Macao Polytechnic Institute
School of Business

Approach B

Exploring university brand communication in social media: An application of text analytics

Supervisor: Dr Linda Lai

<u>Group Members:</u>

CHEN JINGRU P-16-0243-1
HE WENYA P-16-0893-3
ZHU JIAQI P-16-0897-1

Abstract

Purpose of the study

This project aims to apply text analytics to assess brand communication by Chinese universities on social media. While China has now become a superpower with great influence in the international arena, the country's education system is still in need of improvement. China has gradually launched a series of policies to boost the development of the education system in order to increase China's soft power, but there are still fewer international students studying in China than in developed countries. Further, social media has come to have an important role in branding, and it would be useful for Chinese universities to build up their brands as a way of attracting more international students. The driving force for our study is thus to enhance the soft power of China by building strong university brands on social media.

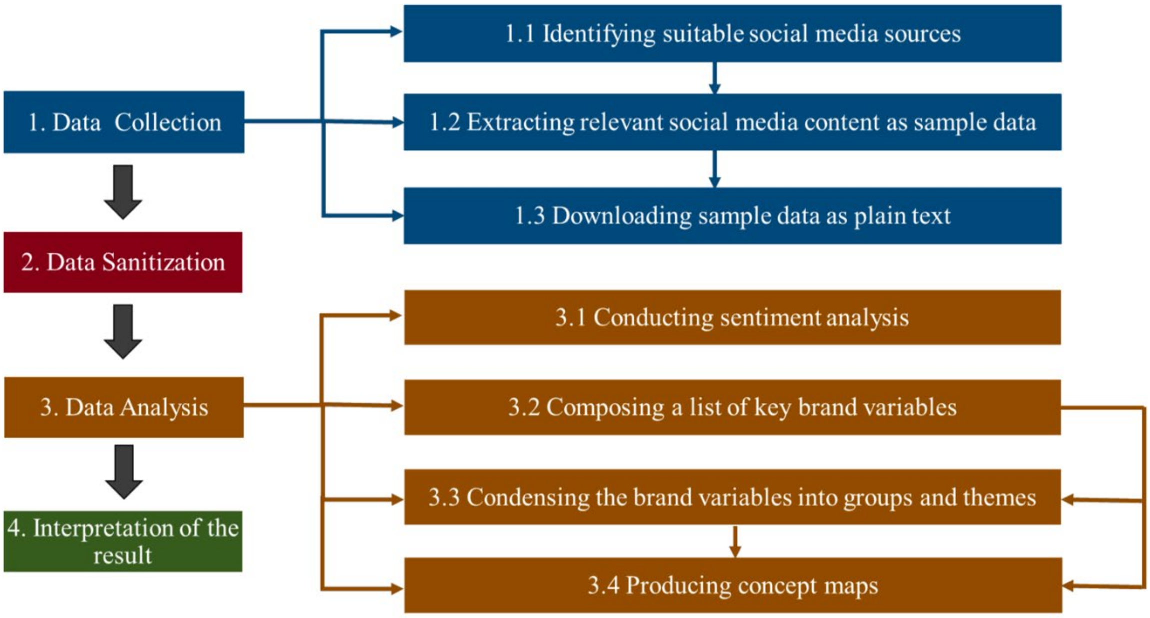
Our research objectives are twofold. First, we hope to learn how universities promote their brands through social media posts and find ways for Chinese universities to use branding to attract more overseas students. Then, we aim to demonstrate the use of social media analytics in generating insights into university brand promotion.

Research design/methodology/approach

We applied our adopted text-analytics methodology to study Chinese university branding. The figure shown below illustrates the flow diagram of our methodology, which involved four main stages: data collection, data sanitization, data analysis and interpretation of the result.

Based on the criteria of international, representative and accessible, we chose Twitter as our data source, and a total of 1,040 pieces of data were extracted from the social media platform. In the data sanitization stage, we found and corrected data to ensure the text used for analysis was correct and had a clear meaning to obtain effective results. The sample data were analysed using the following software tools: Wordsmith, SPSS and Leximancer. We first used Wordsmith to generate a list of key

brand variables based on lexical analysis. SPSS was then used to perform factor analysis, which enabled us to summarize and group our brand variables into dimensions. Last, we used Leximancer to produce concept maps, which provided a holistic view of the university from different perspectives. In the last stage, we gained insights into the branding strategies of Chinese universities on social media.

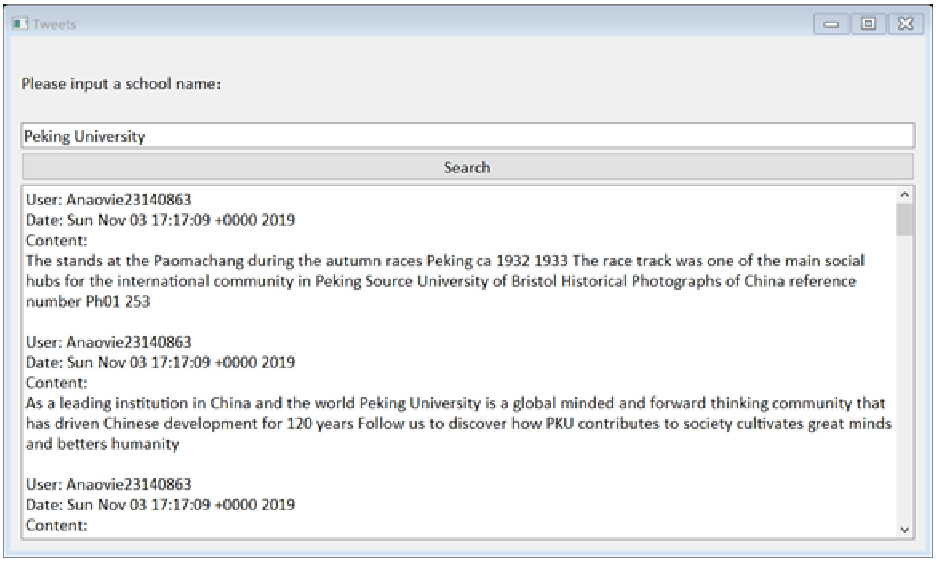


Software development

We developed a software tool to download relevant data on Twitter using Python programming language. Before starting the software development, we had to get a Twitter API by registering for a developer account on the Twitter Developer Community. The API is a permission which allows us to access Twitter data.

After that, we installed Tweepy in our Python code. Tweepy is a Python program that helps to download tweets from Twitter, and it also includes many functions that can help developers to maintain their Twitter accounts or to download relevant tweets with specific criteria. When we have set up the criteria for the tweets that we want to download, the next step is to create several spaces to store our data. In our development, we created several lists to store the content, date, username and other elements that make up the components of the tweet. Tweepy has other functions besides downloading data, such as the ability to check and remove duplicate data when we downloaded and removed all the punctuation and symbols from each tweet.

We also developed a graphic user interface (GUI) to make it easier for users to use this software tool, as seen in the following image. Users can type the keywords they want to search in the input bar. The image gives the example of a search for Peking University as the keywords, meaning that the user wants to search for the tweet contents which mention Peking University. After clicking the 'Search' button, the result will be displayed in the big box below.



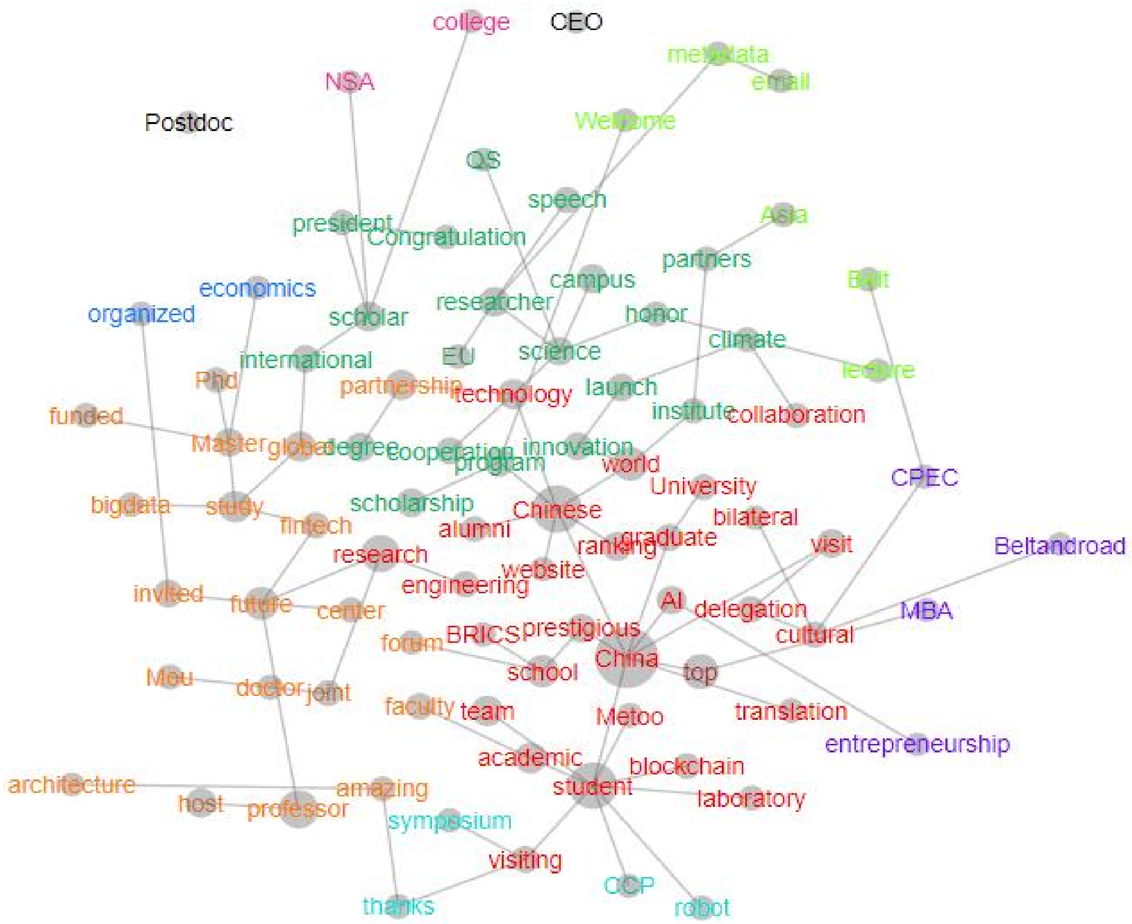
Key findings of the study

We collected a data set of 1,040 tweets on Twitter which were discussing the top seven Chinese universities of the C9 League (Tsinghua University, Peking University, Fudan University, Shanghai Jiao Tong University, Zhejiang University, Nanjing University and University of Science and Technology of China). In addition to the content, we also recorded the engagement metrics for each tweet, such as the numbers of images, videos, likes, comments and retweets. According to the engagement metrics for the tweets, we found that there was high correlation between 'Like' and 'Retweet', 'Like' and 'Comment', and 'Comment' and 'Retweet', but that 'Comment' and 'Image' have negative correlation. We also found that Twitter users prefer to use the 'Like' and 'Retweet' functions but rarely write comments.

We generated 95 keywords based on the content of 1,040 tweets. They represent the high-frequency terms that occur in user-generated content when people discuss the universities. So, a keyword list can be regarded as the key brand variables for the universities. ‘China’ had the highest keyness in the list of key brand variables. The most positive words such as ‘amazing’, ‘great’, ‘welcome’ and ‘prestigious’ became key brand variables. A lot of people discussed university rankings, scholarships, advanced technologies and degrees. Moreover, we generated key brand variables for Tsinghua University (THU) and Peking University (PKU) in order to compare the key brand variables between the top two university in China. This comparison showed that ‘China’ has higher keyness in PKU than in THU. More advanced technology variables appeared at the top ranking of the key brand variable list in THU, and more variables about the university community appeared in PKU.

After generating the key brand variables for the target universities, we grouped these variables by factor analysis using SPSS. We dropped the variables when their rotated loading was under 0.3, and the remaining 66 variables were grouped in 27 factors. Then, we classified the 27 factors into seven branding dimensions: ranking, degree, exchange, administration, discipline, technology and research. According to the distribution of these branding dimensions, ranking, administration, discipline and degree occupy more than 70% of the branding for Chinese universities.

We used Leximancer to generate concept maps based on our collected tweets. The following figure provides a holistic view of the Chinese university brand communicated on social media. The colours represent the importance of the concepts and classify the clusters of the concepts. Hot colours such as red and orange denote more important clusters, while the cool colours denote less important clusters. The size of each node is related to its connectivity in the concept map: the larger the node, the more often that concept is mentioned with other concepts in the map. The most important concept is ‘China’, and the concepts closest to the ‘China’ node are ‘top’ and ‘prestigious’. We also found some concepts which represent positive attitudes in the map, like ‘top’ and ‘prestigious’ in the red cluster.



Discussion and Implication

Our demonstration of text analytics on social media users' comments can help universities better understand the use of social media for their branding. Twitter is a popular and international platform among English-speaking users with plenty of useful data for our study. Using our analysis of the relationship between social media and university branding, Chinese universities will be able to raise the proportion of international students, especially English-speaking students. Hence, the soft power of China would be improved as Chinese universities attract more English-speaking international students by strengthening their branding images.

One thing that needs to be noted is that the users' comments about each Chinese university are in unstructured data form. Since comments that come from social media are dependent on the user's own thoughts, and they are varying and subjective, this textual data should be transformed into structural form. Following our methodology steps and effective analysis approach, the deep value of people's comments for each university could be mined out, and universities could make suitable strategies for their branding improvements by making good use of social media platforms.

Limitation and Recommendation

During the whole research process, we also found several limitations in our study. For each limitation, we have provided a related recommendation to improve future studies.

The data volume was low for our sampled universities, giving us a small data set for analysis. The insufficient data volume may cause the findings to lack generality. We recommend that our data sources could be extended to other social media platforms rather than only focusing on Twitter.

Unstructured data is another challenge for us to handle. Though we transformed the textual data into structured data form, some contents may still be irrelevant and useless for our study. We suggest not only using some advanced applications but also mining the deep value of other types of data, such as image, audio and video.

Also, the language setting limited the available data for our whole research. Since we decided on English as the standard data language form at the beginning, some content in other languages would be dropped, so it is possible that we lost some valuable content. We may need some useful translation tools or analysis methodologies to avoid such a situation in future studies.

Moreover, data veracity may also have affected our study result, as any maliciously manipulated data would cause our findings to be unreasonable. We may need to verify the credibility of Twitter and its users' reputation through a comparison with other platforms, then check the facts and use the relevant advanced technology to avoid such problems.

After overcoming these limitations and following the relevant recommendations, we believe our study could be more meaningful and trustworthy.

Conclusion

Text analytics is the major analysis method we used for this study. This method is useful for textual data, which helped significantly with interpreting the unstructured data.

By using text analytics, we explored the relationship between social media and universities' branding, and we analysed user comments for each university we researched. Then we drew conclusions about how universities could improve their branding through social media communication among users.

Reviewing the whole process of our study, we recognised that we lacked some of the necessary knowledge and skills for effective text analytics. In further studies, we would be able to improve our analytics process with more knowledge and skills.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

In this chapter, we provide an overview of our study.

1.1 The study rationale

Since the 1990s, China has made remarkable achievements in its soft power resources and its ability to turn these resources into advantageous foreign policy outcomes. China has now become a superpower with great influence in the international arena, and its strength should not be underestimated.

However, China's education system is still in need of improvement. As the world's most populous country, China has gradually launched a series of policies to boost the development of the Chinese education system in order to increase the soft power of China. Although many Chinese universities have created programs to attract international students, and the number of international students in China is rising each year, there are still fewer international students studying in China (especially English-speaking students) than in developed countries (Wikipedia, 2019). We should consider whether the attractiveness of Chinese universities is lower than that of foreign universities and examine how to improve the brand of Chinese universities. Furthermore, social media has come to have an important role in branding. It would be useful for Chinese universities to build up their brands. The driving force for our study is thus to enhance the soft power of China by building strong university brands on social media.

1.2 Definition of the problem

In order to promote the image of Chinese universities, we need to listen and understand the brand stories and conversations that are exchanged on international social media platforms such as Twitter. However, social media content is noisy, messy, ill-structured and cannot be processed via traditional data-mining methods. The greatest barrier to using social media data is the lack of appropriate methodology when selecting, collecting, and analysing the vast amounts of user-generated content that is freely available on the Internet. Through our project, we proposed a text analytics approach for social media content analysis.

1.3 Aim and objectives of the study

In our project, we used text analytics to assess brand communication by Chinese universities on social media. The study has a two-fold objective: brand building for Chinese universities and demonstrating the use of social media analytics.

1.3.1 Brand building for Chinese universities

Universities face fierce competition when recruiting students and thus they need to identify ways to stand out among each other. Building a brand is an essential way to enhance the reputation of a university and prove its strength. University branding delivers critical information on the institute's reputation, its quality of education and its campus facilities during the decision-making process of potential students. A strong university brand communicates to potential students that studying in the chosen university can add value to their lives and will be an enjoyable experience (Ali-Choudhury et al., 2009).

We hope to learn how universities promote their brands, in order to understand how China can become the largest study-abroad destination in Asia. We do this by studying the social media posts of Chinese universities. Twitter is our target data source because it has many active users from different nationalities, and it is thus easy to communicate with people and to survey a greater number of opinions. Following this, we want to find ways that Chinese universities can attract more overseas students and create brands that are more international.

1.3.2 Demonstrating the use of social media analytics

Social media allows participants to express opinions freely and honestly. Because of this, many organizations, including universities, use social media data to inform their decisions and to acquire actionable strategic advice. Social media is also a great platform for building brands because it provides a connection for stakeholders of universities. Universities can listen to the true voice of stakeholders. Because it is so popular, universities actively use social media to increase their exposure to stakeholders, including potential students, and to listen to their voices. This study aims to demonstrate the use of social media analytics when generating insights into university brand promotion.

1.4 The study contribution

Our study offers two main contributions. The first is toward the improvement of the branding of Chinese universities. The second is toward the use of text analytics for social media content analysis.

This study could help Chinese universities to better understand how to build a good university brand through the use of social media, and this could then boost the development of China's soft power and increase the country's international influence. As well, the information gathered could enable Chinese universities to attract more international students and create more diverse campuses.

Using analytics allows us to apply a method that is more systematic and efficient. When using text analytics, we have finally developed a web crawler to efficiently collect textual data from social media users and mined these data to form our conclusion. The data we collect from social media are unstructured, so our analysis is convincing. When looking at social media analytics, we can better understand the structure of the social media platforms and how users interact with each other.

1.5 Scope of the study

This project applies text analytics to explore Chinese university branding on social media. Our study targets are the C9 League universities with a focus on Tsinghua University and Peking University. We have selected Twitter, the most popular microblogging online platform, as the source of our social media data. Our sample data are all written in English as it is the international language for global audiences.

1.6 The study workflow

Figure 1 shows the workflow of our study. We first identified the focus of our study and looked at the branding of Chinese universities on social media as our area of investigation. We collected relevant information such as statistics, trends and national policies on higher education. We then conducted a comprehensive literature review of related concepts, such as text analytics, social media and branding. A conceptual framework was formulated based on previous influential studies. Afterward, we studied the text analytics methodology including the incorporated phases, steps and software tools. We chose Twitter as our data source and developed a web crawler using Python to collect data from Twitter. Then, we followed the project methodology to conduct the research. Later, we obtained project results and findings. Learnings were generated from the findings. Finally, we drew conclusions and made suggestions and recommendations.

