyword/Category	Number of Tweets	Theme/Sub-category	Number of tweets	Examples
Study & studying	58	Heavy workload (lots of studying)	31	“back hurts from sitting at the desk studying for so long”; “Study over 30 hours for a test” “up late studying again....it’s no longer unusual, but more of a routine” “going to bed at 3 A.M. Still have about 8 hrs of homework and studying to do....”
		Sacrifice life (also sacrifice the time of other classes)	9	“I don’t have enough photos from college because all I do is study” “in fact, copying solutions for one class so you can study for a different one...”
		Negative feelings	9	“studying is useless” “Study = Fail” “is it bad that before i started studying for my tests today that i considered throwing myself in front of a moving car??” “I don’t want to study anymore!!”
		Difficulties	1	“Studying and still not understanding a damn thing”
		Other	8	“Teacher told us we should embrace our image as nerds”
test	52	Heavy workload (having tests, and having lots of tests)	29	“My head hurts... Raped by test 1, let's see how the other 3 go...”
		Difficulties	20	“i was 15 above the average on one test.... i had 50” “Average emags test grade = 49 My grade = 47. I think that's passing...” “Got a 44 on my engineering test, sounds bad but the average was a 38” “Would it kill him to just put one problem on the test that we've actually seen before?”
		Negative feelings	3	“About to crash. I still have to use all my remaining energy on a test. Wanna just lay down” “that test ruined my week/life”
homework	51	Heavy workload (lots of homework)	14	“so much homework, so little time” “C++ CAE project due Tuesday, Mfg project Wednesday, 25 Page Tech Repot Wednesday + heavy homework load. Huzzah.”
		Sacrifice life	12	“I havent been to a basketball game yet this year. What is wrong with me? Oh yea it's because I always have homework” “Doing homework on my birthday”
		Difficulties	11	“I wish my professors wouldn’t assign homework that they didn’t teach me how to do” “not being able to do your calculus homework because you forgot how to do the algebra” “It took me 4 hours, but i finally got the first homework problem done.”

		Other	14	“Pondering on hooking up with my TA to get my homework average up”
work	36	Heavy workload	24	“Is it rude to laugh when a comm major complains about how much work they have?” “It’s monday! That’s means 8 hours of work followed by 3 hours of Organic Lab!”
		Sacrifice life	5	“Let the shit show begin. And I mean a shit ton of school work. Is it thanksgiving yet?” “Hello Thanksgiving Break! Also, hello excessive amount of work I want to accomplish in a week!”
		Other	7	“Ever feel like all your hard work has gone to waste?”
lab	32	Workload	16	“Literally spending my post-dinner night in the freshman lab. Solid Edge, you kill me.”
		Sacrifice life (and sleep)	6	“Falling asleep in class because you stayed up all night working on a lab report”
		Social observation	4	“Kid next to me in the computer lab passed out on the keyboard.”
		Other	6	“I pity the 1 girl in my lab with 25 guys. It smells like man in here.... And that's not in a good way”

Table 3. Main themes and example tweets under each keyword from the second stage data analysis.

type of assessment methods may not be designed appropriately to measure students’ abilities.

Another issue revealed under this theme is students’ lack of the ability of transferring knowledge to problems they have not seen before. For example, one tweet says “Would it kill him to just put one problem on the test that we've actually seen before?” This tweet expresses anger towards the professor who gave the test, but also reveals that the student was unable to deal with problems he/she had not seen before.

Discussion

Through analyzing data related to engineering students’ college experiences on Twitter, we found a large number of tweets complaining about homework, classes, professors, exams, and studying. Negative tweets are overwhelmingly more numerous than positive tweets. This indicates an imbalance between academic life and social life among engineering students. Poor curriculum design is also revealed. For example, many students complain that sociology and history classes are useless to engineering majors, and chemistry class is useless to computer engineering

majors. These classes need to be better designed and tuned to the needs of engineering students.

As stated in a previous APS study [11], if engineering education wants to attract and retain students, it must find ways to provide the environment and opportunities for students to maintain a more balanced life. We acknowledge that this cannot be solved simply by reducing students’ workload. It is necessary for students to work hard and master a lot of knowledge and skills in order to become successful engineers. However, we cannot ignore these complaints either. It is wise to help students pick the important content to work on, and manage stress and negative feelings. Additionally, the fact that students tend to complain about their imbalanced life on social media sites may indicate that social media is an important venue for releasing stress and seeking emotional support. Future efforts can look into using social media to help students manage stress and build supporting communities.

