What do Transfer Students Have to Say: An Analysis of the Experience of Transfer Students through Topic Modeling

Ms. Claire MacDonald, The University of Texas at El Paso

Claire MacDonald is a research assistant at the University of Texas at El Paso and she is currently conducting data analysis using Natural Language Processing on online surveys. She likes to visit and explore the National Parks nearby her hometown of El Paso, Texas.

Palvi Aggarwal, The University of Texas at El Paso

Dr. Aggarwal is an Assistant Professor in the Department of Computer Science at the University of Texas at El Paso (UTEP). Dr. Aggarwal has focused on socio-technical aspects of cybersecurity using human experiments, machine learning, and cognitive modeling. She is currently leading an interdisciplinary research lab, i.e., Psyber Security Lab at UTEP, that focuses on improving cyber defense by understanding human decision-making processes. At UTEP, Dr. Aggarwal teaches courses on Computer Security, Behavioral Cybersecurity, and Applied Computational Cognitive Modeling to undergraduate and graduate students. Dr. Aggarwal has strong interdisciplinary collaborations with various universities and such collaboration will be beneficial for this project. Dr. Aggarwal published her research work in various conferences including HFES, HICSS, ICCM, GameSec, and journals including Human Factors, Topics in Cognitive Science, and Computers & Security. Her papers in HICSS-2020 and GameSec-2020 received the best paper awards. Her professional activities include journal reviews for Computers & Security, Cybersecurity, Frontiers in Psychology, and conference reviews for HFES, AHFE, HICSS, Euro S&P, and CyberSA. She is also an advocate for the Cybersecurity Community of Practice at UTEP and a member of the Special Cyber Operations Research and Engineering (SCORE) Interagency Working Group.

Xiwei Wang, Northeastern Illinois University

Xiwei Wang is the Department Chair and an Associate Professor of Computer Science at Northeastern Illinois University. He earned his Ph.D. in Computer Science from the University of Kentucky and. His primary research interests include recommender systems, data privacy, data mining, and machine learning. He has served as an associate editor, editorial board member and reviewer of international journals. He also served as a technical program committee member, session chair, and reviewer for many international conferences.

Yun Wan,

Yun Wan is a Professor of Computer Information Systems in the University of Houston- Victoria. His current research includes electronic commerce and information systems in STEM education. His other research includes text analytics, decision support systems, and enterprise systems development. His research is funded by the National Science Foundation (NSF). He serves as senior editor for Electronic Markets and an editorial board member for several journals, such as the Journal of Electronic Commerce in Organizations. He received his Ph.D. in Management Information Systems from the University of Illinois at Chicago.

Dr. Shebuti Rayana, The State University of New York at Old Westbury

Shebuti Rayana is an Assistant Professor of Computer and Information Sciences at the State University of New York at Old Westbury (SUNY OW). She received her PhD from the Department of Computer Science at Stony Brook University. Before moving to the United States for higher studies, she completed BSc from Computer Science and Engineering at Bangladesh University of Engineering and Technology (BUET). Shebuti Rayana's research is to build a safe and secure digital world with the help of cutting-edge Data Mining techniques. During her PhD, she was involved in several projects funded by National Science Foundation (NSF), Defense Advanced Research Projects Agency (DARPA), and R&D grant from Northrop Grumman to develop Anomaly Mining algorithms and apply them to solve real-world problems. She also worked as a Research Intern in the Information Security team at IBM Thomas J. Watson Research

Center. She has been awarded two NSF: Computer and Information Science and Engineering - Minority Serving Institution (CISE-MSI) grants as a Co-PI, (1) to increase the research capacity at SUNY OW by creating the infrastructure for big data research, incorporating course embedded undergraduate research experience, and training undergraduate students in big data research through seminars, workshops, and summer bridge programs, (2) to design an AI-driven counseling system for underrepresented transfer students in collaboration with UTEP, NEIU, UHV, and Cal Poly Humboldt. Moreover, she is working on several projects on misinformation, stigma, hate speech, and cyberbullying detection and sentiment towards chatGPT from social media platforms.

Rudy Caraballo
Dr. Sherrene Bogle, Cal Poly Humboldt

Dr. Sherrene Bogle is a Fulbright Scholar and alumna of the University of Georgia, USA, where she earned her PhD in Computer Science. She is currently an Associate Professor of Computer Science and Program Lead for the BS Software Engineering at Cal Poly Humboldt. Dr. Bogle has a passion for sharing and helping students to improve the quality of their lives through education, motivation and technology. She has published two book chapters, two journal articles and several peer reviewed conference papers in the areas of Machine Learning, Time Series Predictions, Predictive Analytics, Multimedia in Education and E-Learning Technologies.