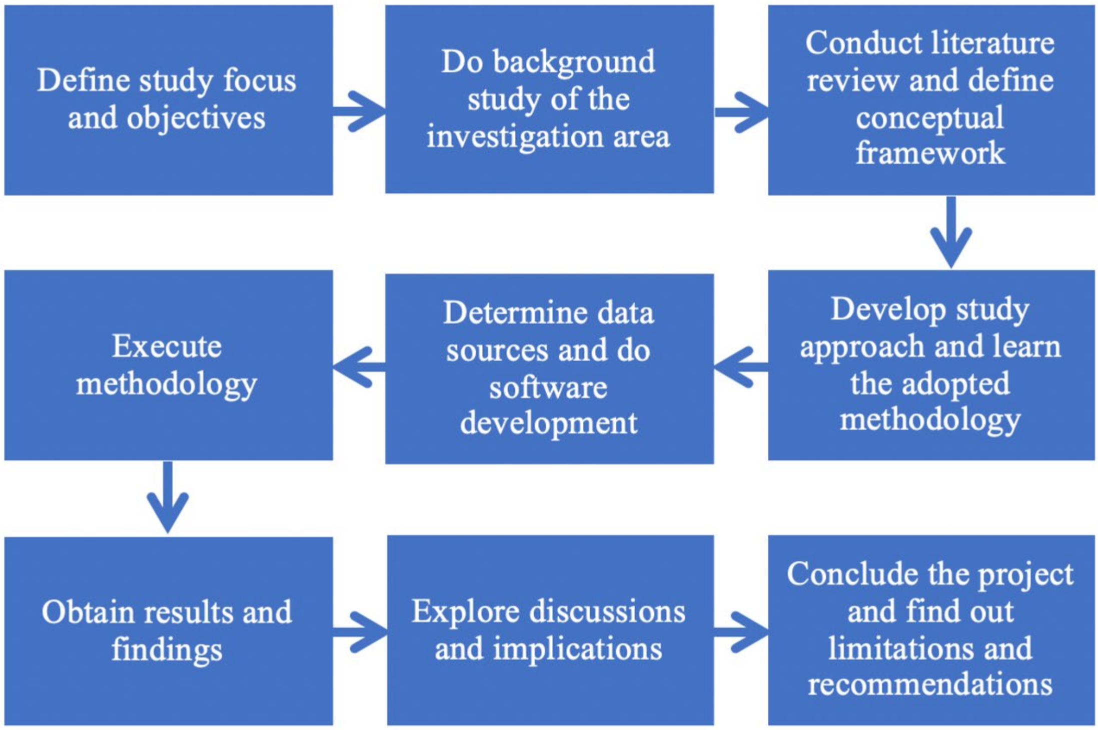


Figure 1: The study workflow

Chapter 2: Background of the Study

2.1 China's rising power

For the last few decades, the world has witnessed China's rise, mostly in terms of hard power, which is the use of military and economic means to influence the behaviour and interests of other political bodies. For instance, the socialist market economy of China is now the world's second-largest economy by nominal GDP (World Bank, 2019a) and the world's largest economy by purchasing power parity (World Bank, 2019b). China's armed forces, the Chinese People's Liberation Army, are the largest military force in the world and have the world's second-largest defence budget (Wikipedia, 2019).

2.2 Higher education as an instrument for China's soft power

However, a number of changes in world politics stand out as especially transformative forces elevating the utility of soft power relative to hard power. China has the advantages of being a nation with rich culture and rising hard power. Thus, more and more international students have been attracted to and enrolled in Chinese universities. The Chinese government believes that soft power, especially through higher education, is vital in the country's development. Accordingly, a series of policies has been implemented to emphasize this. The following are a few milestones for higher education's development in China.

2.2.1 Project 211

Firstly, in 1995, the Ministry of Education of the PRC initiated a project called Project 211 to raise the research standards of high-level universities and cultivate strategies for socio-economic development. The name of the project is an abbreviation of its slogan, "In preparation for the 21st century, successfully managing 100 universities" (Ministry of Education, PRC, 2016).

2.2.2 Project 985

On May 4, 1998, another project, Project 985, named after its date (98/5, or May 1998), was first announced by Chinese President Jiang Zemin to promote the development of the Chinese higher education system by founding world-class universities in the 21st century (Ministry of Education, PRC, 2011). It was noticeable that the original nine founding member universities of Project 985 formed the C9 League (Wikipedia, 2019), which is referred to as the Chinese equivalent of the US Ivy League.

2.2.3 Education Modernization Plan 2035

Education has always been a vital part of Chinese soft power development. Recently, a development plan was made public on 23 February 2019 which outlined blueprints for the sector's modernization and development by 2035 (State Council, PRC, 2019). The plan set the goals and tasks for 2035 and detailed how the education modernization drive would be facilitated from 2018 to 2022. The Education Modernization Plan 2035 was released by the Central Committee of the Communist Party of China (CPC) and the State Council. In the Plan 2035, eight goals were proposed, highlighting the accessibility of high-quality education from preschool to higher education stages. It also specified 10 strategic tasks, including ensuring equal access to basic public education services, building world-class universities, and opening education further to the world.

2.3 C9 League universities

Among China's education policies, the most famous and successful is called the C9 League (Wikipedia, 2019). The C9 League, referred to as C9, is the first university alliance among China's top universities. It was launched in October 2009. The members of the C9 League are Peking University, Tsinghua University, Fudan University, Shanghai Jiao Tong University, the

University of Science and Technology of China, Zhejiang University, Nanjing University, Xi'an Jiao tong University, and Harbin Institute of Technology, which are all the first batch of Project 985 elite universities in the country.

C9 League universities share a special resource arrangement with one another and have been co-cultivating top talents since 2009. The leaders of the nine universities jointly signed an agreement on talent exchange and training cooperation which aims to strengthen cooperation and exchanges in the fields of personnel training and scientific research, and to complement each other. If the content of the subject or the quality of a professor's teaching is not good, a student can apply to the university with the best teaching resources among the nine universities. After the application is approved, the student can go to the other university to study, and the original school will admit the student's grades and credits when he or she finishes the subject and returns to the original school.

2.3.1 Global branding of C9 League universities

At present, the C9 League universities cooperate with other well-known international colleges and universities, such as the British Russell University Group, the Japan Academic Research Forum, the German U15 University Alliance, the Canadian U15 Alliance, the Australian Eight School Alliance, etc., to maintain good relations for better development. After the establishment of the league, a number of substantive activities were launched, such as exchange of international students, convening seminars, summer camps, etc. Through global communication, the C9 brand image has been gradually built up, and the international impact has been expanded.

The members of the C9 League are mostly in a stable rising stage in the world university rankings. Figure 2 shows the changes in the nine universities' world rankings from 2012 to 2020.

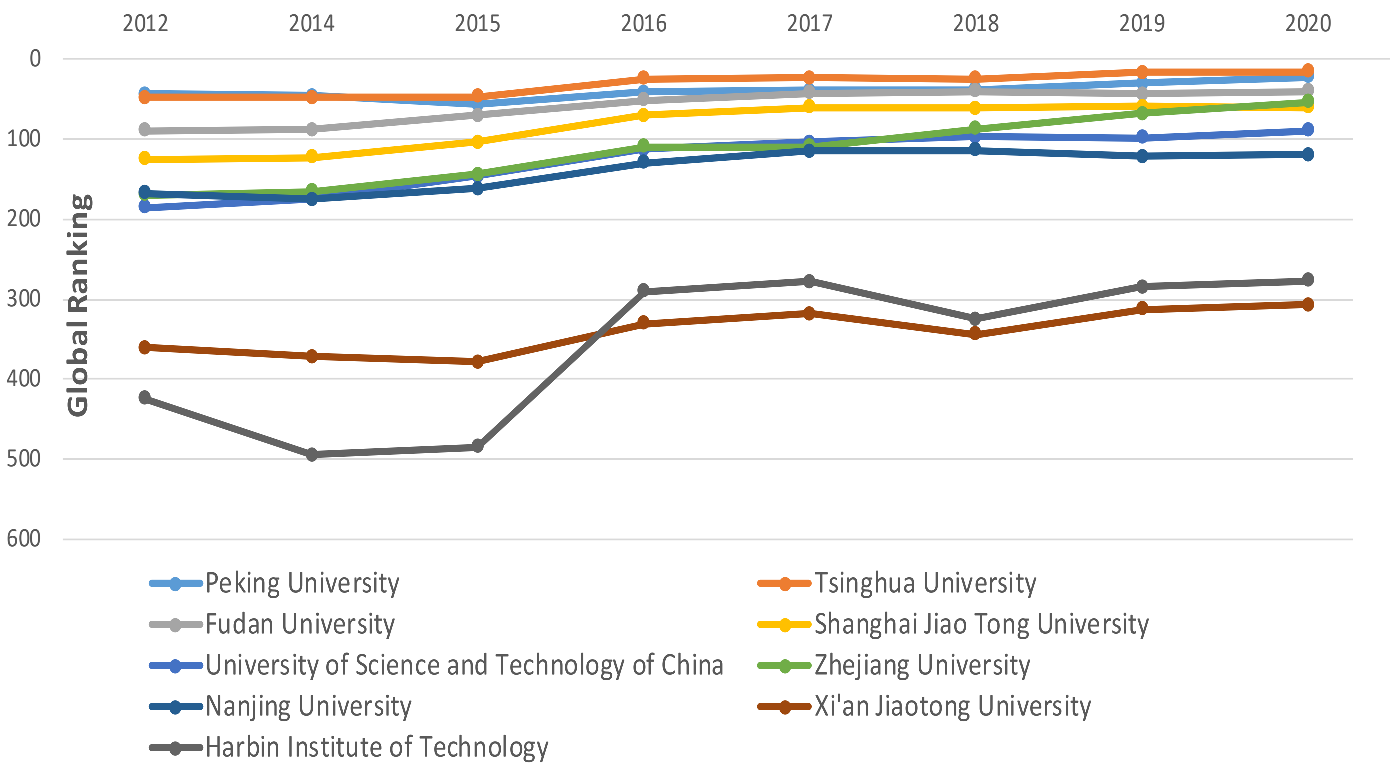


Figure 2: Global ranking of C9 League universities from 2012 to 2020
(Source: https://www.topuniversities.com/qs-world-university-rankings)

2.4 Peking University and Tsinghua University

Peking University, established in 1898, is a member of the C9 League and has educated many of modern China's prominent thinkers, scholars and politicians. Peking University has a long history and is regarded as a symbol of Chinese culture. The development of Peking University and its response to challenges have always been of concern to the Chinese people (Hayhoe, Zha, & Yan, 2012).

Tsinghua University, established in 1911, is also a member of the C9 League and it has been attended by Chinese leaders in politics, business, academia, and culture (Wikipedia, 2019). The remarkable research achievements of Tsinghua University in the fields of science, technology, engineering and mathematics research have become a point of pride in China (The Economist, 2018).

Due to their national and international rankings, Peking University and Tsinghua University are the top Chinese universities within the C9 League. The total number of the top five world-class disciplines at Peking University and Tsinghua University in 2019 were 38 and 32 respectively.

Among Peking University's 38 world-class (top five) disciplines, 16 were in the humanities and social sciences, 18 were in the natural sciences, and 4 were in engineering (see Figure 3).

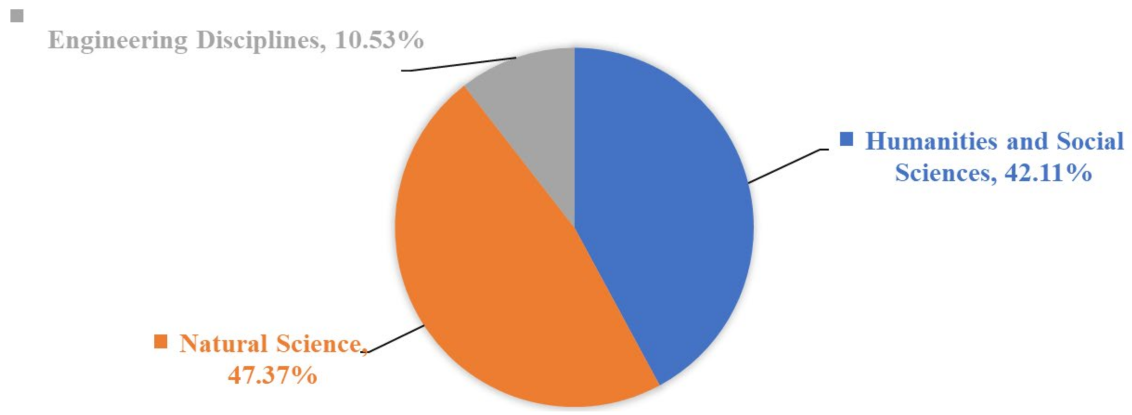


Figure 3: Distribution of Peking University's 38 world-class (top five) disciplines
(Source: http://www.gzgddi.com/index.php?m=content&c=index&a=show&catid=3&id=231)

Among Tsinghua University's 32 world-class (top five) disciplines, 5 were in humanities and social sciences, 6 were in natural sciences, and 21 were in engineering (see Figure 4).

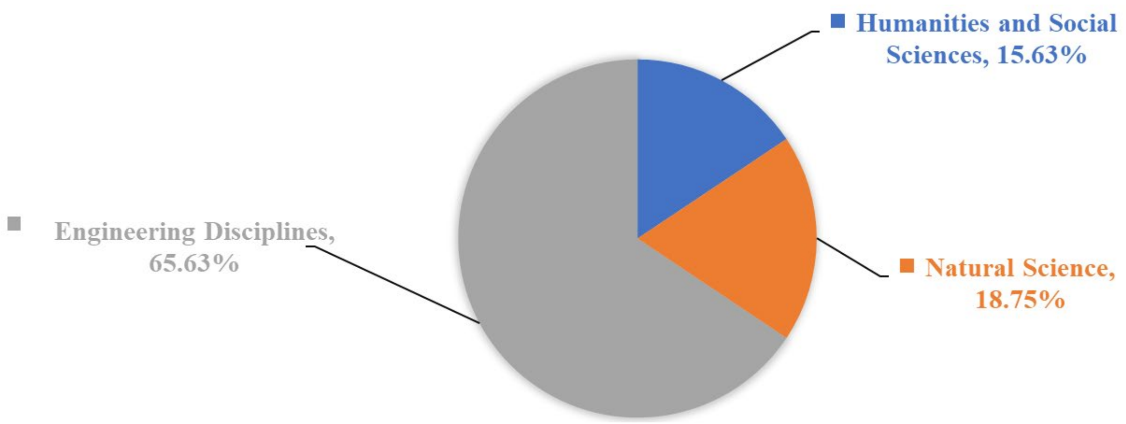


Figure 4: Distribution of Tsinghua University's 32 world-class (top five) disciplines
(Source: http://www.gzgddi.com/index.php?m=content&c=index&a=show&catid=3&id=231)

Therefore, it can be seen that Peking University is focused on the study of liberal arts while Tsinghua University is focus on science and engineering. Hence, the brands of the two universities are different. Peking University is more suitable for international students who study liberal arts while Tsinghua University is more appealing to students who prefer science and engineering.

In addition, the two universities have planned to internationalize their campuses through the recruitment of international students. They have both attempted to attract these students through promotional activities and scholarships. However, the proportion of international students attending Chinese universities is still lower than it is in the top universities of English-speaking countries (Handbook on Globalization and Higher Education, 2011). Hence, the brands of these two universities need to be improved and made more appealing for international students.

2.5 Values of internationalization of higher education to China

Internationalization is a trend in higher education around the globe, including in China. It brings significant benefits to China in areas such as the country's reputation, cultural influence, international relationships, as well as social and economic development.

2.5.1 International relationship

As the growing power in the east, China is sometimes unknown or misunderstood by other countries, especially western ones. By attracting and educating international students, China will be able to showcase the real Chinese culture, values, political system, and economic model. Early in 2010, the Ministry of Education of China issued the Studying in China Plan (State Council, PRC, 2010). In the plan, China set its goals of being the biggest education destination in Asia and having 500,000 international students in China by 2020. In 2018, the number of international students was already 492,185 (Ministry of Education, PRC, 2019).

2.5.2 Economic development

According to the Ministry of Education of China, in 2018, only 63,041 international students, or 12.81% of the total, were on government-funded scholarships, and the rest were self-supporting (Ministry of Education, PRC, 2019). Unlike in other countries, such as the US and the UK, the exact economic contribution of international students to China is not well documented.

However, international students can bring more than tuition fees. An education alliance is part of China's Belt and Road Initiative, a significant policy for economic and other development in China (Wikipedia, 2019). With the founding of the university alliance of the New Silk Road Initiative (Wikipedia, 2019), more international students are coming to China to study, which enhances and strengthens the relationships and ties of countries along the Belt and Road Initiative route. Furthermore, the New Silk Road Initiative could boost the trading between Central and South Asia and enhance the development of China's economy.

In August 2019, 12 more convenient immigration policies were adopted to attract more international talent to work and start business in China (State Council, 2019). International students are an important group that the policy was aimed at because they already have living experience in China, and most of them know China and even can speak Chinese.

2.6 International students in China

In 2018, there were 492,185 international students from 196 countries and areas studying in China (Ministry of Education, PRC, 2019). They came to 31 different provinces and cities to perform different levels of studies.

According to the statistics (Ministry of Education, PRC, 2019), the largest part of students came from Asia, which accounted for 59.95%. The second-largest continent international students came from was Africa, comprising 16.57%. There were also students from Europe, America, and Oceania, although the number was relatively few. Figure 5 shows the distribution of international students in China from different continents.

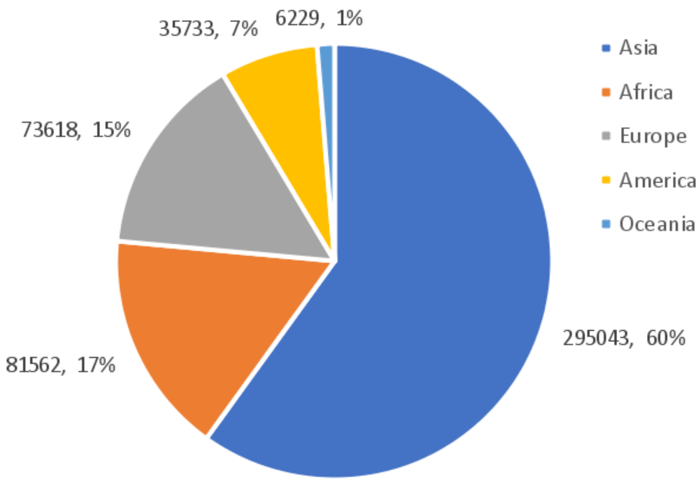


Figure 5: Distribution of international students by continents
(Source: www.moe.gov.cn/jyb_xwfb/gzdt_gzdt http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/)

In terms of nationalities, 50,600 international students were from Korea, which was the largest source of international students in China in 2018. The countries that sent the second- and third-most students were Thailand and Pakistan, respectively (Ministry of Education, PRC, 2019). Figure 6 shows the top 10 source countries of international students in China in 2018.

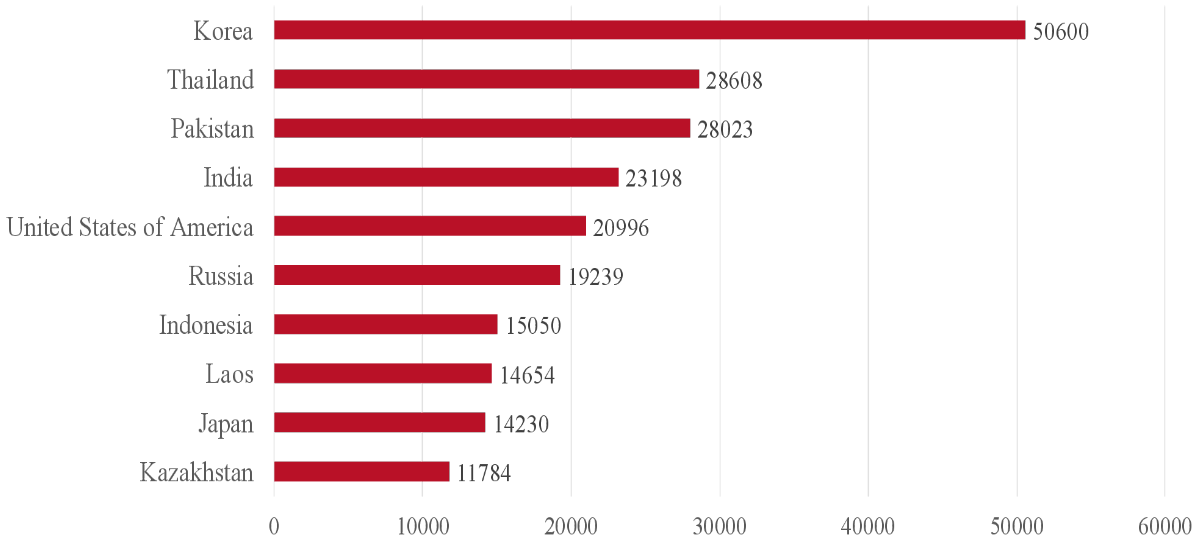


Figure 6: Top 10 source countries of international students in China, 2018
(Source: http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/)

According to the statistics (Ministry of Education, PRC, 2019), international students came to China for different levels of studies. There were 258,122 international students who came to China for degree studies in 2018, an increase of 6.86% compared with 2017. The distribution of levels of studies for international students is shown in Figure 7.