Other issues such as lack of representation of women and other minorities, stereotypical public image of engineers, students’ difficulties in formulating an engineer identity, poorly designed assessment methods, and students’ difficulties in

transforming knowledge to problems they have not seen before were also identified.

Many of the issues discussed above have been identified by previous studies. Because the time cycle of retrieving data from the Internet is much shorter than using other traditional methods, if researchers and policy makers utilize these data, they can become more aware of the most up-to-date issues.

Conclusion

This study is aimed at investigating engineering students' college experiences by analyzing data retrieved from social media site Twitter. We retrieved 179 tweets using social media monitoring tool Radian6 with a set of keywords developed around students' social, academic and institutional contexts. During the process of analyzing these data, we identified a Twitter hashtag #engineeringProblems, which engineering students use to address problems in their learning experiences. We did a further analysis of 213 tweets using this hashtag. The findings confirm many issues identified by previous studies, such as imbalanced life, lack of representation of women and other minority groups, self-identity formulation issues, and assessment issues. Efforts should be made to help students release stress and build supporting communities, both emotionally and intellectually. If researchers and policy makers use social media to identify these issues, they may be able to be aware of the most up-to-date issues and formulate actionable plans faster.

One limitation of this study is that in order to retrieve tweets about engineering students' college experiences and eliminate spams and irrelevant tweets as much as possible, we used search terms specifically pointing to "engineer", "student" and "college". Many tweets relevant to engineering students' experiences may not specifically use these terms, and thus get filtered out of the dataset. It is difficult to define the boundary of the words people use to talk about engineering experiences. This is a limitation of this study and also an on-going research topic in computer and information science on how to effectively index and retrieve information on certain topics.

Another limitation of this study is that the time period we chose was close to the end of the semester and the holiday season. It may be because of the stressful time period, there tended to be more negative than positive posts. However, negative posts are valuable because they help us identify problems and issues better than positive posts.

This study only analyzed data from Twitter. Future work should analyze data from blogs, forums and other types of social media sites. Analysis can also be done for students in specific institutions, but this brings up the challenge of identifying relevant data to specific institutions.

During the data collection process, we tried several social media monitoring tools and chose Radian6 as the relatively more effective one in terms of retrieving data for the purpose of this study. However, we found that because many of these tools, including Radian6, are designed for commercial reasons, the usage of these tools for educational research purposes is very limited, manual analysis is still largely needed, which limits the amount of data can be analyzed. Future opportunities exist to develop social media analytics tools tuned to the needs of educational researchers and policy makers. Ethical issues regarding whether universities should monitor online student discourse and how to do it properly also need to be addressed.

References

1. Committee on Prospering in the Global Economy of the 21st Century *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. (The National Academies Press: 2005).
2. Rising Above the Gathering Storm Committee *Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5*. (The National Academies Press: 2010).
3. Lenhart, A., Purcell, K., Smith, A. & Zickuhr, K. *Social media & mobile internet use among teens and young*

adults. (Pew Internet & American Life Project: 2010).

4. Keeling, R. P. *Learning Reconsidered: A Campus-Wide Focus on the Student Experience*. (National Association of Student Personnel Administrators, American College Personnel Association: 2004).
5. Clark, M. *et al.* Academic pathways study: Processes and realities. *Proceedings of the American Society for Engineering Education Annual Conference and Exposition* (2008).
6. CAEE Academic Pathways Study. at <www.engr.washington.edu/caee/about_APS.html>
7. Atman, C. *et al.* *Enabling engineering student success: The final report for the Center for the Advancement of Engineering Education*. (2010).
8. CAEE Academic Pathways Study Research Brief. at <www.engr.washington.edu/caee/research_brief_student_exp.html>
9. Chachra, D., Kilgore, D. & Loshbaugh, H. G. Being and becoming: gender and identity formation of engineering students. *American Society for Engineering Education Annual Conference* (2008).
10. Matusovich, H. M., Streveler, R., Miller, R. & Olds, B. I'm Graduating This Year! So What is an Engineer Anyway? *American Society for Engineering Education Annual Conference* (2009).
11. Loshbaugh, H. G., Hoeglund, T., Streveler, R. & Breaux, K. Engineering School, Life Balance, and the Student Experience. *American Society for Engineering Education Annual Conference* (2006).
12. Fleming, L., Engerman, K. & Griffin, A. Persistence in engineering education: Experiences of first year students at a historically black university. *American Society for Engineering Education Annual Conference* (2005).
13. Courter, S., Millar, S. & Lyons, L. From the Student's Point of View: Experiences in a Freshman Engineering Design Course. *Journal of Engineering Education* **87**, 283-288 (1998).
14. Demetry, C. & Groccia, J. E. A comparative assessment of students' experiences in two instructional formats of an introductory materials science course. *Journal of Engineering Education* **86**, 203-210
15. Torres, F. *et al.* Experiences with Virtual Environment and Remote Laboratory for Teaching and Learning Robotics at the University of Alicante. *International Journal of Engineering Education* **22**, 766-776 (2006).
16. Grimes, D., Warschauer, M., Hutchinson, T. & Kuester, F. Civil engineering education in a visualization environment: Experiences with VizClass. *Journal of Engineering Education* **95**, 249-254 (2006).
17. World Internet Usage Statistics News and World Population Stats. at <<http://www.internetworldstats.com/stats.htm>>
18. *Power to the People-Social Media Tracker Wave 4*. (Universal McCann: 2009).
19. Goffman, E. *The Presentation of Self in Everyday Life*. (Anchor: 1959).
20. Pearson, E. All the World Wide Web's a Stage: The performance of identity in online social networks. *First Monday* **14**, 1-7 (2009).
21. Bull, G. *et al.* Connecting informal and formal learning: Experiences in the age of participatory media. *Contemporary Issues*