What do transfer students have to say?
An Analysis of the Experience of Transfer Students through Topic Modeling

Abstract

In recent years, there has been a notable rise in an alternative route to achieving higher education: a growing number of students are transitioning from 2-year colleges to 4-year institutions to complete their undergraduate degrees. Transfer students are a minority among the 4-year institution student population, many being first-generation, low-income, and racial minorities. To understand how to assist these underrepresented students, the question is: what are the most significant factors influencing the decision to attend a 2-year institution and transfer instead of immediately attending a 4-year institution? An online survey, which was anonymous and confidential, of 161 students in computing majors provides invaluable information about the transfer process for underrepresented students. This paper analyzed the demographic information along with the five open-ended questions asked to the participants of the survey. Participants' responses reveal the influence of their families, social media, and advisors and how aspects of their identity have affected their decisions. To gain a deeper understanding of this data, NLTK and Pandas libraries are used to clean the data, WordCloud library is used to generate word clouds and three topic modeling algorithms including unsupervised (i.e., Latent Dirichlet Allocation (LDA)), semi-supervised (i.e., Correlation Explanation (CorEx)), and pre-trained (i.e., Bidirectional Encoder Representations from Transformers (BERTopic)) models are used to identify critical issues regarding students' transfer decision. Responses are first cleaned, aggregated, and visualized into word clouds; separate word clouds are generated for each question to reveal critical factors. With the aggregated analysis of word clouds and topic-modeling results, it becomes evident that cost, career opportunities, financial aid, distance from home, and guidance from family are the key factors influencing the decision between 2-year and 4-year institutions. The biggest challenges in the participants' transition were transferring credits, difficult classes, working while attending school and overall adjusting to a 4-year institution. These findings can be used to help transfer students succeed in their 4-year institution and beyond in their careers.

1 Introduction

In recent years, an alternative pathway to higher education attainment has gained prominence: students are increasingly transferring from 2-year institutions to 4-year institutions to complete their undergraduate degrees [1]. Transfer students are a minority among the 4-year institution student population, many being first-generation, low-income, and racial minorities [2]. Community colleges and universities face enormous challenges co-advising transfer students due to the uncertainty surrounding the path from their decision to transfer to post-transfer graduation [3], particularly those who are from disadvantaged minority or underrepresented groups in STEM majors [4, 5]. Understanding the factors influencing their decision-making, the challenges they encounter,

their experiences, and expectations is crucial for providing effective advising and ensuring their academic success.

Student retention is one of the major challenges faced by American Universities and the risk of dropping out is even higher for transfer students [6]. A key aspect of retaining students and ensuring their success is to gain insights into their struggles and offer timely guidance. Much of the prior research has employed traditional statistical methods to understand factors contributing to student success, substantial focus has been on structured data such as GPA, transfer credits, SAT scores, etc. to predict student success. Researchers identified various factors that can negatively impact the persistence, retention, and completion rates of transfer students such as a lack of financial support causing students to attend a cheaper institution (e.g., a community college) before they could transfer to a 4-year institution [7, 8, 9]. While deciding to attend a 4-year university, the distance between old and new institutions impacts students' decision [10, 11]. The physical proximity of institutions not only reduces the hassle and cost of relocation but also makes course credit transfer easier for students. Social isolation and lack of belonging in the new institution have been identified as important factors that hinder the development of a social support system that helps students with their academic journey [8, 12, 9, 13, 14, 15]. Additionally, academic preparedness [12, 13] plays a key role in student success.

To address some of the above-mentioned issues, student advising is crucial. Advising plays a pivotal role in shaping a student's academic journey, from navigating the transfer process at community college to choosing a major at a four-year university [16, 17]. Students expect clear, accurate, and timely information while advisors work tremendously hard to meet student's expectations. However, with the traditional advising methods, it is exceedingly challenging for advisors to be well-versed in the details of each student and address their individual concerns [16, 17]. A deeper understanding of transfer students' concerns and experiences can empower advisors to effectively communicate on specific topics.

Only evaluating structured data may not help in uncovering all the factors that influenced students' decision to transfer, their experience and story in this journey, and success factors after the transfer are also important. Learning directly from transfer students and their experiences is vital to obtaining the holistic perspective of why transfer students either succeed or struggle. In this paper, a survey is conducted to identify the key factors using open-ended questions that influence the transfer decision, particularly for students from traditionally disadvantaged groups. We also perform an exploratory analysis of these factors by inviting students from both community colleges and 4-year universities to a survey that includes a wide range of questions, from demographics, pre-transfer decisions, post-transfer performance, etc.

The major contributions of this paper are, (i) analyzing the survey data through wordcloud to identify the most frequent factors in making the transfer decision, (ii) further analyzing this data with three topic modeling algorithms (i.e., LDA, CorEx, and BERTopic) to identify both broad and subtle topics, (iii) provide semi-supervision to the CorEx topic model through anchor words identified from the wordclouds which are less frequent words to identify more specific subtle topics which do not solely depend of word frequency regarding general information, important information, challenges, expectations and experience of transfer students in the transfer process.