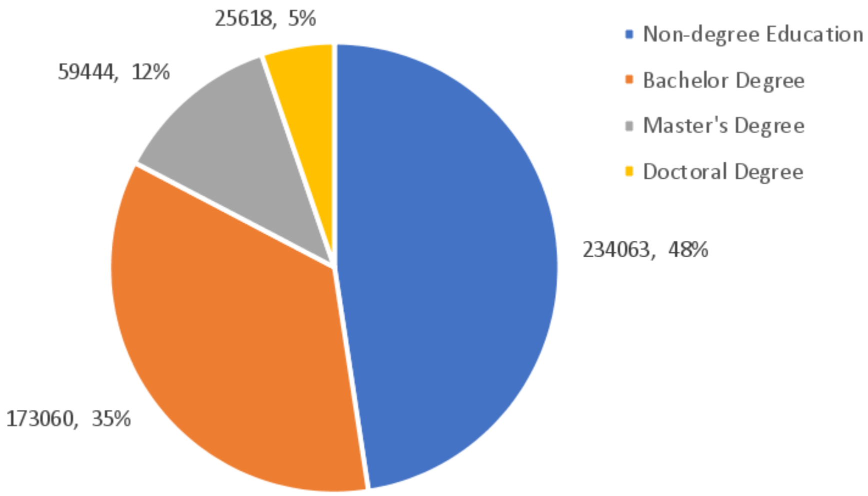


Figure 7: Distribution of international students by levels of study
(Source: www.moe.gov.cn/jyb_xwfb/gzdt_gzdt http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/)

As illustrated by Figures 5 and 6, the international students coming from mainstream western countries were relatively fewer than those from other countries. However, the attraction of students from those mainstream western countries is extremely vital considering the value of higher education to the building of soft power of China.

2.7 Information sources on Chinese universities for international students

International students gather information on Chinese universities from three information sources: social media, traditional media, and word-of-mouth (Rutter, Roper & Lettice, 2016).

- **Social media** is a revolutionary way of communication (see Chapter 3). It facilitates online communities of people to share information, knowledge, and opinions using conversational media. The most popular social media sites include Facebook (and its associated Facebook Messenger), YouTube, WeChat, Instagram, QQ, Q-Zone, Weibo, Twitter, Tumblr, Telegram, Baidu Post Bar, LinkedIn, LINE, Snapchat, Pinterest, Viber, VK, etc.
- **Traditional media** is generally regarded as the media that has nothing to do the Internet. Universities release information to the public through roadshows, school visits, newspapers, radio and televisions. Even traditional media is an old media, it is still used by most Chinese universities for communicating their brands to international students.
- **Word-of-mouth,** or viva voce, refers to oral communication between people. It can be simply telling a story or a fact to spread information from person to person.

Based on these three major sources, international students can search the Internet, communicate with people online, read a news article, or listen to their parents' advice to learn information about Chinese universities.

2.7.1 Twitter as an important information source on Chinese universities

Overall, social media is the major information source where international students learn about Chinese universities. They search a popular social network for the content they want, and it will give them as many results as possible. There are many useful and famous social platforms that provide various function and tools to help their users to find out what they like, what they need, and what they care of.

As Veletsianos (2012) indicated, schools are participating on Twitter for engagement and contribution to network practices. They can share information and resources about their classrooms and students, request assistance from and offer suggestions to others, and engage on Twitter. International students may discover relevant information posted by scholars. Twitter is a useful information resource among social media.

Twitter is a representative and popular platform with around 320 million active users (Clement, 2019) at present that aims to help users find the information they are interested in and allows them to interact with other users. Also, users can comment on other people's posts or browse other people's comments. The main function of Twitter is the stream of tweets that users see when they log in. These tweets are posted by the people they follow and are sorted in reverse chronological order. When international students search for information about Chinese universities on Twitter or follow a relevant account, they will gather a huge number of tweets based on their needs.

Our project aims to investigate how Chinese universities' brands are communicated to international audiences, including students. We thus plan to mine Twitter data for an understanding of Chinese universities' branding on social media.

Chapter 3: Literature Review and Intellectual Framework

In this chapter, we define important concepts related to our study: text analysis, social media and university branding. We also provide a comprehensive literature review of previous studies and formulate an intellectual framework to guide our investigation.

3.1 Text analytics

Text analytics is the process of acquiring knowledge from textual data through linguistic and statistical techniques. It helps enterprises transform large numbers of unstructured data into useful information to assist them in decision-making. Social media content, emails and questionnaire responses are examples of unstructured textual data. There are several methods of text analytics: information extraction, text summarisation, question-answering and sentiment analysis (Gandomi & Haider, 2015).

Rooted in text mining, the objective of text analytics is to explore patterns that people usually do not notice. It is another form of data mining, also known as intelligent text analysis, text data mining or knowledge-discovery in text. Since major information is stored in the form of text, text analytics is considered a high-potential, commercially valuable technology (Gupta & Lehal, 2009).

3.1.1 Application areas of text analytics

In our study, we intend to examine text analytics from a business point of view. In other words, we consider this technique a method to discover business opportunities. In this sense, the technology behind text analytics is just a means to an end. We are concerned about how analytic results can be interpreted and how actionable recommendations can be formulated from them to facilitate the decision-making process (Edgington, 2011).

Text analytics is mainly applied in four areas: security applications, marketing applications, analysing open-ended survey responses and automatic processes for emails and messages.

- Security applications: Text analytics is used to monitor and analyse illegal or unethical behaviour on online platforms such as social media, online stores, news websites, etc.
- Marketing applications: Marketers can ascertain consumers' preferences and experiences by analysing textual data, which allows them to improve their products or services.
- Analysing open-ended survey responses: Companies administer surveys to their customers for their opinions or feedback and then analyse the resulting textual data.
- Automatic processes for emails and message: Text analytics can help enable systems to filter and classify emails through big data analysis. This allows the email system to sort emails into their respective categories (Verma, Agrawal, Patel & Patel, 2016).

3.1.2 The importance of text analytics

Regarding why we have decided to use text analytics as the foundation of our investigation, we first consider the value of text compared to numerical data. Text created by people in natural language delivers information that goes beyond facts, including themes and topics, events and relationships, and opinions and emotions (Grimes, 2008).

The generation of big data (Diebold, 2012) has overtaken businesses, and it is always of great importance for an organisation to process data with the purpose of generating valuable implications. In fact, it is estimated that about 80% of enterprise-relevant data comes in the form of text and other unstructured information (Grimes, 2008). If an organisation can extract value from textual data, it will have a competitive advantage that benefits its business activities and contributes to its sustainable growth as a whole.

Given the significance of analysing unstructured data, a helpful tool to manipulate data is necessary. However, conventional data analysis methods do not offer solutions to this problem. The main challenge of unstructured data is searching, as the conventional method only involves finding content that people already know instead of discovering material people are not yet aware of (Grimes, 2008). Text analytics improves the search process by providing more relevant content as well as more closely linked data from textual and transactional sources.

3.2 Social Media

A social media platform is a content-based production and exchange online platform based on user relationships. It allows users to interact with other people by creating and sharing information, ideas, personal messages, career interests and other forms of expression via virtual communities and networks. The information social media platforms have spread has become prominent in online searches.

According to Our World in Data (2019), there are 7.7 billion people in the world and a third of these people are using social media platforms. People can receive information from social media through their mobile phones at any time and in any place. Generally, younger people are more active on social media platforms. The popularity of social media provides a great potential platform for marketing because it reaches many consumers who use it to discuss products and brands within their communities. Their opinions may even impact other people's consumption behaviours and improve the customer experience. 73% of marketers think that social media marketing is effective.

Social media offers a revolutionary way for organisations to communicate with stakeholders. In the past, organisational communication was mainly one-way communication, as organisations were active generators of messages and stakeholders were passive receivers of them. With the birth of social media, organisations now have limited control over the communication process in terms of how, when and from whom a message is communicated, as well as its content.

3.2.1 Use of text analytics in social media

Monitoring social media is a common application area for text analytics. However, there are issues related to social media text analytics, such as the following:

- Time sensitivity: Along with the rapid development of the communication style on social media, the style of textual content is changing. Textual data is time-sensitive, and people's thoughts also change over time.
- Short length: Short messages often appear on social media because they allow people to communicate more efficiently, so being able to process short textual data is critical in text analytics.
- Unstructured phrases: The big difference between traditional media (such as newspapers, television and print advertisements) and social media is the quality of content. Social media users can freely structure their content as desired to express their thoughts, feelings and knowledge. For example, emojis are popular in social media to express feelings, but they are not words (Verma, Agrawal, Patel & Patel, 2016).

3.3 Branding

According to the American Marketing Association, a brand is a "name, term, sign, symbol, or design, or combination of them, intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competition" (Keller 2008, p. 2) and "is a shorthand way of communicating critical data to the market to influence decisions" (Sinha 2007, p. 139).

Branding is usually relevant in the marketing field. The process of branding involves the accumulated effects of the products and services provided, advertising promotions and outside comments (American Marketing Association, 2019).

3.4 University branding

With intense competition in the education field, many universities now perceive branding as a critical task and brand as an asset to their organisation. In higher education, a university's brand is an emergent property of the institution that differentiates it from others. The brand reflects the institute's ability to fulfil its mission, maintain trust from the general public and satisfy students' educational needs (Anctil, 2008). A university brand also helps potential recruits make enrolment decisions (Ali-Choudhury, Bennett & Savani, 2009). In other words, branding a university involves defining what the university is, what it stands for, and what it wants to be known for (Waeraas & Solbakk, 2008). The attributes of a university brand (Cubillo, Sanchez & Cervino, 2006) are shown in Table 1 as follows:

Dimension	**Attribute**
Institute	Institution prestige, rank, brand reputation, academic reputation, research reputation, quality reputation
Faculty	Expertise of teaching staff, professional experience of teaching staff
Facilities	Campus atmosphere, social life at university, safety and security, library facilities, availability of computers, availability of quiet areas, availability of areas for self-study, sport facilities

Table 1: Attributes of a university brand

3.4.1 Advantages of branding a university

Through branding, universities can differentiate themselves from their competitors in terms of sports, facilities, academic strengths and geographic location (Stephenson, Heckert & Yerger, 2016). University brands can enhance the institute's reputation as well as become a symbol of excellence (Judson, Aurand, Gorchels & Gordon, 2009) so that members associated with the university can gain a sense of superiority and belonging from the symbol (Mael &

Ashforth, 1992). It is easier to recruit students if the university has a famous brand (Moogan, Baron & Bainbridge, 2001).

3.4.2 University brands for international students

Choosing to study abroad requires prudent decision-making. In addition to tuition fees, the reputation of the destination country has a significant impact on students' choices. They prefer to make "safe" decisions regarding their major, university and country that can bring them better career opportunities. Therefore, universities should provide adequate certification for the quality of their education to attract prudent international students. Since the main reason why students choose to study abroad is that they believe the degree will give them better career opportunities, universities should provide evidence that a degree from them can lead to job offers and promotions in the respective fields. They can provide this evidence through their alumni associations (Gray, Fam & Llanes, 2013).

Understanding the key factors that students consider in selecting a university can help universities effectively set goals when building a brand. The key factors are as follows:

- Major: Generally, students are attracted to a college because it offers a major, they want to study that is not available at other schools. Besides majors, honours colleges are also attractive to students.
- Price: If the tuition exceeds students' expectations, they will consider other universities even if the expensive one was their first choice. However, prospective students with a higher socioeconomic status have a different awareness of the brand and are less sensitive to price (Narayana & Markin, 1975).
- Visits: Even if students regard other universities as their first choice, inviting them to visit the school can increase their sense of belonging and even influence them to change their mind.

- Perceptions of others: Others' positive or negative views of a university affect prospective students' emotions toward the institute. This aspect is related to the reputation of the school. Prospective students will feel comforted if others think it is a good college, and similarly, they will feel upset if others have negative comments about it.
- Size and location: These factors are related to students' preferences for their living environment. For example, students from small towns may prefer to attend to larger campuses, and students may choose a specific university because it is close to their home.
- Campus aesthetics: A beautiful campus, convenient facilities, brand-new dormitories, recreation centres and so on are all important factors for students when choosing a school.
- Friendly and comfortable campus: A friendly and pleasant campus atmosphere, including both the people and environment, not only makes students feel comfortable but also promotes students' sense of belonging (Stephenson, Heckert & Yerger, 2016).

3.5 Social media and branding

Brands on social media can set up their own communities, and these communities influence brand loyalty. According to Laroche, Habibi, Richard and Sankaranarayanan (2012), building brand communities on social media has a positive effect on brand loyalty and brand trust. These special online communities allow community markers to value creation practice (i.e., social networking, community engagement, impressions management and brand use) and enhance brand loyalty through brand-use practice.

Regarding how the brand communities based on social media affect brand loyalty, in the findings of Laroche, Habibi and Richard (2013), these communities were a positive factor in customers' relationships with the product, brand, company and other customers. The relationships affect brand trust positively, which positively affects brand loyalty. Overall, brand trust plays a mediating role in enhancing relationships in brand communities. Therefore, social media can help brands develop customer loyalty if they set up brand communities with suitable relationships between customers, with the key factor being brand trust.

3.5.1 Brand performance on social media

The rise of social media has turned consumers from listeners into participants in marketing. Consumers can present their ideas to businesses on social media, allowing businesses and consumers to work together to build a brand on social media. There are three facts related to building a brand on social media: (1) the process of building a brand is more important than the outcome, (2) managing the brand is necessary to keep it alive, and (3) understanding consumers and their roles is the key factor in ensuring successful brand performance (Singh & Sonnenburg, 2012). According to research by Rutter and Roper (2016), using social media interactively in relation to branding has a positive impact, especially if a university acquires a large number of "likes" or followers on social media.

3.6 Prior studies on university branding and social media

Table 2 provides a chronological view of prior studies about university branding on social media.

Author	Year	Contribution
Pringle, J., & Fritz, S.	2019	The study suggested that the content about universities shared on social media is largely trustworthy, albeit filled with shades of grey. Universities still lack opportunities to build and strengthen their brand and counter negative messages.
Rutter, R., Roper, S., & Lettice, F.	2016	The study suggested that using social media, especially amassing a large number of "likes" on Facebook and followers on Twitter, has a positive effect on university brand and recruitment performance.

Huynh, T., & Nguyen, N.	2015	The study suggested that the objectives and operations of Lahti UAS to promote an international image through social media were generally effective, but room for improvement still existed.
Bélanger, C. H., Bali, S., & Longden, B.	2014	The study suggested that Facebook is the preferred platform for university-initiated posts, while Twitter is a more popular platform to generate conversations about Canadian universities, and that students and third parties are becoming the dominate message generators.
Qi, B., & Mackie, L.	2014	The study suggested that there is little evidence that the use of social media to engage with students and strengthen brand awareness is effective.
Chauhan, K., & Pillai, A.	2013	The study suggested that the content type and agility on social media for Indian higher education institutes have a significant impact on the amount of student engagement.
Wijaya, B. S., & Putri, D. M.	2013	The study aimed to verify the impacts of social media on university brand image. It suggested that the roles of 'community' towards brand identity and 'connectivity' towards brand benefits were the most powerful influences. In addition, 'openness' and 'conversation' did not have significant effects on brand image.
Farshid, M., Pitt, L., Botha, E.	2011	The study suggested that the visibility of social media in South Africa university branding is not distinct, and none of the studied South African universities have an unequivocal strategy for engaging with consumers on a particular social media platform.
Constantinides, E., & Zinck Stagno, M. C.	2011	The study identified the behaviour of prospective students on social media and the role of social media in their choice of higher education programmes and institutions in the Netherlands.

Mattson, E., & Barnes, N. G.	2009	The study showed the current situation of social media in university branding. It also suggested that most university admission departments consider social media important to their future strategy.

Table 2: Prior studies about university branding on social media

3.7 Prior studies on Chinese university branding

Because of region and language limitations, there has only been one international study on Chinese university branding. In the study, the author defined 'university spirit' and 'university brand' under the background of Chinese higher education and also suggested the conditions under which to realise university brand effects in the marketing economy in China (Zun, 2002). This suggests the need for a study on Chinese university branding.

3.8 A conceptual framework for study

We have developed a conceptual framework (Figure 8) based on our understanding of the key concepts as well as our comprehensive literature review. The framework serves as a mental guideline for our investigation. It demonstrates how text analytics can be applied to the study of university brand communication on social media platforms such as Twitter.

As discussed earlier, branding is about delivering critical data in order to influence decision making. University brands consist of general perceptions about the institute, faculty and facilities.

Since social media provides a platform for two-way communication between the stakeholders of an organisation, it is a powerful tool for universities to use when building their brand. Social media participants are the brand ambassadors and storytellers of universities. Brand stories that are spread through social media are more influential because they are digital, visible and dynamic. Through social media, universities can listen directly and authentically to the participants, such as students, for brand assessment and management.

However, social media data are unstructured and extensive, as well as time-sensitive and full of noise. Text analytics is a powerful tool for managing unconventional data. There are several methods used in text analytics such as mining, extracting, question answering and sentiment analysis. As a result, universities can convert messy textual social media data into insights about their online brands. Universities can also develop more specific branding strategies for internationalisation and global promotion.

The feedback loop of the text analytics system is also shown in figure 8. The university branding process can be used to evaluate the effectiveness of the applied text analytics approach, as well as to build and engage loyal communities on social media.