in *Technology and Teacher Education* **8**, (2008).

22. Barnes, N. *The 2010 Inc. 500 Update: Most Blog, Friend And Tweet But Some Industries Still Shun Social Media*. (UMass Dartmouth: 2011).at <<http://www.umassd.edu/cmr/studiesandresearch/industriestillshunsocialmedia/>>
23. Marcus, A. *et al.* Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration. 227 (2011).doi:10.1145/1978942.1978975
24. Gaffney, D. #IranElection: Quantifying Online Activism. *WebSci10: Extending the Frontier of Society On-Line* (2010).
25. Sakaki, T., Okazaki, M. & Matsuo, Y. Earthquake shakes Twitter users: real-time event detection by social sensors. *Proceedings of the 19th international conference on World wide web* 851–860 (2010).doi:10.1145/1772690.1772777
26. Weber, R. P. *Basic Content Analysis*. (Sage Publications, Inc: 1990).
27. Berg, B. L. & Lune, H. *Qualitative Research Methods for the Social Sciences*. (Prentice Hall PTR: 2011).
28. Neuendorf, K. A. *The content analysis guidebook*. (SAGE: 2002).

Biographical Information

Xin Chen is a Ph.D. student in the School of Engineering Education at Purdue University. She received her B.S. in Electrical Engineering from East China Normal University in 2010. Her research focuses on social media data analytics in the context of engineering education and research, and web interface development and personalization.

Dr. Mihaela Vorvoreanu is an assistant professor in Computer Graphics Technology and Technology Leadership & Innovation at Purdue University. She studies the socio-cultural impact

of new communication technologies. Before joining Purdue, she was an assistant professor in the Department of Communication Studies at Clemson University, SC, and the Department of Communication at the University of Dayton, Ohio. While at Clemson and UD respectively, Dr. Vorvoreanu taught various public relations and communication courses, and did academic research in the area of public relations and new Web technologies. . She has published research articles in the *Journal of New Communications Research*, *Public Relations Review* and the *Journal of Website Promotion* and a book about online public relations: *Web Site Public Relations: How Corporations Build and Maintain Relationships Online*. Dr. Vorvoreanu holds a Ph.D. in Communication from Purdue University.

Dr. Krishna P. C. Madhavan is an Assistant Professor in the School of Engineering Education at Purdue University. Prior to his arrival at Purdue, he was an Assistant Professor with a joint appointment in the School of Computing and the Department of Engineering and Science Education at Clemson University. Dr. Madhavan also served as a Research Scientist at the Rosen Center for Advanced Computing, Information Technology at Purdue University where he led the education and the educational technology effort for the NSF-funded Network for Computational Nanotechnology (NCN). Dr. Madhavan continues to work actively on the NCN Education Team as the Education Director. He specializes in the development, deployment, and assessment of advanced cyber-infrastructure tools in day-to-day engineering and science curricula. One of the key aspects of his work focuses on how semantic grid-based technologies and tools can co-exist with students' lifestyles, learning patterns, and technology choices. Dr. Madhavan was the Chair of the IEEE/ACM Supercomputing Education Program 2006 and was the curriculum director for the Supercomputing Education Program 2005. In January 2008, he was awarded the NSF CAREER award for work on transforming engineering education through learner-centric, adaptive cyber-tools and cyber-environments.