The remainder of this paper is organized as follows. Section 2 gives the summary of the related

work, and Section 3 outlines the methodology of the study. Section 4 presents the data analysis on the datasets collected via surveys and interviews. Some concluding remarks and future research directions are given in 5.

2 Literature review

As the number of transfer students increases in 4-year institutions, understanding the factors that either hinder or propel the success of transfer students in their academic and professional careers has become increasingly critical [1]. In previous literature on the subject of higher education transfer students, two categories of transfer students were identified regarding their subsequent success in their careers [10, 11]. The critical difference is that students who plan the transfer before attending community college were more likely to succeed than those who did not plan early[18]. However, even with planning, there are still factors that are imperative to transfer student success post-transfer. Previous literature considers personal [11, 8, 9, 19] and academic factors about these differences [11, 8, 9, 19]. Major factors identified in the past research include social isolation, finance, advising, GPA, number of credits transferred etc. However, an in-depth understanding of how these factors impact the students is still to be explored. For example, there was no mention of issues such as how commuting to the university impacts students' decision of course selection, and cope with limited resources at institution campuses. Some studies show that the greater the distance between the two institutions, the greater the struggle for the student, however, this literature focuses on the transfer of credits and not on the actual commuting to the class itself [10]. Additionally, factors such as on-campus events like hackathons, research or teaching assistance positions, and internships, allow students to not only learn but also connect with other students on-campus and gain technical experience on campus are not well explored.

Past research has mostly conducted surveys to allow understanding of the experience of transfer students. Recent research has started exploring other ways of analyzing factors such as analyzing social media data, one-on-one interviews, meeting scripts from advisors, etc. Deng et al. [20] focused on the influence of social media on the decision to transfer through Revealed Casual Mapping (RCM) to study the relationships between students in social media spaces [20]. Harper et al. [17] conducted interviews to better understand the concerns with the advising. Without a deeper understanding of students' concerns, advising or overall decision support systems are difficult to develop. Along with traditional approaches of data collection and analysis, employing innovative methods such as using word clouds to visualize the most frequent words, and relying on topic modeling methods to reveal key issues or factors on students' success would help benefit this process.

This paper's purpose is to explore the underlying reasons for transfer student success and struggle through five questions on an online survey. The usage of Natural Language Processing will allow for Topic Modeling methods to find key issues that are not frequently talked about to learn how to let transfer students grow and to understand other pressing issues for them.

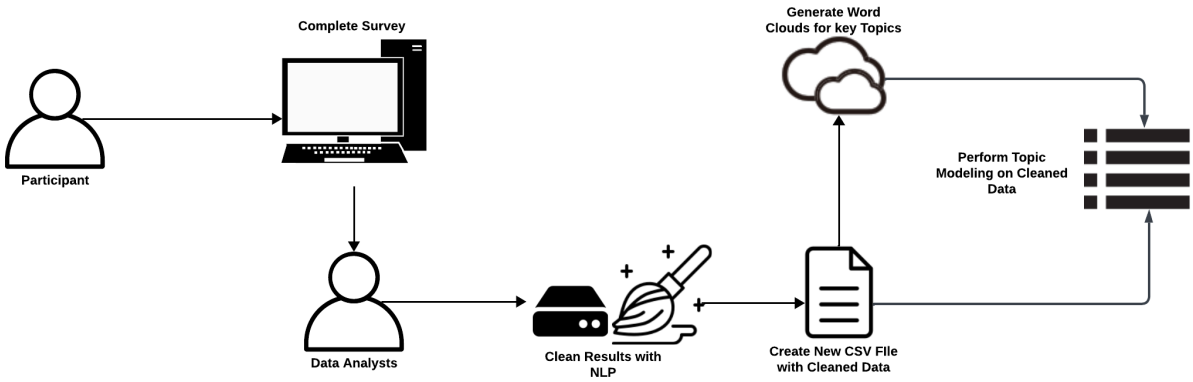

Figure 1: Process of data collection, cleaning, visualizing, and identifying topics in participant's responses

.

3 Methodology

The methodology used in this paper consists of several phases, (i) data collection from an online survey, (ii) data cleaning, (iii) uncovering the important factors through high frequency words using word clouds, and (iv) identifying important topics using topic modeling. Figure 3 shows the overall workflow of the methods used. Within several phases, there was the usage of multiple kinds of qualitative, frequency, and manual analysis. The employment of these three kinds of analysis was critical for uncovering subtle issues that would not be revealed through traditional analysis. The usage of Natural Language Processing enabled the ability to analyzed the 805 responses given by the 161 participants and the five questions given in the survey. Natural Language Processing was used within most of the phases of data analysis. Natural Language Processing has also been used in previous literature pertaining to students [21].