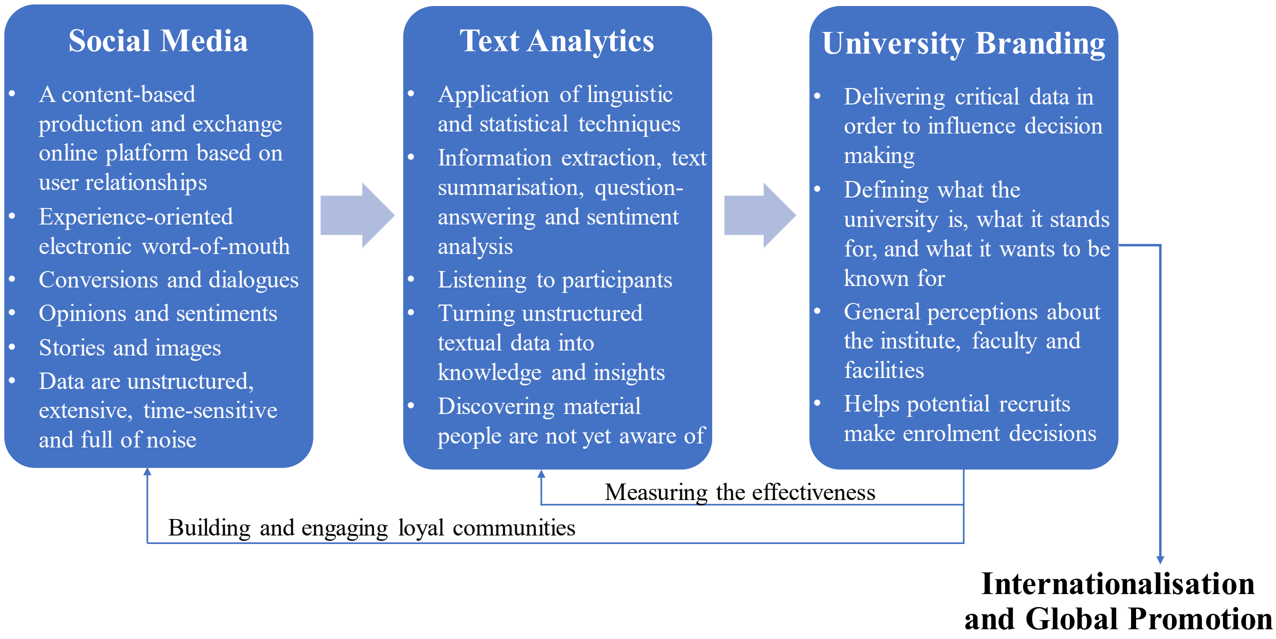


Figure 8: Using text analytics to assess university branding on social media

Chapter 4: Methodology

As discussed in the previous chapters, our project applies text analytics to explore Chinese university branding on social media. Our study targets are the C9 League universities with a focus on Tsinghua University and Peking University. The text analytics methodology we have adopted to process social media data is shown in Figure 9. The approach consists of three main stages: data collection, data sanitisation and data analysis.

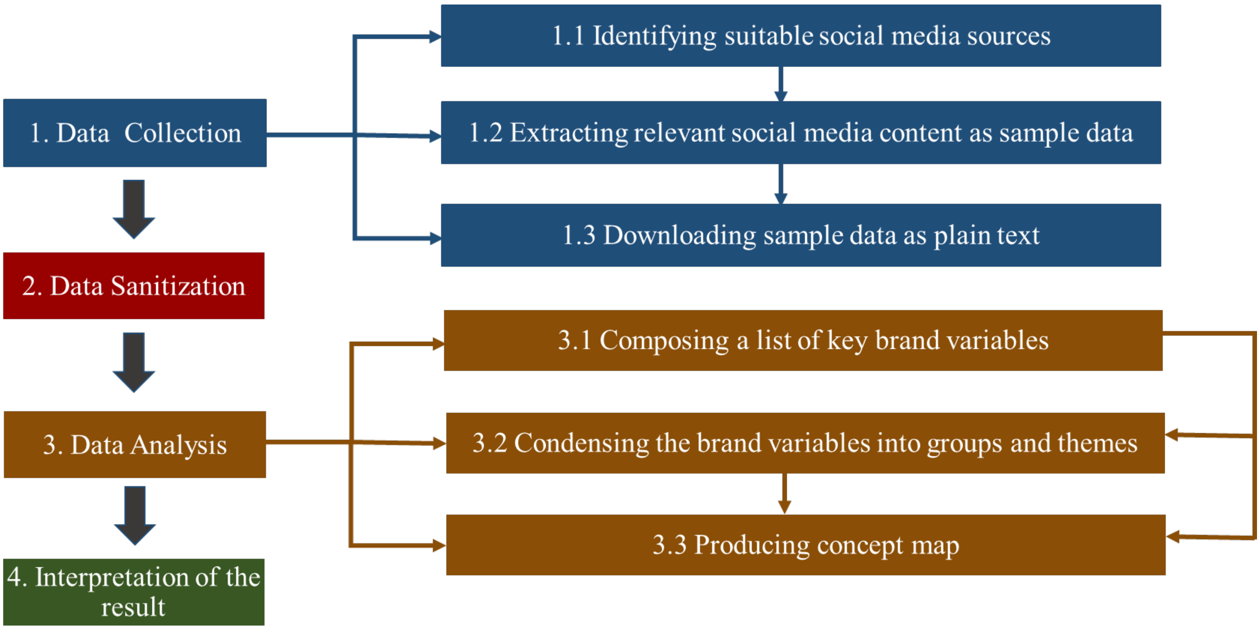


Figure 9: The adopted text analytics methodology

4.1 Data collection

Data collection is the first stage of the methodology, and it consists of three steps.

4.1.1 Identifying suitable social media sources

A relevant data source is essential for an investigation of Chinese university branding on social media. The target social media must be international, representative and accessible. In terms of being international, the language used in the target data sources should be English, the international language, and the users should be from all over the world. In terms of being representative, the target social media should have a large number of users and a broad userbase. In terms of being accessible, the target social media should provide access to content such as reviews, comments, discussions, opinions and experiences concerning the C9 League Chinese universities. Based on these three criteria, we have set Twitter as our target data source.

4.1.2 Extracting relevant social media content as sample data

Because Twitter protects its data, we first applied for Twitter application programming interface (API) keys to get access to Tweet data. Then we developed a web crawler (see Chapter 5) to help us find useful content using keyword searches.

4.1.3 Downloading sample data as plain text

All the extracted Tweets, including the username, date and content, are automatically downloaded as plain text. We are also manually double-check the relevance of the downloaded Tweets. The screened Tweets are saved in a master file as our sample data.

4.2 Data sanitisation

Data sanitisation is the procedure of finding and correcting data. Since the words people use on Twitter are informal, data sanitisation ensures that the text used for analysis is correct and has a clear meaning to obtain effective results. The means of data cleaning are based on the source and content of the data. There are some steps that are typically performed during data cleaning:

- Ensuring that all action or descriptive terms (e.g., ‘good’ and ‘like’) are used in a positive context without negation (e.g., ‘no good’ and ‘don’t like’).
- Changing multi-word phrases into one-word formats (e.g., ‘Hong Kong’ to ‘HongKong’).

- Turning plural forms into singular forms (e.g., ‘students’ to ‘student’).
- Using consistent spelling for all words in the sample data (e.g., use ‘program’ instead of ‘programme’).

4.3 Data analysis

After two data preparation stages, we continue with the data analysis process. Analytic tools are required at this stage. There are four steps to data analysis.

4.3.1 Composing a list of key brand variables

We compose a list of key brand variables using the WordSmith software program. Key brand variables are words that appear at an unusually high frequency in the sampled Tweets. To determine especially high frequencies of appearance for each word, we refer to the British National Corpus, which lists 100 million English words and their normal frequency in language use.

4.3.2 Condensing the brand variables into groups and themes

Based on the brand variable list, we group the variables into factors and themes and carry out a factor analysis using SPSS in this task. WordSmith software help us obtain the count frequencies of all identified variables in each sample text. The resulting frequency table is transferred into the SPSS database for the factor analysis.

4.3.3 Producing concept maps

In the last step, we produce concept maps of the C9 League universities using Leximancer. After inputting our collected Tweets into Leximancer, the software automatically generates high-level concepts and their relationship based on semantic analysis. The resulting concept maps provide a holistic view of the university brand make-up.

4.4 Interpretation of the results

After the data analysis process, we gain insights into the Chinese university branding on social media. We interpret the analysis results with reference to our study questions: (a) What are the brand identities of Chinese universities on social media? (b) What are the similarities and differences of Tsinghua University and Peking University? (c) How do the brands of Chinese universities compare with those of international elite universities?

4.5 Three levels of computer analysis

In our project, we use three software tools to conduct the text analysis: WordSmith, SPSS and Leximancer.

4.5.1 WordSmith

WordSmith is a software package primarily used for lexical analysis. The software tool has a collection of modules for searching for patterns in a collected text. The core areas of the software package include three modules (see Figure 10):

- Concord produces a concordance from plain files. To use Concord, one specifies a search word or phrase for Concord to locate in all selected text files. A concordance allows us to see many examples of a word or phrase in context.
- WordList lists all the words and word forms that are included in the selected corpus and the statistical data that are different from the text corpus. By using this module, we can learn which words or word forms occur frequently or rarely.
- Keyword creates a list of all words and word forms whose frequency is unusually high. The software compares two pre-existing word lists: one list generated from the sample

data and one large word list that acts as a reference file, like the British National Corpus.

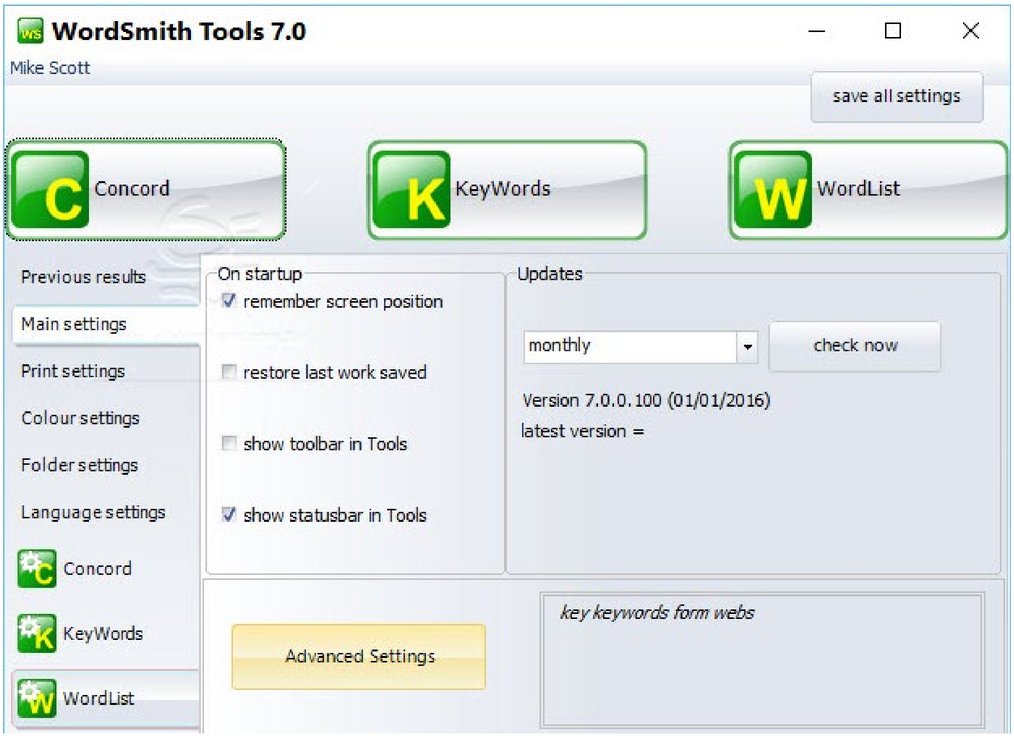


Figure 10: The three modules of WordSmith

4.5.2 SPSS factor analysis

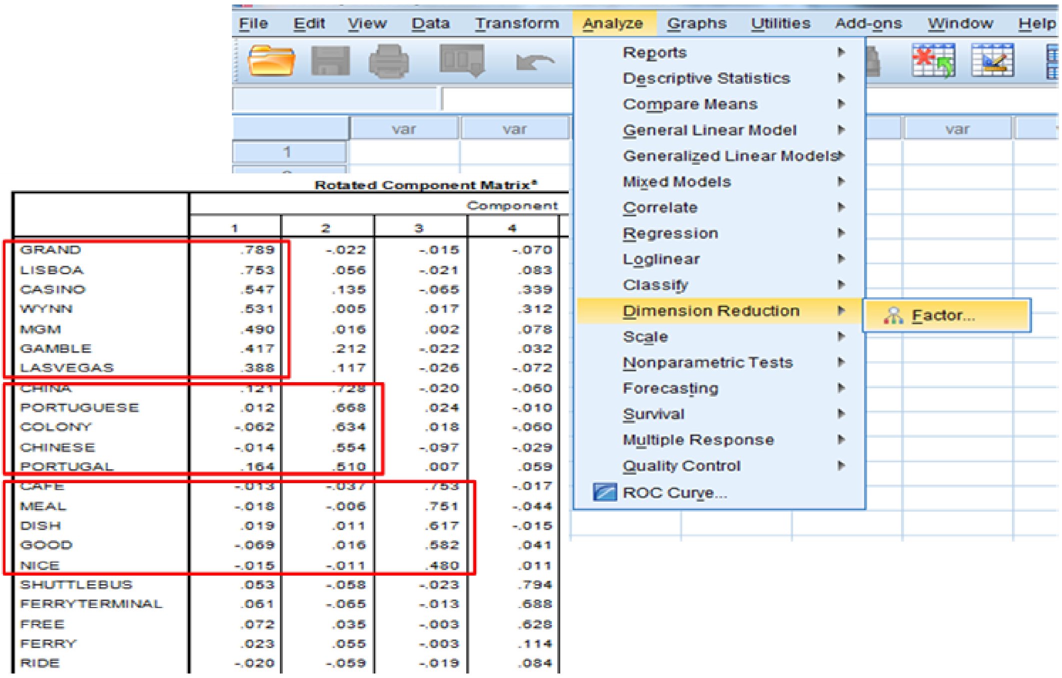
Statistical Package for the Social Sciences, often abbreviated as SPSS, is a software program widely used for statistical analysis in social science. The program is also used when dealing with big data. It provides advanced statistical analysis and a vast library of algorithms.

In this study, we use SPSS to perform factor analysis (Figure 11). Factor analysis is mainly used to summarise related variables into factors for data deduction. To conduct factor analysis, one must acquire a large sample size because the analysis relies on the correlation matrix of the variables involved, and a large sample size guarantees the accuracy or stability of the correlations. Factor analysis involves the following three stages:

- Computing the correlation matrix of all variables; SPSS generates a correlation matrix for all variables and eliminates those that are unrelated to other variables.
- Extracting factors. Initial factors are estimated through a principal component analysis,

which is the most commonly used extraction method.

- Rotating factors. The most popular rotational method is the varimax rotation, which yields uncorrelated factors. This process enhances the interpretability of the factors.



Rotated Component Matrix[a]

	Component			
	1	2	3	4
GRAND	.789	-.022	-.015	-.070
LISBOA	.753	.056	-.021	.083
CASINO	.547	.135	-.065	.339
WYNN	.531	.005	.017	.312
MGM	.490	.016	.002	.078
GAMBLE	.417	.212	-.022	.032
LASVEGAS	.388	.117	-.026	-.072
CHINA	.121	.728	-.020	-.060
PORTUGUESE	.012	.668	.024	-.010
COLONY	-.062	.634	.018	-.060
CHINESE	-.014	.554	-.097	-.029
PORTUGAL	.164	.510	.007	.059
CAFE	-.013	-.037	.753	-.017
MEAL	-.018	-.006	.751	-.044
DISH	.019	.011	.617	-.015
GOOD	-.069	.016	.582	.041
NICE	-.015	-.011	.480	.011
SHUTTLEBUS	.053	-.058	-.023	.794
FERRYTERMINAL	.061	-.065	-.013	.688
FREE	.072	.035	-.003	.628
FERRY	.023	.055	-.003	.114
RIDE	-.020	-.059	-.019	.084

Figure 11: SPSS factor analysis

4.5.3 Leximancer

Leximancer is a software tool that is designed to analyse textual data in natural languages. It can extract semantic meaning and relational information, as well as produce outputs that include concept maps and concept thesauruses. To produce a concept map, this tool computes the frequency at which each word is used and then calculates the distance between each term. The algorithms used are statistical, but they employ non-linear dynamics and machine learning. Concepts are semantically contextualised on a Leximancer concept map (Figure 12). The semantic network (Figure 12) is a heat map. The degree of importance of the various concepts are represented by different colour clusters. The colours range from warm to cool and denote importance from strong to weak, respectively.

Figure 12: A Leximancer concept map

Using Leximancer involves the following four main processes (see Figure 13):

- Selecting the documents. We load the textual data collected from Twitter.
- Generating concept seeds. Seed words in our sampled text are identified as potential concepts.
- Generating a thesaurus. The settings for the concept seeds can be edited. Leximancer generates a list of weighted terms (a thesaurus) for the concepts.
- Generating a concept map. The settings for the concept coding, compound concepts and project outputs can be edited. Leximancer generates concepts maps (see Section 4.3.4) that approximate human perceptions.

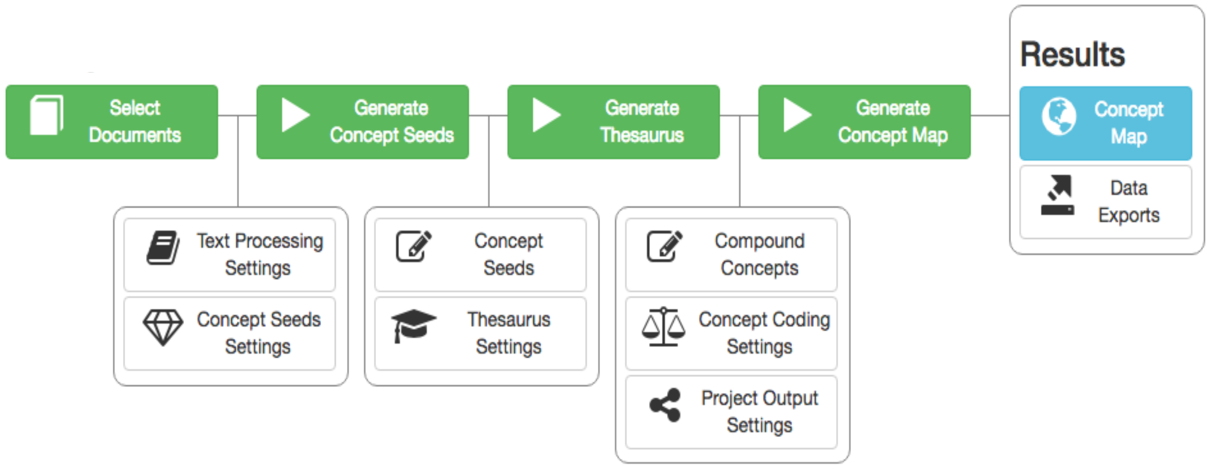


Figure 13: Functions of Leximancer

Chapter 5: Software Tool Development

5.1 Web crawlers

A web crawler, also known as spider or spider-bot (Wikipedia, 2019), systematically downloads and indexes publicly available webpages so that users can recover them when necessary. This kind of software is widely used in search engines. Web crawlers collect data by crawling the web and producing a list of hyperlinks relevant to the user's requirements. To build a high-quality search engine and provide efficient results for the user, the index criteria are generally based on the number of other pages connected to the target page, the amount of traffic the page attracts, etc (Cloudflare, 2019). Crawlers expand the target website's resources when they visit the site; therefore, web crawlers need to ensure the schedule and load to not overburden the target site and increase the cost of broadband, which may affect the user experience.

5.1.1 Crawling and scraping

Web crawling typically refers to collecting data from world wide web. Crawlers go through many different websites and download information on the web such as text and metadata. Web scraping refers to locating data and downloading them from a website. Web crawling and scraping always go together. When web crawling, a crawler goes through various websites like a spider crawling around its web. Once the crawler arrives at the target, it filters the information it needs and extracts the data using web scraping (ProWebScraping, 2016). Figure 14 illustrates the similarities and differences of web crawling and web scraping.

Our project aims to study online brand communication of Chinese universities by analysing social media content. We have developed a software tool to crawl the web and scrape relevant Tweets (Twitter posts) based on keywords such as school name. The scrapped Tweets were then downloaded as sample social media data for text analytics.

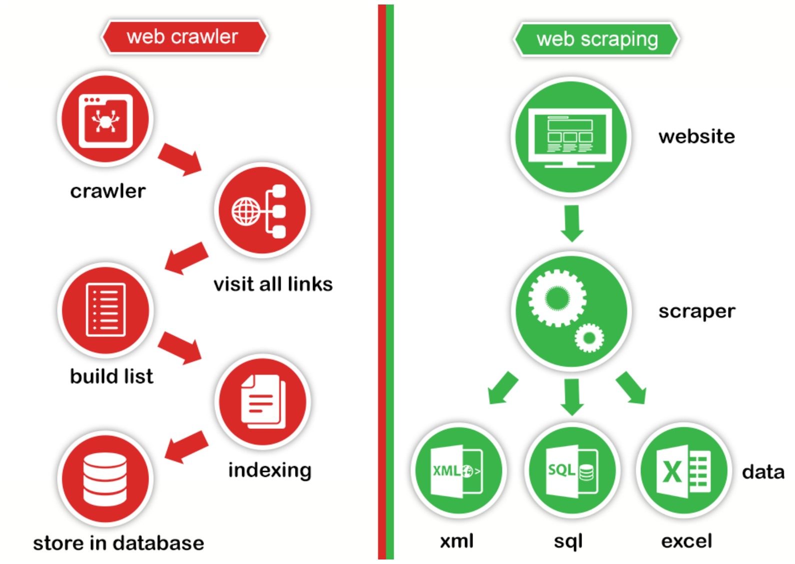


Figure 14: Web crawling and web scraping
(Source: http://prowebscraping.com/web-scraping-vs-web-crawling/)

5.1.2 Web crawling steps

Figure 15 illustrates the steps of web crawling. When the spider crawls the pages on a website, it usually starts at the homepage. It then follows the links on the homepage to navigate to the subsequent pages in the second and third tiers. While crawling, the spider records certain parameters on each page. It draws up a list of words and notes where they appear. In the next stage, the search engine builds an index of the page based on its own weighting system for each parameter. Once it is finished building the index, the search engine encodes the data, mainly to save server space and for privacy. The route is completed when the data are stored in somewhere that can be accessed by the search engine to retrieve information that will show up in the search results in response to a search query by the user.

Since the Internet changes and updates rapidly, there are a lot of websites that any given crawler does not know. Therefore, the spider starts web crawling from a list of known URLs. When the spider crawls those websites, it finds hyperlinks that link to other websites and then clicks to visit those new websites. If the spider arrives at a new website that it has never visited before, it appends the URL to its list. This process can be looped infinitely, so developers should consider what websites they want to crawl as well as the crawling schedule and update their database regularly so that they can obtain efficient information and reduce duplicate data.

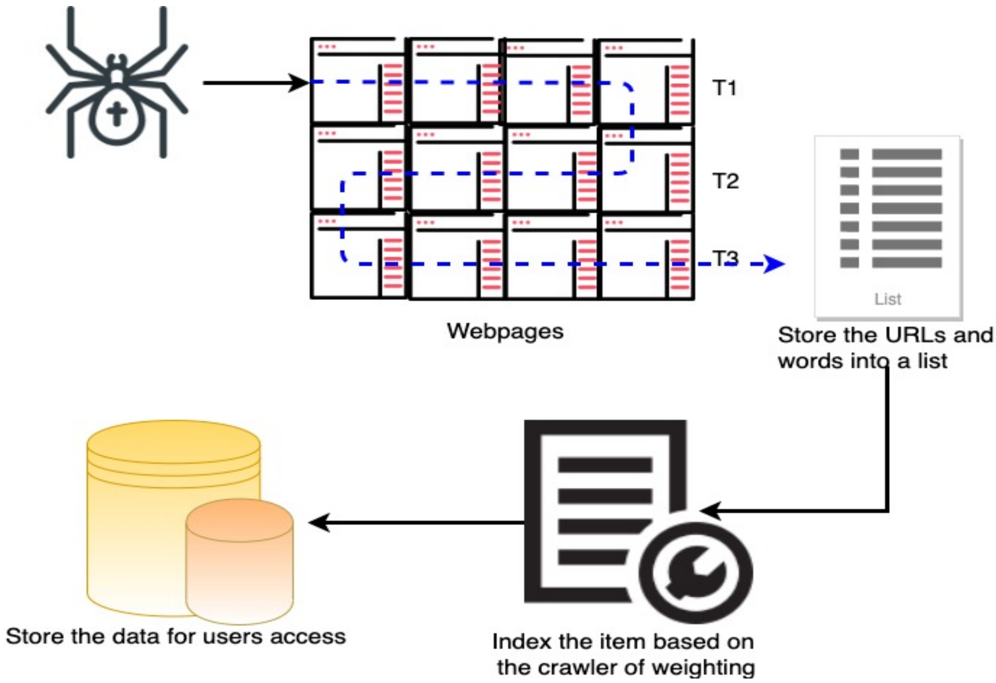


Figure 15: Web crawling steps

5.2 Python

Our web crawler program is written in Python, an interpreted, high-level, general-purpose programming language. It is compatible with multiple programming languages, such as Java and C++ (The Python Software Foundation, 2019). Python uses English keywords instead of symbols, which greatly improves the readability of the code and makes it easier to learn. Python Package Index (PyPI) is a repository that allows developers to find and install Python software (packages) shared in the Python community by others. These packages contain functions developed by the authors that developers can use directly after installing the package instead of developing all the functions themselves, which improves development efficiency and keeps code cleaner. We can use these packages to help us easily and systematically develop a crawler software.

5.3 Preparation for the program development

In preparation for the program development, we downloaded the required Python packages and related tools. We also obtained data access permission from Twitter. The Twitter authorization was carried out keys and tokens from the Twitter application programming interface (API).

5.3.1 Python installation and setup

Python 3.0 is the latest version of the Python programming language. We first downloaded the appropriate Python installer from the official Python website. Once the installer was downloaded, we ran it. After that, Python was installed on our computer (see Figure 16).

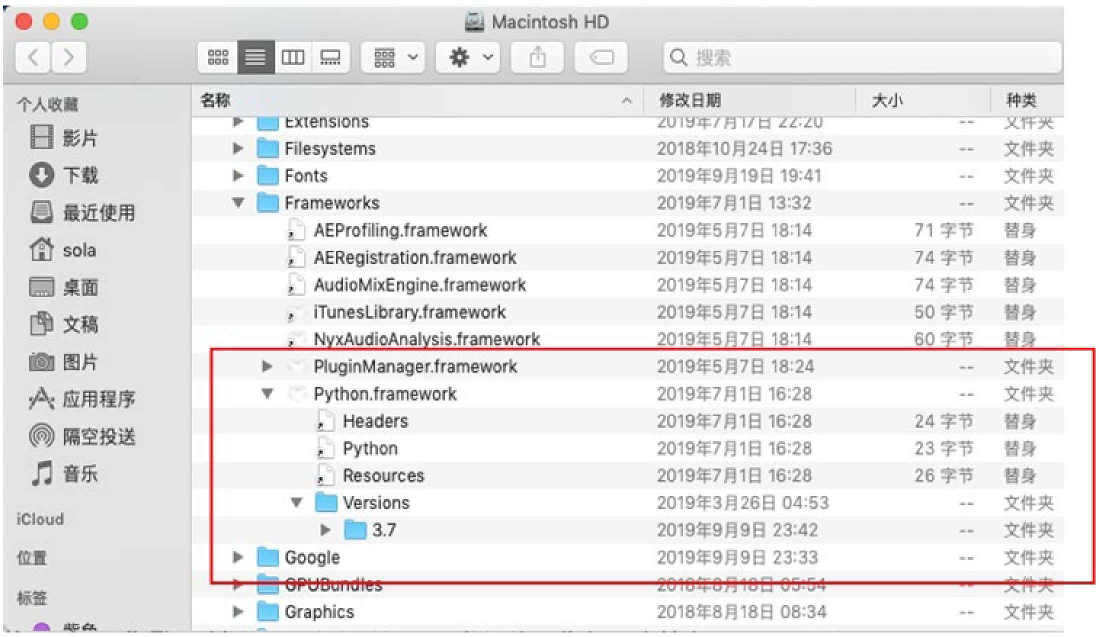


Figure 16: Python elements

As shown in Figure 16, the installed elements include the following:

a) IDLE: an integrated developing environment written in Python with the Tkinter GUI package. It provides a programming and testing environment for Python developers.

b) Standard library: a collection of packages we can use.

5.3.2 Installing the package manager

So that we could more easily download the packages we need, the next step of the setup was to download a package manager. We downloaded a third-party package tool by calling the package manager in the terminal. Pip, which includes a large package index, is a popular manager for developers. We just typed the command into the terminal to install the package manager (Figure 17).

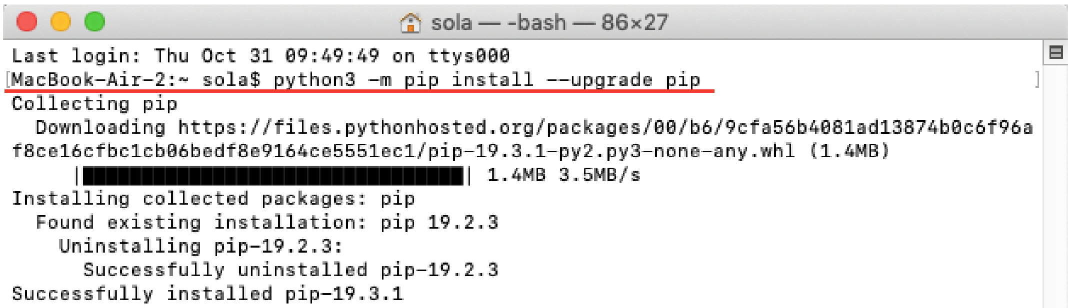


Figure 17: Pip, a package manager

5.3.3 Twitter API

To scrape Twitter content, we had to register for a developer account on the Twitter platform. After that, we received a collection of API keys after receiving permission to access Twitter data. In the process of applying for the API, we needed to elaborate on how we would use the Twitter data.

5.3.4 Installing Tweepy

Tweepy is a Python program developed with Twitter. It helps download Tweets from Twitter with queries. Beside accessing Twitter data, Tweepy has many functions, such as allowing developers to manage their own Twitter accounts by posting Tweets using Python.

5.4 Web crawler program development

There were three major stages of our web crawler development: 1) gaining access to Twitter data; 2) coding the web crawling function; and 3) using regular expressions to pre-process the data.

5.4.1 Gaining access to Twitter data

We applied the set of API keys and tokens provided by Twitter to gain access to Twitter data. Figure 18 illustrates the authorization process.

```
consumer_key = '[redacted]i'
consumer_secret = 'rR[redacted]xqK'
access_token = '118[redacted]MK'
access_token_secret = 'Fbi[redacted]aF'

auth = tweepy.OAuthHandler(consumer_key,consumer_secret)
auth.set_access_token(access_token,access_token_secret)

api = tweepy.API(auth)
```

Figure 18: Twitter API keys and tokens

5.4.2 Coding the Web-crawling function

We then coded the web-crawling function (see Figure 19), which was broken down into five steps:

1) Setting the search criteria. The extraction of Tweets is based on search keywords. Other query criteria, such as the number, language, or content of Tweets, can also be specified.
2) Creating lists to store the search results (data)
3) Extracting the full content of all messages, including re-Tweets and forwarded Tweets, and storing the data in ‘rlist’.
4) Keeping only one copy of Tweets with duplicated content and storing the data in ‘resultlist’.
5) Displaying the results.

```
def getTweet(self,search_key):
    search_requirement = api.search(q=search_key,
                        result_type = 'recent',
                        count = 10,
                        lang = 'en',
                        tweet_mode="extended",
                        since = "2019-10-25")
    ulist = []
    rlist = []
    dlist = []
    resultlist = []

    for tweet in search_requirement:
        if 'retweeted_status' in tweet._json:
            screen_name = tweet._json['user']['screen_name']
            content = tweet._json['retweeted_status']['full_text']
            date = tweet._json['created_at']
            content1 = strip_all_entities(strip_links(content))
        else:
            screen_name = tweet._json['user']['screen_name']
            content = tweet.full_text
            date = tweet._json['created_at']
            content1 = strip_all_entities(strip_links(Remove_Emoji(content)))

        rlist.append(content1)

    for i in rlist:
        if not i in resultlist:
            resultlist.append(i)
            dlist.append(date)
            ulist.append(screen_name)

    fin_result = ''
    count_resultlist = len(resultlist)
    for r in range(0, count_resultlist):
        result = 'User: '+ulist[r]+'\n'*1+'Date: '+dlist[r]+'\n'*1+'Content: '+'\n'*1+resultlist[r]+'\n'*2
        fin_result += result
```

Figure 19: Web-crawling source code

5.4.3 Data pre-processing

We developed some functions using regular expressions for data pre-processing (Figure 20). A regular expression is a text string used to describe a search pattern. It is considered a special kind of wildcard.

We eliminated hyperlinks, emoji, and some meaningless symbols, such as ‘@’ and ‘#’. People use ‘@’ to tag other users, so we also eliminated the username after ‘@’. Similarly, users add hashtags to their Tweets to categorize them, so we removed the ‘#’ symbol but retained the words within the hashtag.

```python
def strip_links(text):
    link_regex    = re.compile('((https?):((//)|(\\\\))+([\w\d:#@%/;$()~_?\+-=\\\.&](#!)?)*)', re.DOTALL)
    links         = re.findall(link_regex, text)
    for link in links:
        text = text.replace(link[0], ', ')
    return text

def strip_all_entities(text):
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t])|(\w+:\/\/\S+)"," ",text).split())

def Remove_Emoji(text):
    emoji_pattern = re.compile("["
          u"\U0001F600-\U0001F64F"  # emoticons
          u"\U0001F300-\U0001F5FF"  # symbols & pictographs
          u"\U0001F680-\U0001F6FF"  # transport & map symbols
          u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
          u"\U00002702-\U000027B0"
          u"\U000024C2-\U0001F251"
          "]+", flags=re.UNICODE)
    new_text = emoji_pattern.sub(r'',text)
    return new_text
```

Figure 20: Data pre-processing

5.5 Graphic user interface development

We used PyQt5 to develop a graphical user interface (GUI) for the web crawler as shown in Figure 21:

```python
def initUI(self):
#create a GUI window
    font = QtGui.QFont()
    font.setFamily("Calibri")                                        1
    font.setPointSize(12)

    search_label = QLabel("Please input a school name: ")
    search_item = QLineEdit()                                        2
    btn1 = QPushButton("Search", self)
    search_result = QTextEdit()

    search_label.setFont(font)
    btn1.setFont(font)                                               3
    search_result.setFont(font)

    grid = QGridLayout()
    grid.setSpacing(5)
    grid.addWidget(search_label, 1, 0)
    grid.addWidget(search_item,2, 0)                                 4
    grid.addWidget(btn1, 3, 0)
    grid.addWidget(search_result, 4, 0, 5, 0)
    self.setLayout(grid)

    #setting the search button
    def search():
        search_result.setText(self.crawfin(search_item.text()))
    btn1.clicked.connect(search)                                     5

    #setting the window
    self.setGeometry(450, 250, 1080, 607)
    self.setWindowTitle("Tweets")
```

Figure 21: GUI source code

Qt is an open-source widget for creating GUIs that can run on various systems. PyQt5 enables the use of the Qt GUI structure from Python. As shown in Figure 21, we compiled five sets of code for the GUI development:

1) Setting the font style.
2) Adding interface components.
3) Applying the font style to the components.
4) Setting the configuration of the components in the frame.
5) Setting the function of buttons to execute certain tasks when the user clicks them.

5.6 The Web crawler in use

Figure 22 shows the input and output sections of our web crawler program.

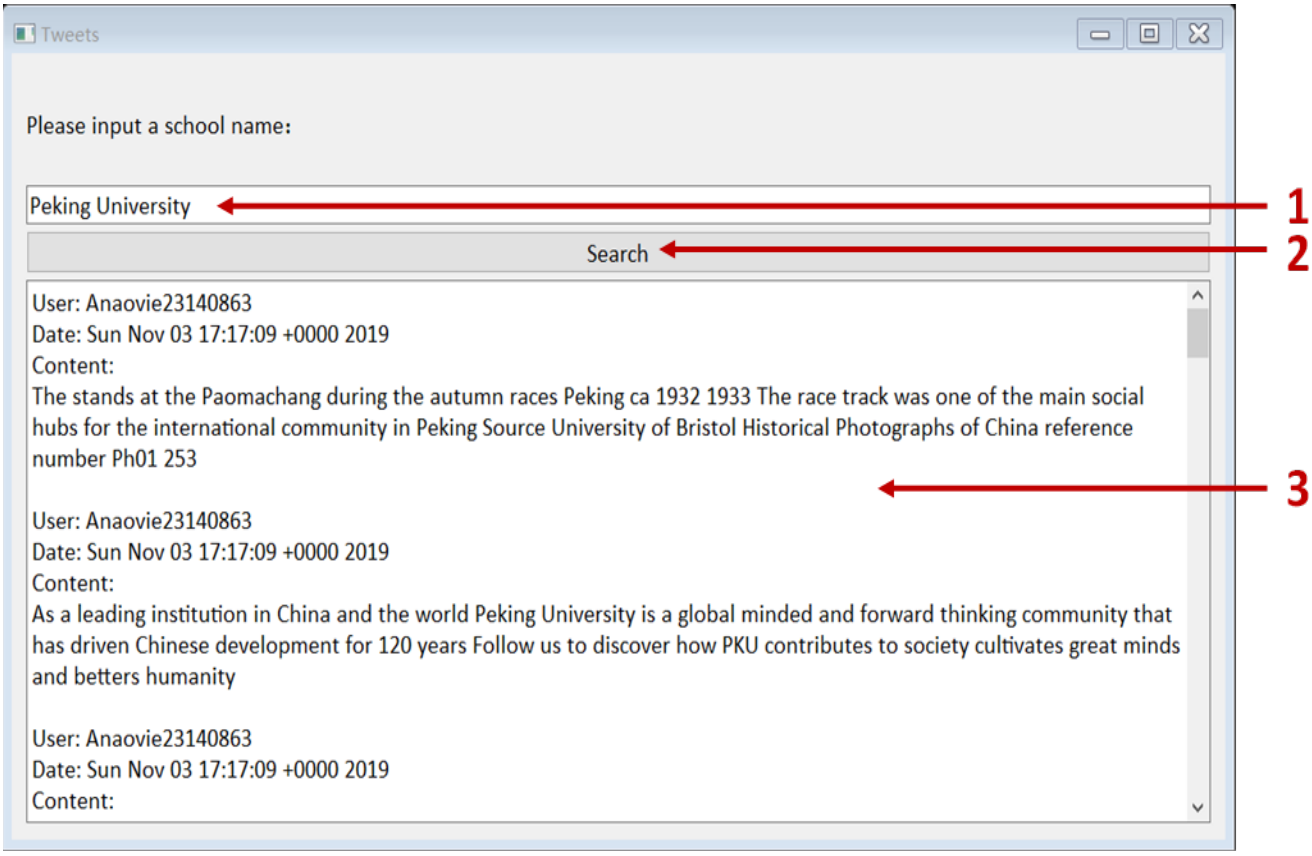


Figure 22: The web crawler user interface

As shown in Figure 22, there are three steps to using the program:

1) The user inputs the keyword for Tweet extraction.

2) The user clicks the ‘search’ button for confirmation.

3) The software displays the extracted Tweets.

Chapter 6: Results and Findings

In this chapter, we follow the steps of the methodology shown in Figure 9 to accomplish our text analytics about exploring the effects of social media communication on university branding. The results and findings from the data collection, data sanitization, and data analysis are given below.

6.1 Data collection

This process consists of three stages: selection of target website, extraction of relevant comments, and descriptive data analysis.

6.1.1 Target website

As described in Chapter 2, Twitter is a social media platform that is popular worldwide in which schools exhibit high engagement and participation. It perfectly meets the 'international, representative, and accessible' standard for our research. Users can easily get information about the college they like and communicate with others, and the language used in Twitter is majorly English. Given that Chinese university data collected on this platform are convincing, we decided to collect relevant data on C9 League Chinese universities on Twitter.

6.1.2 Extracting relevant content

After choosing Twitter as our target social media source, we started extracting relevant content on the top seven C9 League Chinese universities. According to Figure 2 in Chapter 2, the last two Chinese universities, Harbin Institute of Technology and Xi'an Jiao Tong University, are part of the top 200 world-ranking universities from 2012 to 2020. We decided on the top seven C9 League Chinese universities as the representative universities for our text analytics and extracted relevant content on them.