3.1 Data collection

For this project, a Qualtrics survey is designed to understand the decision-making processes of transfer students in their journey from community college to a 4-year institution. There were a variety of types of questions, such as multiple-choice and open-ended questions. For this paper, the responses to the open-ended questions and demographic information were analyzed. The participants were recruited from 5 Minority Serving Institutions (MSIs) across the United States. With the field of computer science growing exponentially, computing majors are becoming rapidly popular. In this paper, we focus on the students enrolled in computing majors. There are challenges that are specific to computing majors that need to be explored. The success rate of STEM students in particular are lower than their transfer rate, meaning that while many may transfer, their success rate is low [22].

3.1.1 Open-ended Survey Questions

Response data is collected using the following five open-ended questions.

- Q1: *General Information:* Please list the information that you used when deciding between a community college and a 4-year university. Please also mention the information that you feel was needed but not available. (For example 4-year university nearby, better job opportunities at 4-year university, ease of transfer, cost, guidance from family, friends, etc.)
- Q2: *Important Information:* Please indicate the most important information from the above list that would affect the decision of transfer.
- Q3: *Challenge:* What challenges did you encounter as you were transferring from your previous college or university
- Q4: *Expectation:* How did your expectations or career plan change after transferring to your current institution?
- Q5: *Experience:* Please share any particular experience you feel is important as a transfer computing major student.

These questions were written in consideration of the journey of a transfer student and how they were successful throughout the process. The success of the transfer student is considered in their positive experience transferring and after transfer to 4-year university.

3.1.2 Demographics

In this study, one hundred and sixty-one students from various institutions were surveyed. All respondents were computer science or computing-related majors. The majority of these transfer students were males with a smaller percentage of females. The rest of the participants identified as non-binary or preferred not to say. Overall, this group of participants was incredibly diverse in terms of race and ethnicity. The majority of these participants are Hispanic and Latino although by a small margin. The next biggest group is Asian, followed by White, Mixed, Black, and participants who responded with others for race and ethnicity.

Category	Sub-category	Percentage
Gender	Male	68.09%
	Female	27.60%
	Others	4.29%
Race & Ethnicity	White	19.8%
	Hispanic, Latina/o	28.22%
	Asian	22.36%
	Black, Afr. Am.	18.61%
	Mixed/Other	11.18%

Table 1: Demographics

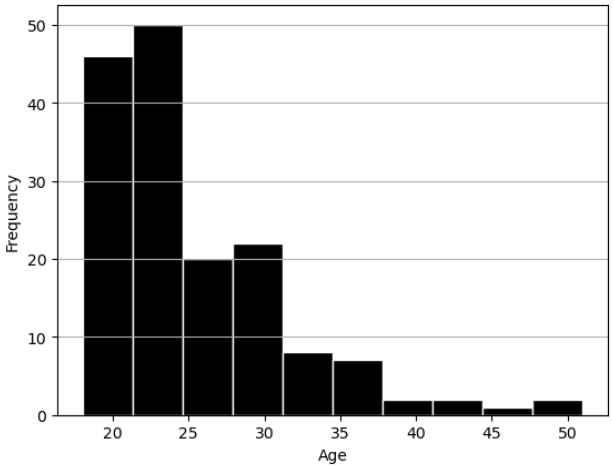

Figure 2: Histogram distribution of student age

The average age of the students was 25, however, there was a high level of standard deviation ($\sim$6.22) suggesting that there was significant variation in this category. Figure 2 shows the histogram distribution of their age, which indicates that the majority of participants are between

the ages of 20 and 30. The majority of participants were in-state residents with 1.84% being out-of-state, 7.36% being international students, and the 1.24% of students presenting as undocumented.

3.2 Data Cleaning

The goal of this work is to reveal the issues that are most important for taking transfer decision for a college student. Hence the text response data from the survey was cleaned and pre-processed. NLTK and Pandas libraries were utilized for data cleaning through Python. The Pandas library was used to process the CSV files by organizing the data in Pandas' Dataframes. The NLTK library enabled cleaning of English stopwords that were indefinite articles and nouns as these words are irrelevant for our analysis such as "might", "the", "be", "in", institution names, etc. Additionally, a list of irrelevant words was added to the stop words list imported from the NLTK library. In this list, there are words, phrases, and singular characters that were not cleaned by the original stopword list in the NLTK library. These words are not considered stop words but they are more frequent and could influence the wordcloud visualization of the responses by blurring some important words representing interesting topics, such as "college", "school", "computing", "semester", "university", etc. The NLTK library also cleaned out punctuation and further cleaned the data through lemmatization which is a linguistic process of reducing a word to its root form. In developing visualization of the findings from the data, the WordCloud library, and matplotlib.pyplot library were used. These libraries were utilized for the creation of the word clouds where the size of each word indicates its frequency or importance within the given text data from these interviews. Cleaned data was put through three topic modeling algorithms to generate the general topics from these responses explained in the following section. The topic modeling algorithms were then able to generate topic words for some key topics per question.

3.3 Topic Modeling

Three topic modeling algorithms are utilized in our survey data analysis, (i) Latent Dirichlet Allocation (LDA) [23], (ii) Correlation Explanation (CorEX) [24], and (iii) Bidirectional Encoder Representations from Transformers (BERTopic) [25].

LDA is an unsupervised algorithm, specifically used for topic modeling. It helps in identifying topics present in a collection of text documents such as an article or a collection of user responses. It assumes that documents are a mixture of topics, and topics are a mixture of words. A "topic" is defined in LDA as a distribution over a fixed vocabulary. Each word in the vocabulary is assigned a probability of being in a particular topic. Each document is assumed to be a mixture of these topics. The specific implementation of LDA that was used was LdaModel provided by the Gensim package in Python. It was chosen as it can take advantage of the CoherenceModel, also provided in Gensim. CoherenceModel quantifies the coherence or quality of the topics discovered by a topic model. Coherence measures help in comparing different topic models or configurations of the same model to find the one that best fits the data.

Anchored Correlation Explanation(Anchored CorEx) is a topic modeling technique that is semi-supervised. This means you can provide "anchors" – a set of words or phrases that guide the topic modeling process. These anchors can help the model focus on specific themes or concepts, making

the resulting topics more relevant to your particular area of interest. The specific anchor words used for the model were the smaller more uncommon words identified in the word clouds generated for the questions. The words being small suggest there could be hidden topics being overlooked by more frequent words. This helped us uncover the broader topics as well as subtle topics related to transfer decisions.

The third model used was BERTopic which leverages the power of pre-trained BERT embeddings [26] to discover meaningful topics within a collection of text documents. It operates in two main steps: document embedding and topic clustering. After converting text data into dense numeric vector representations, BERTopic applies a clustering algorithm, typically UMAP (Uniform Manifold Approximation and Projection) [27] for dimensionality reduction, followed by HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) [28] for clustering. This process groups similar texts together into topics. Although BERTopic also provides a cluster of words like LDA, it derives those clusters contextually - similar to how humans would analyze topics.

4 Results

4.1 Exploratory Analysis using Word Clouds

We analyzed the participant's responses to five open-ended questions mentioned above. The word cloud generated from the responses to question 1 (*general information*) reveals the key influencing factors when community college students are deciding to transfer to a 4-year institution: Cost, ease, opportunity, nearby, time, and tuition (Figure 3(a)). Cost and tuition are significant factors considering that the majority of the participants were in the lower bracket of income. Opportunities regarding careers and jobs are important for students who need to increase their income. Location is also important to many community college students when these students need to be close to family, especially when they are older with a family to consider. This was also confirmed when participants were asked about the most *important information* in The word cloud generated from the responses to question 2. The word cloud analysis divulges the most important information that affected the decision to transfer to a 4-year institution with these keywords: Cost(money, tuition), credit, location, and job (Figure 3(b)). Similar to the responses to the previous question, cost is essential to a student's decision to transfer to a 4-year institution. Credit refers to credits obtained from classes from the student's respective community college, the importance that credits are transferred in full also intertwines with the factor of affordability, so that transfer students do not have to pay for classes that they took at their community college at their 4-year institution. The location of the 4-year institution in regards to their proximity to their home, friends, and family. Jobs in many contexts are important as many students work part-time and attend community college and a 4-year institution to obtain the necessary prerequisites for high-paying jobs in the computing field.

Next, participants were specifically asked about their *challenges* in question 3. The challenges that students encounter while transferring from a community college to a 4-year institution are: Credit, transcript, time, major, and financial (Figure 3(c)). Depending on what type of 4-year institution that students transfer to (ex. private or public 4-year institutions) credits may transfer quickly or may be difficult to process. The responses vary, as many students had issues with

(a) Q1: General Information (b) Q2: Important Information (c) Q3: Challenge

(d) Q4: Expectation (e) Q5: Experience

Figure 3: Wordcloud generated from the responses of five open-ended questions Q1-Q5

credits while other students said that the transfer was smooth in all facets. Major may also be a difficulty, depending on whether the 2-year institution and 4-year institution match in curriculum or majors. With many possible obstacles to a smooth transition, the process can be incredibly time-consuming. With time being a major factor for students in the two previous questions, a time-consuming transfer has the potential to deter students from transferring at all.

Next, participants were asked about their *expectations* in question 4. The expectations and career paths of students had changed for transfer students in these key areas: Classes, majors, jobs, and for some students nothing changed for them (Figure 3(d)). The classes that students take at 4-year institutions may expose them to new areas or fields for them to potentially align with their careers: taking specialized classes that focus on a particular field such as cyber security has the potential to change their career path. Students also change their major as a result of being exposed to other options for their careers and to align their major with their new career path. For some students, with they were learning only solidified their career plans.