We downloaded the content mentioned regarding the seven universities from Twitter as well as detailed information about each tweet – date, number of likes, number of comments, number of retweets, number of images, and whether it included a video.

6.1.3 Descriptive data analysis

After extracting relevant content on the top seven C9 League Chinese universities, we performed descriptive analytics on the sample data (1,040 tweets). As shown in Figure 23, the universities, ordered according to the number of samples (from most to least), are as follows: Peking University (PKU), Tsinghua University (THU), Fudan University, Nanjing University, Zhejiang University, Shanghai Jiao Tong University, and the University of Science and Technology of China.

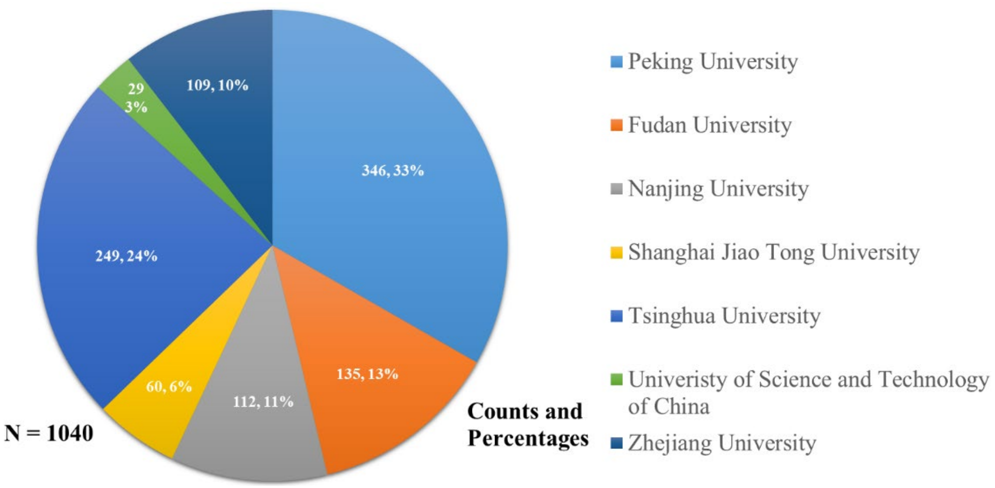


Figure 23: Data distribution based on sample universities

We collected tweets which were posted from 2015 to 17 January 2020 (see Figure 24). Since 2015, the number of tweets increased yearly, meaning that more people talked about China universities during that period.

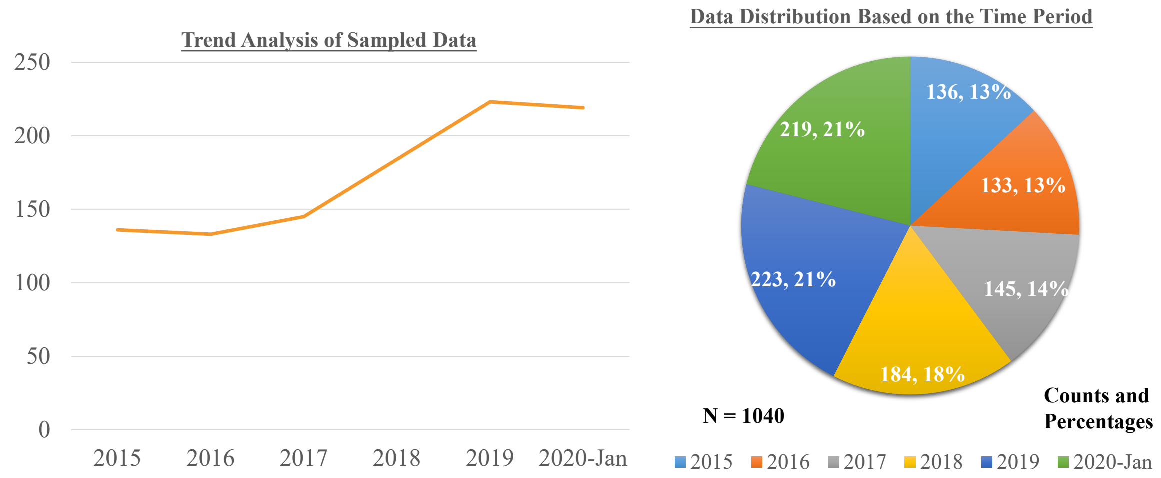


Figure 24: Distribution and trend of sample data from 2015 to 17 January 2020

Figure 25 shows the distribution of tweets according to whether they included images and videos. Most of the tweets we had collected did not include any pictures or videos, and more than half of them (63%) were in plain text. Only 27% of the tweets included videos, and 17% of the tweets had two or more pictures.

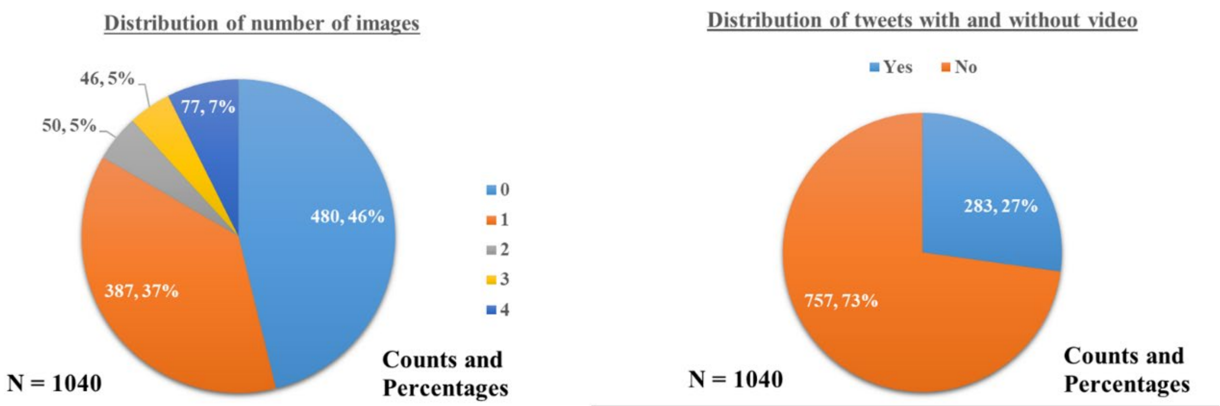


Figure 25: Data distribution based on the number of images and videos

After analysing the distribution of samples, we explored the engagement of the universities on Twitter. We focused on the number of likes, retweets, or comments of the tweets of these universities. Figures 26 and 27 show the average rate of engagement of each tweet. THU and PKU usually got a huge amount of likes and retweets per tweet, which represents a high rate of engagement on Twitter. Although the University of Science and Technology of China had the lowest average amount of likes and retweets for each tweet, it had a higher average amount of comments than Zhejiang University, Nanjing University, and Shanghai Jiao Tong University.

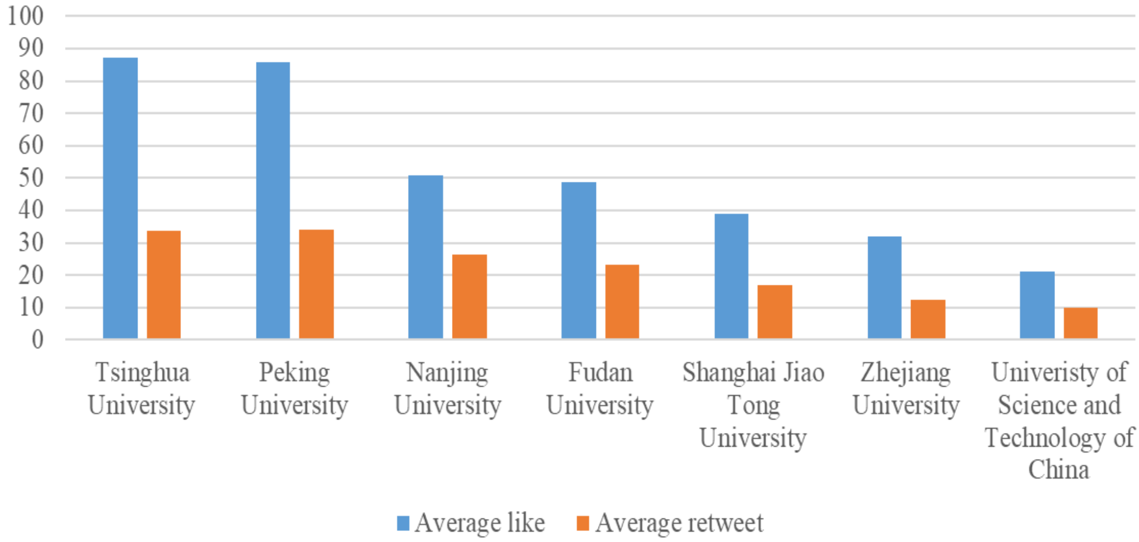


Figure 26: Average number of likes and retweets per tweet of sample universities

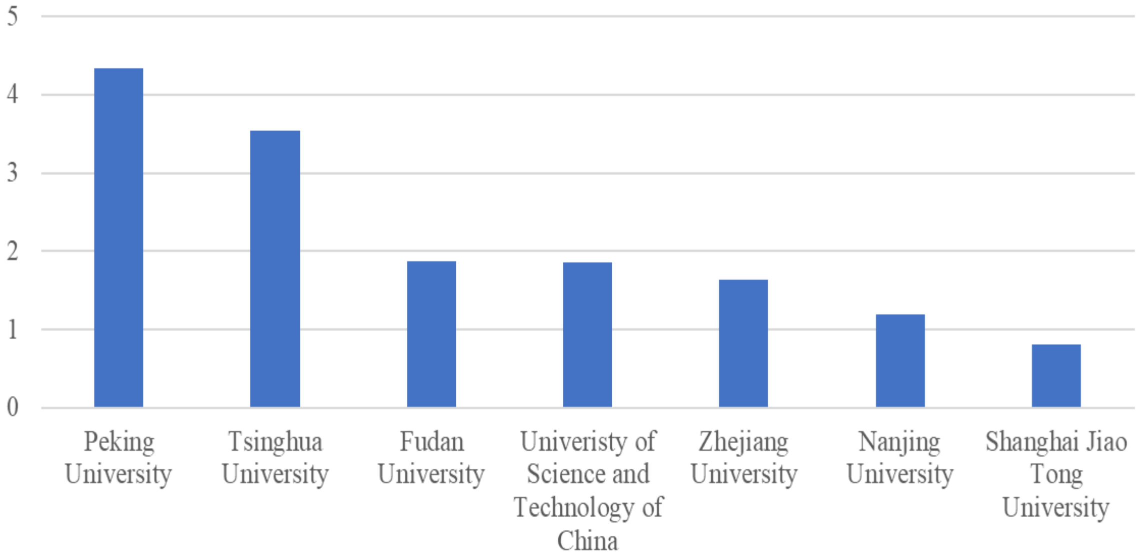


Figure 27: Average number of comments per tweet of sample universities

Some tweets did not receive likes, retweets, or comments, representing zero engagement. In Figure 28, almost half of the sample tweets did not receive comments. The Twitter users' number of likes was greater than that of their retweets, showing that they seemed to prefer liking and retweeting tweets but rarely wrote comments.

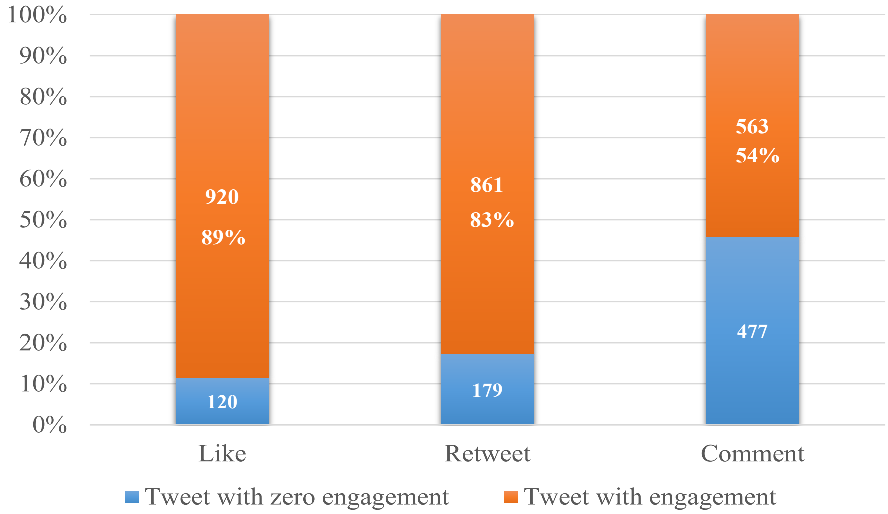


Figure 28: Tweets with and without engagement

6.1.4 Correlation analysis

Correlation analysis is a statistics method used to evaluate the relationship between two quantitative variables as well as the strength of this relationship. Different correlation coefficients present their own range of usability and characteristics (Wikipedia, 2020). We used the Pearson correlation measure to examine how social media elements are correlated with one another in enhancing the viral effects of university brand communication. We applied this analysis on the five most relevant social media elements of Twitter: comments, retweets, likes, images, and videos (see Table 3).

	Comment	Retweet	Like	Image	Video
Comment	1				
Retweet	.572**	1			
Like	.631**	.972**	1		
Image	−.006	.047	.044	1	
Video	.092**	.024	.027	−.093**	1

** — Correlation is significant at the 0.01 level (two-tailed).

Table 3: Correlation Analysis of Social Media Elements

6.2 Data sanitisation

In this stage, we performed the following tasks:

1. Changing multi-word phrases into single words (e.g. 'Beijing University' to 'BeijingUniversity')
2. Turning plural forms into singular forms (e.g. 'students' to 'student')
3. Spelling out abbreviations to maintain consistency (e.g. 'prof' to 'professor')

When we came across an elusive word, we confirmed its meaning through its relationship with the content and by searching for its definition.

6.3 Data analysis

After cleaning the content data, we followed the analytics methodology (Figure 9) to conduct data analysis using Wordsmith, SPSS, and Leximancer to explore the university brands of the seven C9 League universities.

6.3.1 Composing a list of key brand variables

First, we used Wordsmith to generate a wordlist (a vocabulary list used in data content) from the 1,040 tweets we had collected. Then we used the wordlist to generate a keyword list by referencing the British National Corpus (BNC). These keywords are words which have high relative (to BNC) frequency. In other words, the keyword list contained keywords that users often used when discussing the seven universities on Twitter. They provided a very useful method to feature the content of the collected tweets. Thus, a keyword list can be regarded as a list of key brand variables for the seven C9 League universities.

As shown in Table 4, 95 keywords were generated as the key brand variables. The word with the highest keyness was 'China' as the universities represented China. Most of the key brand variables were positive, such as 'amazing', 'great', and 'welcome'. In addition, a lot of people discussed university rankings, scholarships, advanced technologies, and degrees.

N	Keyword	Freq.	%	Keyness
1	CHINA	381	1.59	3,595.77
2	CHINESE	131	0.55	995.52
3	STUDENT	146	0.61	988.37
4	PROFESSOR	104	0.44	707.83
5	UNIVERSITY	124	0.52	605.26
6	BLOCKCHAIN	32	0.13	514.99
7	CAMPUS	49	0.21	447.41
8	SCHOLARSHIP	48	0.20	402.10
9	CENTER	35	0.15	306.29
10	INSTITUTE	50	0.21	256.47
11	AI	28	0.12	235.67
12	SCHOLAR	28	0.12	216.26
13	RESEARCH	74	0.31	208.20
14	CONGRATULATION	18	0.08	201.84
15	HONOR	18	0.08	187.38
16	SCIENCE	50	0.21	184.93
17	RESEARCHER	26	0.11	178.13
18	PROGRAM	36	0.15	171.60
19	STUDY	61	0.26	169.24
20	GLOBAL	34	0.14	166.04
21	BIGDATA	11	0.05	164.94
22	TECHNOLOGY	48	0.20	163.59
23	RANKING	20	0.08	149.67
24	QS	12	0.05	144.52
25	SCHOOL	67	0.28	138.76
26	PHD	17	0.07	133.56
27	INTERNATIONAL	52	0.22	125.75
28	LABORATORY	25	0.10	116.06

29	BRICS	8	0.03	114.93
30	METOO	8	0.03	114.93
31	CPEC	8	0.03	114.93
32	ACADEMIC	29	0.12	112.88
33	PARTNERSHIP	26	0.11	110.52
34	MOU	9	0.04	109.68
35	FINTECH	8	0.03	108.66
36	MBA	13	0.05	104.55
37	ASIA	24	0.10	103.42
38	TOP	50	0.21	103.00
39	ALUMNI	12	0.05	96.44
40	MASTER	28	0.12	94.65
41	INNOVATION	19	0.08	90.41
42	FACULTY	17	0.07	85.55
43	VISITING	21	0.09	83.51
44	ACADEMY	17	0.07	83.27
45	HOST	21	0.09	82.12
46	ROBOT	13	0.05	81.19
47	PRESTIGIOUS	14	0.06	76.21
48	PRESIDENT	35	0.15	74.84
49	DEAN	17	0.07	74.31
50	EMAIL	8	0.03	70.64
51	DELEGATION	16	0.07	70.62
52	VISIT	30	0.13	66.39
53	BELTANDROAD	5	0.02	64.93
54	LAUNCH	18	0.08	61.37
55	POSTDOC	5	0.02	59.52
56	CEO	8	0.03	59.06
57	GRADUATE	13	0.05	57.42
58	SYMPOSIUM	9	0.04	53.15

59	DOCTOR	24	0.10	51.73
60	ENGINEERING	19	0.08	51.23
61	UNV	4	0.02	48.26
62	WEBSITE	4	0.02	48.26
63	ENTREPRENEURSHIP	7	0.03	46.88
64	ECONOMICS	15	0.06	45.23
65	TWITTER	5	0.02	45.06
66	EU	6	0.03	44.93
67	NSA	6	0.03	40.50
68	SPEECH	20	0.08	40.39
69	COOPERATION	11	0.05	38.68
70	COLLABORATION	11	0.05	37.82
71	AMAZING	12	0.05	37.40
72	WORLD	48	0.20	37.38
73	DOCTORAL	6	0.03	36.88
74	CCP	6	0.03	36.88
75	CULTURAL	18	0.08	36.88
76	METADATA	4	0.02	34.80
77	FUNDED	11	0.05	34.75
78	BILATERAL	9	0.04	33.69
79	FUTURE	29	0.12	32.09
80	FORUM	11	0.05	32.04
81	COLLEGE	20	0.08	31.16
82	TEAM	26	0.11	30.77
83	TRANSLATION	10	0.04	29.50
84	INVITED	14	0.06	29.01
85	BELT	11	0.05	28.60
86	DEGREE	19	0.08	27.49
87	LECTURE	10	0.04	24.49
88	WELCOME	15	0.06	23.09

89	ARCHITECTURE	11	0.05	22.26
90	GREAT	38	0.16	20.80
91	JOINT	15	0.06	20.68
92	ORGANIZED	10	0.04	19.52
93	THANKS	14	0.06	18.47
94	CLIMATE	10	0.04	17.01
95	PARTNERS	10	0.04	10.87

Table 4: Key Brand Variables for the Seven C9 League Universities

We also generated a keyword list each for THU and PKU, which are the top two universities in China. We wanted to know about Twitter users' perception of these universities. See Tables 5 and 6 for the two keyword lists.