Finally, participants were asked about their *experience* in question 5. The experiences that transfer students considered to be important as computing majors were focused on: Experience, credit, resources, professors, and opportunity (Figure 3(e)). Experience in general for computing majors has an impact on how well students perform in classes and the probability that they will be hired after graduating from the 4-year institution. Transfer students have more opportunities at a 4-year institution for numerous reasons: More resources, employers having events on campus to hire students, networking with other students and mentors, etc. Professors are essential to a student's success, not only in the classroom but in mentorship. Mentorship allows students to grow in their technical and soft skills, which are essential to students thriving after graduation. The frequent words found within the word clouds were then given as parameters to all three topic modeling methods to reveal more discrete and subtle issues that plague transfer students.

Qn	Topics titles	Topic words	LDA	CorEx	BERT
Q1	Attending cost	cost, cheap, expense, year, tuition	✓	✓	✓
Q1	Career Opportunities	career, opportunity, better, job	✓	✓	
Q1	Financial Aid	financial, support, aid, assistance	✓	✓	
Q2	Saving Money	save, money, cost, cheap	✓	✓	✓
Q2	Proximity from Home	location, closeness, home, nearby	✓	✓	✓
Q2	Guidance from Family	friend, family, guidance, ease, transfer	✓	✓	✓
Q3	Harder Classes	credit, class, different, hard, time, management	✓	✓	
Q3	Difficulty Transferring Credits	credit, transcript, challenge, acceptance	✓	✓	
Q3	Adjusting to a New School	challenge, acclimating, adjusting		✓	
Q4	No Change in Plan	computer, nothing, change, none	✓	✓	✓
Q4	Major Career Change	change, major, career, class, computer, science	✓		✓
Q4	Quit Job	working, job, managing, stop, quit		✓	
Q5	Practical Experience	research, internship, summer, work	✓	✓	✓
Q5	Early Academic Advice	taking, computer, science, class, credit, early		✓	
Q5	Networking Advice	network, exposure, peers, clubs		✓	

Table 2: Aggregated topics with titles across three topic modeling algorithms (LDA, CorEx, BERTopic) with most frequent and relevant topic words (union of three sets). The checkmark indicates the algorithm on that specific column was able to detect the corresponding topic.

4.2 Topic Modeling Results

This section explains the findings from three topic modeling algorithms utilized in our study to identify important factors, challenges, expectations and experiences of students related to college transfer. Table 2 shows the top three topics with their titles and list of most frequent and relevant topic words generated by three topic modeling algorithms (union of three sets) from the responses to the five open-ended questions of the survey. The check marks in the last three columns under the three algorithms indicate if the algorithm was able to detect that corresponding topic. The first step in determining which topics to explore and report was to look at the topic's coherence score. A coherence score in topic modeling is a metric used to evaluate the quality of the topics generated by a topic modeling algorithm. The closer a score is to 1, the more the model considers it a "coherent" topic. A coherent topic in this context means that the topic has enough correlation between the words inside it to be better understood by a human reader. The coherence scores of the topic models reported in Table 2 are within range between $0.48 - 0.54$.

From the Table 2 it is evident that LDA and CorEx performed quite similarly in finding topics. Out of the three topic modeling algorithms, LDA required the most amount of data cleaning as it is necessary to filter enough general words, other than just the stop words to dig out meaningful topics. However, CorEx had a slight edge in being able to find a wider array of specific topics

because of the semi-supervision through the anchor words identified from the initial word cloud analysis. Some of the anchor words utilized for CorEx are, "social", "support", "living", "experience", "saving", "money", "course", "major", "housing", "stress", "anxiety", "study", "resource", "information", "internship", "hire", "career", "research" etc. While BERTopic was able to find some similar topics, it struggled overall to produce as much meaningful topics as the other two, because BERT is a transformer based model, hence removing stop words as a pre-processing step is not advised as the transformer-based embedding models that we use need the full context to create accurate embedding. As such, some topics generated through BERTopic end up having some irrelevant words, hence in future further filtering is needed in the word embedding before clustering to generate better topics with BERTopic. The topics were manually labeled with a topic title by two students from the list of words generated by each topic model. Any disagreement in the topic labels were resolved by the faculty mentors.

Some of the topics generated from the modeling algorithms were very broad and general. For example, almost all models produced topics involving attending costs, saving money, distacne of college from home, guidance from family, no change is plan regarding transfer, and practical experience. However, there were some topics which were only identified by CorEx, for example, some students found it challenging to adjust to the new environment of a 4-year university, some students had to quit their job to attend a 4-year university as a full time student, and some students emphasized about networking to increase exposure which is critical for a successful student life. While these help understand their decision making process and the types of challenges transfer students are facing, using these topics to further look into the data manually would provide more context to these issues.

4.3 Manual Topic Analysis

The survey responses were further analyzed by filtering the responses with the corresponding topic words and manually reading them to understand the context of the topics better. In Q1, when students were asked about the information that helped them in deciding between community college and a 4-year university, most students' reply contains "cost" as community colleges are cheaper, and some said they knew that they would transfer to a 4-year university later for better career opportunities. Some students mentioned that they were not eligible for a financial aid because of their age for a 4-year university, which is why they selected community college. In Q2, students were asked what was the main factor that would affect the decision of transfer. As the survey was conducted in five MSIs most students' response was focusing on saving money as community colleges are less costly, so completing two years of their education would cost less. Their responses include their influence of the guidance of their family and friends who also took similar path. Another aspect of cost reduction was choosing a community college nearby their home. In Q3, they were asked about the challenges they encountered during transferring, some faced difficulty in transferring credits as the courses would not match between the two institutions, some mentioned about hard classes in 4-year universities compared to community college, and some struggled to adjust in a new environment as acclimation was difficult with the general track students. One response uncovered by topic modeling had to do with misinformation as students were informed that certain credits would be accepted in a particular institution, but that was not correct. Some of the responses also mentioned the difficulty in studying while working part-time or full-time jobs. In Q4, students

were asked if their plan changed after transfer, some mentioned nothing changed as the transfer was pre-planned, for some students their whole career plan changed because of the rise of artificial intelligence, and cybersecurity, and overall computer science they changed their career path and transferred to 4-year institution for a Computer Science degree for a better career. Three responses involved having to quit their jobs to be able to attend school full-time. One response explicitly stated the classes they needed were not available after their work hours, which became the reason they left the job they were currently in. In Q5, students were asked to mention any particular experience they want to share which is important for computing major students, and some students emphasized on getting practical experience though research with faculty and summer internships, some gave the advice of taking lower division computer science courses early to be on track for timely graduation, and some gave networking advice through different clubs to get better exposure to different opportunities. One response claimed transfer computer science students will have a harder time obtaining job opportunities and internships because they believe companies "*prefer to hire students who have completed their entire degree program at a single institution*." There is no proof that being a transfer student will put them at any statistical disadvantage in the job or internship market as a student who finished their whole degree at one university. These issues reveal deeper insights outside of registering for classes and choosing a major. They reflect a lack of flexibility and support for non-traditional students who juggle employment and education, as well as misconceptions that can negatively influence students' perceptions and decisions.

5 Conclusion and future work

Rising costs at 4-year universities are bringing a shift in acquiring a bachelor's degree by attending a community college and then transferring to a 4-year institution. With this shift, many advisors and academic support have not enough knowledge of the process of transferring from the community college. There is also a lack of resources to guide students who are exploring transferring in their higher education career. This paper explores the responses from a survey of 161 students who have either transferred or are in the process of transferring to a 4-year institution. With these responses, academic advisors and transfer students seeking guidance have the ability to understand the issues that transfer students face during their transition from a 2-year institution to a 4-year institution. These students brought to light the influence of family, socioeconomic status, time, and proximity of the institution in the decision to transfer to a 4-year institution. From these responses, word clouds were generated to visualize the responses. Along with word clouds, three topic modeling algorithms (LDA, CorEx, and BERTopic) revealed the struggles of transfer students and the key factors in transfer students' success as well. The analysis identified that the cost, career opportunities, financial aid, distance from home, and guidance from family are the key factors influencing the decision between 2-year and 4-year institutions. The usage of multiple kinds of analysis, such as qualitative, frequency, and manual analysis, was necessary to discover multiple issues that plague transfer students. Frequency analysis was unable to find subtle issues that was found in qualitative and manual analysis. With the amount of participants and the number of their responses, automated analysis was also necessary to be able to cover all responses in terms of being analyzed. New issues that were revealed were the distance between the university and the university social scene (in the context of on-campus clubs). The impact of a distant commute for students is negative, with students staying on campus less and not becoming involved hinders

their success. By not participating in student organizations, the ability to network is impacted negatively. Participants also attributed proficiency in C and Java as essential to their success which were not addressed by their community college advisors before transfer. Among the 161 participants, 2 undocumented participants expressed that they had more difficulties transferring their citizenship status.

In the future, we will conduct more research to better understand some of the identified factors. For example, how staying on campus versus off campus impacts student success; more insights on undocumented students' challenges and struggles. In addition, we plan to look further into the difference of experiences between male and female transfer students. We also plan to conduct focus group interviews with advisors at various universities and seek their insights about how these findings can be incorporated into advising. Additionally, this research work will contribute to the AI-driven Counseling System for Transfer Students (ACOSUS) in the future [29].

6 Acknowledgments

This research was supported by the National Science Foundation CISE-MSI under Grant No. CNS-2219623. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] D. Lederman, "Progress, and finger pointing, on student transfer: A survey," Oct 2020.

[2] N. L. Sánchez, "An answer to increasing diversity at selective schools? community college transfer," *Forbes*, 2021.

[3] K. Reid, "Reaching the uncertain student," *Eduventures Principal Analyst at Encoura*, 2021.

[4] J. D. Edwards, R. S. Barthelemy, and R. F. Frey, "Relationship between course-level social belonging (sense of belonging and belonging uncertainty) and academic performance in general chemistry 1," *Journal of Chemical Education*, vol. 99, pp. 71–82, 2022.