N	Keyword	Freq.	%	Keyness
1	CHINA	87	1.42	791.74
2	BLOCKCHAIN	19	0.31	349.99
3	SCHOLAR	23	0.38	227.89
4	UNIVERSITY	32	0.52	143.09
5	CHINESE	22	0.36	133.97
6	AI	14	0.23	127.69
7	STUDENT	23	0.38	118.49
8	FINTECH	7	0.11	111.29
9	QS	8	0.13	106.99
10	BIGDATA	6	0.10	97.92
11	CONGRATULATION	8	0.13	89.59
12	TECHNOLOGY	20	0.33	75.97
13	GLOBAL	13	0.21	62.35
14	LABORATORY	12	0.20	60.94
15	RESEARCH	23	0.38	60.20
16	SCIENCE	17	0.28	59.93

17	UNV	4	0.07	59.14
18	MBA	7	0.11	58.39
19	NSA	6	0.10	56.80
20	RANKING	8	0.13	56.11
21	PROFESSOR	13	0.21	54.22
22	TOP	21	0.34	51.29
23	PROGRAM	12	0.20	51.13
24	ALUMNI	6	0.10	47.24
25	SCHOLARSHIP	8	0.13	45.16
26	INNOVATION	9	0.15	44.05
27	CENTER	7	0.11	43.43
28	ASIA	10	0.16	41.87
29	CAMPUS	7	0.11	40.43
30	HONOR	5	0.08	40.27
31	CPEC	3	0.05	39.75
32	POSTDOC	3	0.05	35.26
33	LAUNCH	9	0.15	33.15
34	ACADEMY	7	0.11	30.03
35	WORLD	22	0.36	27.49
36	CEO	4	0.07	25.82
37	RESEARCHER	6	0.10	25.81
38	ENGINEERING	9	0.15	25.12
39	BELT	7	0.11	23.86
40	IOT	3	0.05	20.05
41	DOCTOR	10	0.16	19.81
42	ECONOMICS	7	0.11	19.39
43	INSTITUTE	8	0.13	18.31
44	DEAN	6	0.10	18.10
45	EU	3	0.05	17.45
46	INTERNATIONAL	13	0.21	17.05

47	BRI	3	0.05	16.82
48	MASTER	8	0.13	15.56
49	ORGANIZED	6	0.10	14.15
50	RANKED	4	0.07	12.77
51	VISITING	6	0.10	11.99
52	ARTS	7	0.11	11.06
53	PARTNERSHIP	6	0.10	10.15

Table 5: Key Brand Variables for Tsinghua University

N	Keyword	Freq.	%	Keyness
1	CHINA	113	1.35	1,022.20
2	STUDENT	70	0.84	508.66
3	CHINESE	61	0.73	489.35
4	PROFESSOR	50	0.60	362.74
5	UNIVERSITY	53	0.63	269.00
6	CAMPUS	18	0.22	155.31
7	METOO	8	0.10	131.73
8	BLOCKCHAIN	7	0.08	112.96
9	CENTER	13	0.16	104.28
10	HOST	18	0.22	99.29
11	SCHOOL	30	0.36	65.08
12	ACTIVIST	7	0.08	44.76
13	TOP	22	0.26	43.99
14	PRESTIGIOUS	8	0.10	43.49
15	TRANSGENDER	3	0.04	37.89
16	CPEC	3	0.04	37.89
17	FACULTY	8	0.10	35.23
18	ACADEMIC	11	0.13	33.13
19	PRESIDENT	16	0.19	31.97

20	STUDY	18	0.22	31.43
21	AI	6	0.07	30.46
22	RESEARCH	19	0.23	29.23
23	ASIA	9	0.11	28.53
24	CONGRATULATION	4	0.05	27.85
25	MOU	3	0.04	25.67
26	HONOR	4	0.05	24.32
27	CEO	4	0.05	23.33
28	MARXISTS	5	0.06	22.71
29	ALUMNI	4	0.05	19.69
30	DORM	3	0.04	17.79
31	FUTURE	15	0.18	17.64
32	SPICY	4	0.05	17.05
33	PROFOUND	6	0.07	16.89
34	SCHOLARSHIP	5	0.06	13.61
35	LECTURE	6	0.07	13.59
36	CONVERSATION	8	0.10	13.06
37	SPEECH	9	0.11	12.22
38	SCIENCE	10	0.12	11.86
39	FANCY	6	0.07	11.83
40	GRADUATE	5	0.06	11.72
41	VOLLEYBALL	3	0.04	11.60
42	CCP	3	0.04	11.47
43	RANKING	4	0.05	10.90
44	CENSORED	3	0.04	10.60

Table 6: Key Brand Variables for Peking University

Comparing the keyword lists of THU and PKU, ‘China’ had a higher keyness in PKU than in THU. In the keyword list of THU, we found that most of the advanced technology keywords were ranked highly, such as ‘blockchain’, ‘fintech’, ‘bigdata’, and ‘technology’. On the other hand, many keywords about campus facilities and people from PKU appeared on its keyword list, such as ‘student’, ‘professor’, ‘campus’, ‘center’, ‘dorm’, and ‘faculty’. People exhibited a positive perception towards both universities as some keywords represented positive attitudes, such as ‘prestigious’ and ‘congratulation’, which were ranked highly in the keyword lists.

6.3.2 Condensing the brand variables into groups and dimensions

After generating the key brand variables using Wordsmith keyword lists, we grouped these variables via factor analysis and divided the groups into several dimensions.

6.3.2.1 Frequency of key brand variables

Before conducting factor analysis, we created a table to record the frequency (in terms of appearance) of the key brand variables in the content. This frequency can be obtained using the Concord function in Wordsmith, which counts the number of hits for each variable. Afterwards, we generated a similar frequency using the VLOOK UP function in Excel.

6.3.2.2 Dimensionality reduction of opinion variables

After developing the frequency table, we ran SPSS factor analysis to group the 95 university-branding variables into groups and finally divided them into dimensions.

6.3.2.3 KMO Test and Bartlett’s Test results

Before performing factor analysis, the KMO Test and Bartlett’s Test should be done to examine whether the data fit into the analysis. The KMO Test stands for the Kaiser-Meyer-Olkin measure of sampling adequacy, first introduced by Kaiser (1970) and later modified by Kaiser and Rice (1974). The KMO statistic, which ranges from 0 to 1, indicates how well each variable in the data set is predicted by other variables. The larger the KMO, the stronger the correlation among the variables. Bartlett's Test is used to measure the degree to which the variables are related. A significance level lower than 0.05 according to Bartlett's Test is suitable for factor analysis.

Table 7 shows the KMO Test and Bartlett's Test results of our data. The KMO measure of our data was 0.564, which is higher than the acceptable 0.5 cut-off line. Our Bartlett's Test of sphericity was statistically significant (p = 0.000, less than 0.05). Moreover, the original data we had used was unstructured text data randomly posted on Twitter; thus, our sample was considered appropriate for factor analysis.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.564
Bartlett's Test of Sphericity	Approx. chi-square	7,296.617
	df	2,145
	Sig.	0.000

Table 7: KMO Test and Bartlett's Test Results

6.3.2.4 Grouping university brand variables

In this stage, we performed several iterations of SPSS factor analysis. In this analysis, we dropped variables with rotated factor loadings under 0.3, and we had 66 variables remaining, which were clustered into 27 factors by SPSS.

The 27 factors explained 60.725% of the variances in the university branding tweets based on content from global Twitter users. The alpha values of the 27 condensed factors ranged from 0.769 to 0.202 (see Appendix 2), indicating a reasonable level of internal consistency, especially for unstructured sample data. Table 8 shows the 27 factors from the SPSS factor analysis and the related statistical index.

Factor	Variable	Loading	Alpha	Eigen-values	Variance	Cumulative %
F1	RANKING	0.842	0.654	2.199	3.332	3.332
	WORLD	0.732				
	UNIVERSITY	0.683				
	TOP	0.635				
F2	PHD	0.746	0.579	2.046	3.099	6.431
	FUNDED	0.743				
	SCHOLARSHIP	0.586				
	MASTER	0.549				
	STUDY	0.382				
F3	DOCTORAL	0.735	0.434	1.845	2.795	9.226
	PROGRAM	0.641				
	DEGREE	0.546				
F4	CULTURAL	0.841	0.614	1.806	2.737	11.963
	BILATERAL	0.771				
	BELTANDROAD	0.662				
F5	EMAIL	0.927	0.769	1.721	2.608	14.571
	METADATA	0.926				
F6	CONGRATULATION	0.683	0.372	1.577	2.389	16.960
	GLOBAL	0.598				
	SCHOLAR	0.522				
F7	COLLABORATION	0.882	0.711	1.566	2.373	19.334
	UNV	0.871				
F8	HOST	0.803	0.374	1.541	2.335	21.668
	PROFESSOR	0.618				
	FUTURE	0.561				
F9	SCIENCE	0.727	0.419	1.536	2.328	23.996
	TECHNOLOGY	0.685				

F10	ORGANIZED	0.756	0.329	1.514	2.295	26.291
	BLOCKCHAIN	0.597				
	INVITED	0.395				
F11	ENGINEERING	0.719	0.372	1.465	2.219	28.510
	COLLEGE	0.676				
F12	ECONOMICS	0.792	0.28	1.427	2.162	30.672
	SCHOOL	0.552				
F13	JOINT	0.765	0.25	1.425	2.159	32.831
	MOU	0.646				
	RESEARCH	0.402				
F14	BIGDATA	0.683	0.38	1.424	2.158	34.989
	FINTECH	0.638				
F15	BRICS	0.836	0.464	1.405	2.129	37.118
	QS	0.665				
F16	BELT	0.846	0.479	1.404	2.127	39.245
	CPEC	0.704				
F17	LAUNCH	0.714	0.232	1.398	2.118	41.363
	PARTNERS	0.605				
F18	INTERNATIONAL	0.717	0.264	1.378	2.087	43.451
	SYMPOSIUM	0.625				
	CENTER	0.462				
F19	EU	0.779	0.358	1.365	2.068	45.518
	COOPERATION	0.639				
F20	CEO	0.737	0.275	1.328	2.012	47.530
	ALUMNI	0.595				
F21	TRANSLATION	0.725	0.345	1.326	2.009	49.539
	ACADEMY	0.722				
F22	DEAN	0.761	0.258	1.315	1.992	51.531

	DOCTOR	0.401				
F23	THANKS	0.783	0.271	1.265	1.917	53.448
	INSTITUTE	0.571				
F24	ENTREPRENEURSHIP	0.739	0.254	1.257	1.905	55.353
	INNOVATION	0.558				
F25	GRADUATE	0.744	0.202	1.205	1.826	57.179
	WEBSITE	0.665				
F26	ARCHITECTURE	0.792	0.244	1.185	1.796	58.975
	AMAZING	0.625				
F27	ROBOT	0.734	0.205	1.155	1.750	60.725
	TEAM	0.695				

Table 8: Grouping of Branding Variables via SPSS Factor Analysis

6.3.2.5 Dimensions of university branding

After classifying the 27 factors into dimensions, we employed triangulation, which is the process of applying more than one method in data analysis, to obtain better insights.

We first classified the 27 factors individually and separately based on the nature of the factors. We then compared the individual classification results. We accepted a resultant classification only when all three of us came up with the same factor group. For example, factors 2, 3, 25 mentioned and described university degrees, such as ‘PHD’, ‘master’, ‘doctoral’, and ‘graduate’. Thus, we grouped these three factors into one dimension named ‘degree’. Variables in factors 10, 14, and 27 were related to technology, such as ‘bigdata’, ‘fintech’, ‘robot’, and ‘blockchain’. Thus, we grouped these three factors into the ‘technology’ dimension.

Eventually, with several rounds of grouping, we classified the 27 factors into seven dimensions, which are shown in Table 9. The alpha values of the seven dimensions ranged from 0.598 to 0.213 (see Appendix 3), indicating a reasonable level of internal consistency, especially for our unstructured tweet data.

Dimension 1: Ranking					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F1	RANKING	0.842	0.519	0.778	0.882
	WORLD	0.732			
	UNIVERSITY	0.683			
	TOP	0.635			
F6	CONGRATULATION	0.683			
	GLOBAL	0.598			
	SCHOLAR	0.522			
F15	BRICS	0.836			
	QS	0.665			
Dimension 2: Degree					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F2	PHD	0.746	0.598	0.712	0.844
	FUNDED	0.743			
	SCHOLARSHIP	0.586			
	MASTER	0.549			
	STUDY	0.382			
F3	DOCTORAL	0.735			
	PROGRAM	0.641			
	DEGREE	0.546			
F25	GRADUATE	0.744			
	WEBSITE	0.665			
Dimension 3: Exchange					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F4	CULTURAL	0.841	0.339	0.291	0.539
	BILATERAL	0.771			

	BELTANDROAD	0.662			
F7	COLLABORATION	0.882			
	UNV	0.871			
F16	BELT	0.846			
	CPEC	0.704			
F18	INTERNATIONAL	0.717			
	SYMPOSIUM	0.625			
	CENTER	0.462			
Dimension 4: Administration					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F5	EMAIL	0.927	0.265	0.437	0.661
	METADATA	0.926			
F8	HOST	0.803			
	PROFESSOR	0.618			
	FUTURE	0.561			
F20	CEO	0.737			
	ALUMNI	0.595			
F22	DEAN	0.761			
	DOCTOR	0.401			
F23	THANKS	0.783			
	INSTITUTE	0.571			
Dimension 5: Discipline					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F9	SCIENCE	0.727	0.213	0.393	0.627
	TECHNOLOGY	0.685			
F11	ENGINEERING	0.719			
	COLLEGE	0.676			
F12	ECONOMICS	0.792			

	SCHOOL	0.552			
F21	TRANSLATION	0.725			
	ACADEMY	0.722			
F26	ARCHITECTURE	0.792			
	AMAZING	0.625			
Dimension 6: Technology					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F10	ORGANIZED	0.756	0.295	0.214	0.463
	BLOCKCHAIN	0.597			
	INVITED	0.395			
F14	BIGDATA	0.683			
	FINTECH	0.638			
F27	ROBOT	0.734			
	TEAM	0.695			
Dimension7: Research					
Factor	Variable	Loading	Alpha	Variance	Std. Deviation
F13	JOINT	0.765	0.318	0.283	0.532
	MOU	0.646			
	RESEARCH	0.402			
F17	LAUNCH	0.714			
	PARTNERS	0.605			
F19	EU	0.779			
	COOPERATION	0.639			
F24	ENTREPRENEURSHIP	0.739			
	INNOVATION	0.558			

Table 9: Seven Dimensions of University Branding

6.3.2.6 Distribution of brand dimensions

For further insights, we computed the scale mean (see Table 10 and Appendix 3) of each dimension of the Chinese university brands via SPSS to explore their weights.

Dimension	Theme	Std. Deviation	Mean	%
1	Ranking	0.882	0.37	22%
2	Degree	0.844	0.26	15%
3	Exchange	0.539	0.17	10%
4	Administration	0.661	0.31	18%
5	Discipline	0.627	0.29	17%
6	Technology	0.463	0.12	7%
7	Research	0.532	0.18	11%
Total			1.7	100%

Table 10: Scale Mean of Each Dimension of Chinese University Brands

The pie chart below (Figure 29) illustrates that the dimensions ‘ranking’ (22%), ‘administration’ (18%), ‘discipline’ (17%), and ‘degree’ (15%) occupy more than 70% of the branding for Chinese universities, which implies that people often mention university ranking, administration, discipline, and degree information when discussing Chinese universities. Moreover, the dimension ‘research’ occupies 11% of the branding for Chinese universities, which suggests that people often mention research when discussing Chinese universities. These people also frequently discuss foreign exchange and technology since the last two dimensions are ‘exchange’ (10%) and ‘technology’ (7%).

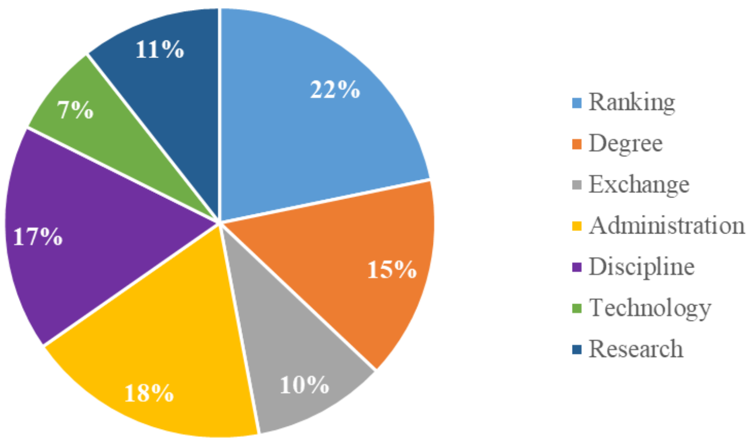


Figure 29: Distribution of the seven dimensions of Chinese university brands

6.3.3 Producing concept maps

To understand what people discuss among the collected content, we used Leximancer to generate different concept maps based on such content. Concept maps provide holistic views on Chinese university brands as communicated on social media.

The colours of the concept clusters indicate their levels of importance. Hot colours (such as red and orange) denote more important clusters, and cold colours (such as blue and green) denote less important ones. The size of a concept's dot reflects its level of connectivity in the concept map. The larger the concept dot, the more often the concept is related to other concepts in the map. Connectivity, in this sense, is the sum of all the co-occurrence counts of the concept with every other concept on the map. Details of the Leximancer concept mapping algorithm can be found in Appendix 1.

6.3.3.1 Overall concept map for seven universities

Figure 30 shows the concept map of the seven universities, generated according to the 1,040 tweets we collected. The concepts – that is, the nodes on the concept map – are the key brand variables we obtained using Wordsmith analysis. The most important concept, 'China', includes the following important concepts which are close to the 'China' node: 'top', 'prestigious', 'AI', and 'translation'. This means that when people talked about the seven universities, they often made references to China. 'Professor' is the second most important concept being discussed. Through the important concepts 'BRICS', 'CPEC', and 'beltandroad', we can also see that the university is one of the developments of Chinese foreign policy.

We also found some concepts which represent positive attitudes in the map, such as 'top' and 'prestigious' in the red cluster. The concept 'amazing' was linked to 'Chinese', 'China', 'campus', 'architecture', and 'student', and the concept 'prestigious' was linked to 'AI', 'China', 'Chinese', 'school', 'blockchain', and 'scholarship'.