[5] E. Höhne and L. Zander, "Belonging uncertainty as predictor of dropout intentions among first-semester students of the computer sciences," *Zeitschrift für Erziehungswissenschaft*, vol. 22, pp. 1099–1119, 2019.

[6] O. Almatrafi, A. Johri, H. Rangwala, and J. Lester, "Retention and persistence among stem students: A comparison of direct admit and transfer students across engineering and science," in *American Society for Engineering Education Annual Meeting, Columbus, OH*, 2017.

[7] M. R. Anderson-Rowland and A. A. Rodriguez, "Sophomore transfers: Who are they and what support do they need?," in *2015 ASEE Annual Conference & Exposition*, pp. 26–1384, 2015.

[8] M. Blekic, R. Carpenter, and Y. Cao, "Continuing and transfer students: Exploring retention and second-year success," *Journal of College Student Retention: Research, Theory & Practice*, vol. 22, no. 1, pp. 71–98, 2020.

[9] T. T. Ishitani and L. D. Flood, "Student transfer-out behavior at four-year institutions," *Research in Higher Education*, vol. 59, no. 7, pp. 825–846, 2018.

[10] S. L. Dika, K. Siarzynski-Ferrer, K. Galloway, and M. M. D'Amico, "Predicting the persistence of undeclared first-year and transfer students," *Journal of College Orientation, Transition, and Retention*, vol. 22, no. 2, 2015.

[11] P. D. Umbach, J. B. Tuchmayer, A. B. Clayton, and K. N. Smith, "Transfer student success: Exploring community college, university, and individual predictors," *Community College Journal of Research and Practice*, vol. 43, no. 9, pp. 599–617, 2019.

[12] D. Chamely-Wiik, E. Frazier, D. Meeroff, J. Merritt, W. R. Kwochka, A. I. Morrison-Shetlar, M. Aldarondo-Jeffries, K. R. Schneider, and J. Johnson, "Undergraduate research communities for transfer students: A retention model based on factors that most influence student success," *Journal of the Scholarship of Teaching and Learning*, vol. 21, no. 1, 2021.

[13] J. M. Lakin and D. C. Elliott, "Steming the shock: Examining transfer shock and its impact on stem major and enrollment persistence," *Journal of The First-Year Experience & Students in Transition*, vol. 28, no. 2, pp. 9–31, 2016.

[14] F. C. Lane, G. L. Martin, and R. K. Henson, "A multidimensional comparison of traditional, transfer, and online students' university attachment," *Journal of College Student Development*, vol. 56, no. 7, pp. 746–751, 2015.

[15] K. Y. Walker and C. Okpala, "Exploring community college students' transfer experiences and perceptions and what they believe administration can do to improve their experiences," *The Journal of Continuing Higher Education*, vol. 65, no. 1, pp. 35–44, 2017.

[16] J. M. Allen, C. L. Smith, and J. K. Muehleck, "Pre-and post-transfer academic advising: What students say are the similarities and differences," *Journal of College Student Development*, vol. 55, no. 4, pp. 353–367, 2014.

[17] R. Harper and H. Thiry, "Advising from community college to university: What it takes for underrepresented transfer students in stem to succeed," *Community College Journal of Research and Practice*, vol. 47, no. 9, pp. 582–601, 2023.

[18] G. Crisp and C. Delgado, "The impact of developmental education on community college persistence and vertical transfer," *Community College Review*, vol. 42, no. 2, pp. 99–117, 2014.

[19] C. Clausen and R. D. Wessel, "Transfer shock: Predicting academic success after transition," *Journal of College Orientation, Transition, and Retention*, vol. 23, no. 1, 2015.

[20] X. Deng, Y. Fernández, and M. Zhao, "Social media use by first-generation college students and two forms of social capital: a revealed causal mapping approach," *Information Technology & People*, vol. 35, no. 1, pp. 344–366, 2022.

[21] C. Robinson, M. Yeomans, J. Reich, C. Hulleman, and H. Gehlbach, "Forecasting student achievement in moocs with natural language processing," in *Proceedings of the sixth international conference on learning analytics & knowledge*, pp. 383–387, 2016.

[22] R. Rincon, "A descriptive study of community college transfers in engineering and computer science in texas," in *2018 CoNECD-The Collaborative Network for Engineering and Computing Diversity Conference*, 2018.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[24] R. J. Gallagher, K. Reing, D. Kale, and G. Ver Steeg, "Anchored correlation explanation: Topic modeling with minimal domain knowledge," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 529–542, 2017.

[25] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022.

[26] J. D. M.-W. C. Kenton and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of naacL-HLT*, vol. 1, p. 2, 2019.

[27] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.

[28] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering.," *J. Open Source Softw.*, vol. 2, no. 11, p. 205, 2017.

[29] "AI-driven counseling system for transfer students." https://cs.neiu.edu/acosus/. Accessed on 15 September 2023.