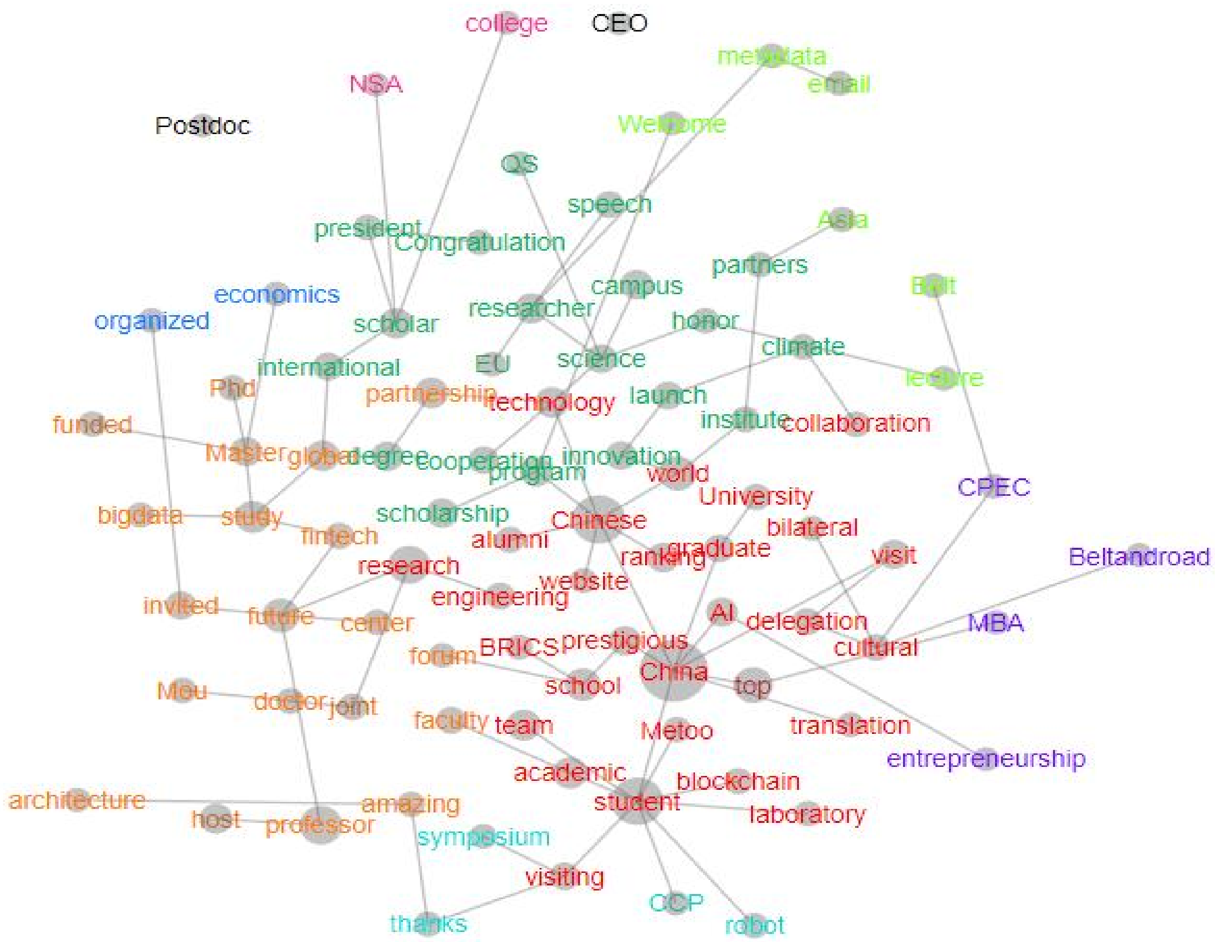


Figure 30: An overall concept map of the Chinese university brands

6.3.3.2 Concept map for Tsinghua University

Figure 31 shows the concept map of the THU brand. We can find that people seemed to think that THU is the best university in China because 'top' is the most important concept. In addition, the most important subjects of THU are 'science', 'engineering', and 'technology'.

6.3.3.3 Concept map for Peking University

Figure 32 shows the concept map of the PKU brand. In contrast to THU, which promotes science and technology, PKU promotes the liberal arts. The faculty and students of PKU were always mentioned together in the tweets. The prevalent concepts 'host', 'speech', and 'conversation' mean that PKU actively communicates with others.

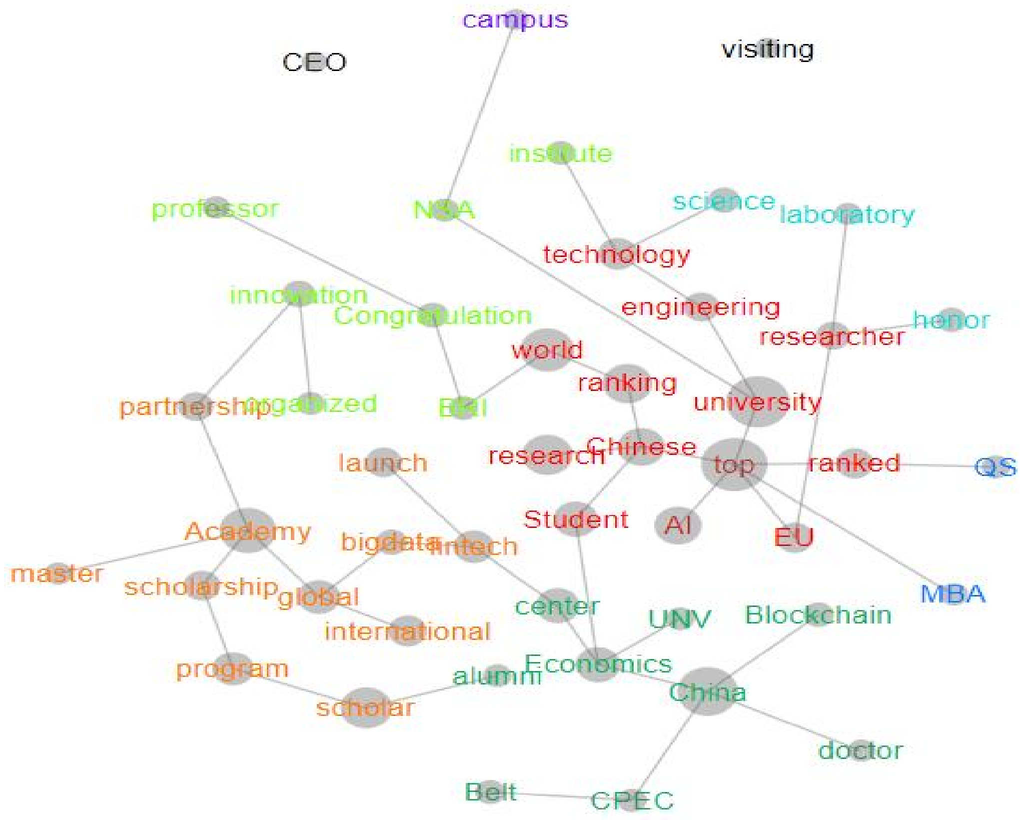


Figure 31: The concept map of the Tsinghua University brand

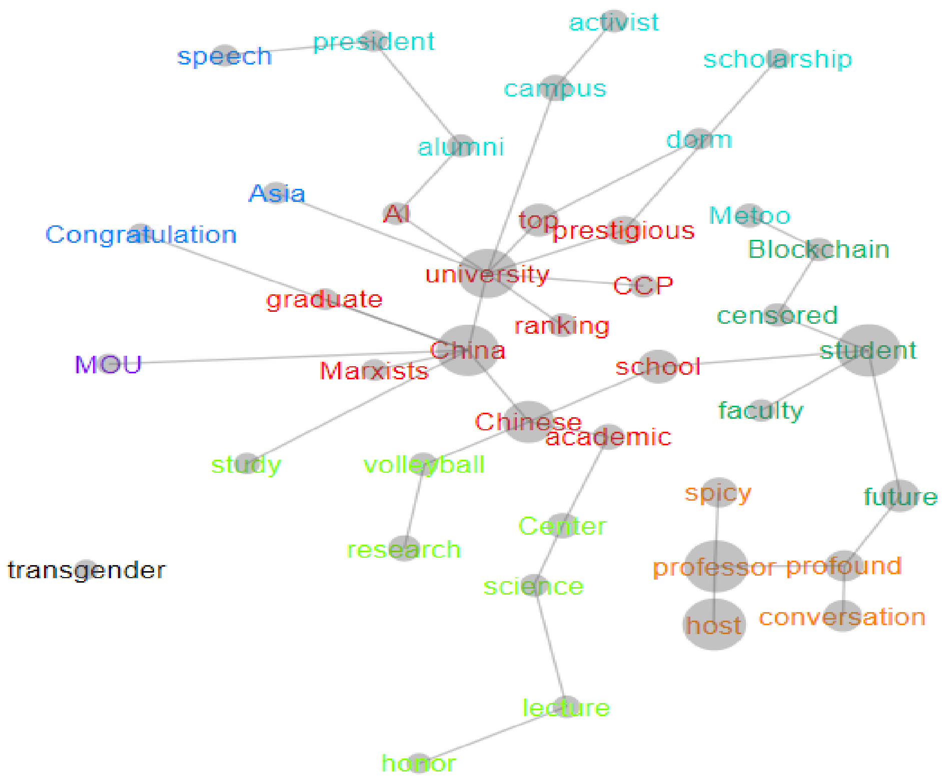


Figure 32: The concept map of the Peking University brand

6.3.3.4 Concept maps for the seven universities based on period

We have made two concept maps (Figure 33 and 34) based on two different periods. One is from 2015 to 2018, and the other is from 2019 to January 17, 2020. We compared the two concept maps to understand the changes in people's perceptions of the Chinese universities' brands. In the two concept maps, some concepts are shown to become more important over time, as observed by the colour changes. Some concepts attracted more attention from the past to the present, such as 'degree', 'economics', 'launch', 'program', and 'scholarship'. These concepts changed from green (from 2015 to 2018) to red (from 2019 to January 2020), becoming the most important concepts on the map. Second, 'AI', 'graduate', 'joint', 'research', 'technology', and 'world' also became more important concepts in 2019 to January 2020, changing from orange (from 2015 to 2018) to red (from 2019 to January 2020).

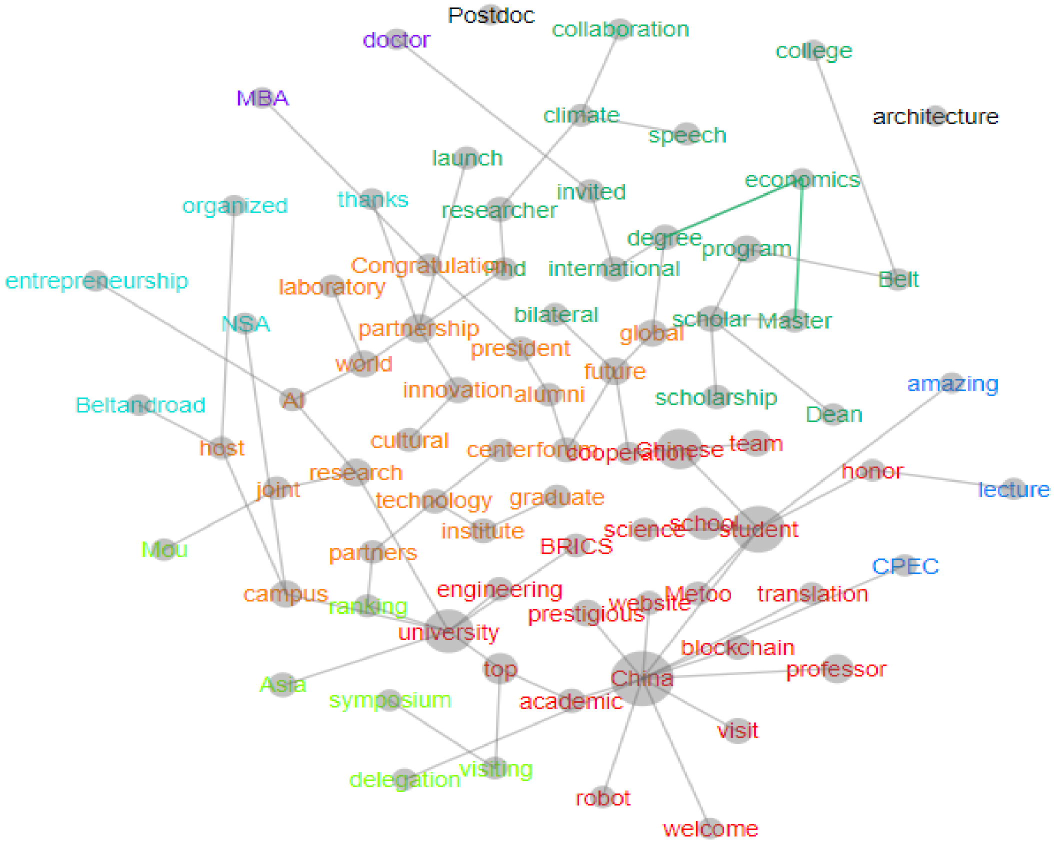


Figure 33: The concept map of the seven universities from 2015 to 2018

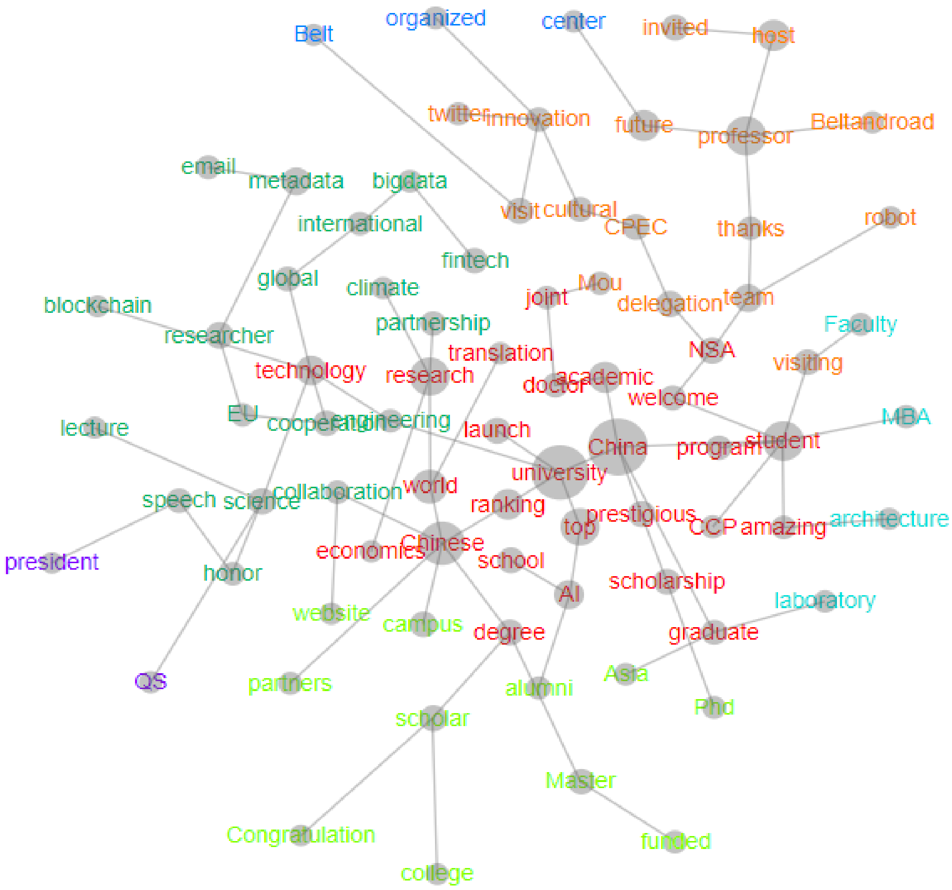


Figure 34: The concept map of the seven universities from 2019 to January 17, 2020

6.4 Interpretation of the results

First of all, 'China' and 'Chinese' are the two most important brand variables based on both our Wordsmith and Leximancer analyses. This means that the brands of these seven universities have become representative of China. Whenever people talk about these schools, they also talk about China or vice versa, making China's brand closely associated with the seven universities' brands. Therefore, as the top seven universities in China, they have the responsibility to promote the brand image or soft power of China.

According to the distribution of the seven dimensions, more than 70% of the brand communication is related to the ranking, administration, discipline, and degrees of the universities. This means that Twitter provides suitable information for people about universities.

When students choose the universities, they want to study in, they can get detailed information about these universities from Twitter and hear other users' opinions about the universities.

According to the two concept maps generated based on period, we can see that from 2015 to early 2020, people paid more attention to non-academic concepts. The exposure topics of these universities range from academic research to high-tech, diplomatic strategies and even participation in international and social issues. It can be seen that the strength of a university needs to be measured in diverse ways, not only in terms of academic performance.

We also used non-textual data to analyse user engagement. We found that users prefer liking and retweeting tweets but leave few comments. However, comments are important sources of people's opinions about the universities and even about China. Therefore, when organizations make promotions on Twitter, they can appropriately raise topics on social media for people to discuss. This way, they can establish communication and connection with online users rather than just post information.

In addition, we found that likes had a strong correlation with comments and retweets as well as a significant positive correlation between comments and retweets. These results could provide advice for organizations when they promote their brands on social media. Collecting as many likes as possible can lead to more retweets and comments, which can increase the visibility of the content and the collection of comments.

Chapter 7: Discussion and Recommendations

In this chapter, we discuss the implication of our methodology in our study, that is, the impacts on branding research. We also conclude by identifying the current limitations and making corresponding recommendations for further studies in the future.

7.1 Discussion and implications

In this report, we made use of text analytics to explore university branding on Twitter. We demonstrated a method to discover how social media users perceive the brands of Chinese universities, and how to process unstructured data. In the process of analysis, we found that the content we collected from Twitter is different from the content on general review websites such as Amazon. The content on social media is very diverse, as each tweet is on a different topic and from a different perspective. This diversity allowed us to discover more concepts related to the university brands. From the results of the analysis, we can know which majors, colleges and dimensions of China's universities are of greater concern. Twitter provides enough detailed information about the universities because some universities have already set up accounts on Twitter and are actively building relationships with users.

With the growing popularity of social media, what is said on these platforms can influence and even control readers' perception of things or ideas. Therefore, it is important that organizations monitor and promote their brand images on social media in order to find more potential customers and understand what they should improve. In order to know what people are saying about the brands, we need to process unstructured data and turn it into structured data. This is because people's opinions are usually expressed in the form of unstructured data, and only after converting it into structured can we see the direction and focus of the content generated by online users. In addition to listening to people's voices, organizations also need to make their voices heard appropriately on social platforms. Organizations can determine the theme of their promotional efforts according to the results of the analysis to gain more effective engagement.

7.2 Limitations

As we collected the relevant comments about the top seven Chinese universities of the C9 League, we conducted multiple analyses of these data to explore how communication on social media platforms affects university branding. Although we were careful to avoid making mistakes during the whole research process, we still found that our study had some limitations.

a) **Data volume:** We extracted a total of 1,040 sample data from Twitter for our text analytics. However, these data include each of the top seven Chinese universities in the C9 League. The volume of the sample data is a little too small for analysing so many subjects, which means the findings may lack generality.

b) **Unstructured raw data:** Since our major analysis method is text analytics, we focus on extracting users' comments about the target Chinese universities. However, comments from social media users are random. We could not make sure that every user is giving their opinions in a neutral and objective manner. In addition, some users' comments were irrelevant to the target Chinese universities, which means some data are useless for our analysis and make our research process slow. Therefore, the branding image of the Chinese universities that comes from analysing these comments may have deviations.

c) **Language setting:** We had decided to extract data in English, as the international language for global audiences, but some people have used other languages to communicate content about Chinese universities of the C9 League. Because the tools we used are primarily designed for English users and most people speak English on Twitter, our target website, we had to exclude content written in languages other than English. In other words, we lost some important comments from other languages. However, in chapter 2 we had indicated that the international students coming from mainstream western countries were much fewer in number than those from other countries. Thus, by focusing on content in English, our study can be of use for helping Chinese universities to brand themselves on social media platforms in ways that will attract more international students from mainstream western countries and contribute to the building of China's soft power.

d) **Data veracity:** On social media, since everybody can express their opinions, their words sometimes influence and even control others' thinking. For example, if a university wanted to promote their image on social media, they would post a large amount of similar and repeated promotional information on these platforms. Seeing the same content so many times, people would be more likely to think that it represents the truth without verifying it. We can infer that the university manipulated the content or data about their images. Therefore, the social media data we collected from Twitter may be manipulated, and the results and findings would be not be reasonable.

7.3 Recommendations

Based on the limitations suggested in the last section, we have provided several recommendations for further research.

a) **Data volume:** To get insights into Chinese university branding, we chose Twitter as our data source in this research. However, we believe that with an increase in data volume, more extensive and convincing results can be generated. In order to achieve that, more data content needs to be obtained. Data sources could be extended to other social media platforms, such as Facebook, Instagram and YouTube.

b) **Unstructured raw data:** In our research, unstructured data were processed and analysed, and we only focused on text data. However, we believe that deviations exist and that the results might be affected by irrelevant or biased datasets. Moreover, other types of data such as images, audio and video also contain meanings and insights. Thus, we suggest that more advanced applications can be used to do data processing in order to improve performance.

c) **Language setting:** In our research, we used English, the international language, as both data language and research language. However, we encountered several relevant tweets posted in other languages, which we simply dropped. Because those tweets also have their meanings, a multilingual dataset could generate more insights. In other words, with certain translation tools or analysis methodologies, better results can be obtained.

d) **Data veracity:** Given the amount of manipulated content on social media nowadays, we need to verify the veracity of this content using a number of strategies. One solution is to look at the opinions on other platforms. However, for our research, if there is a large amount of duplicated content in our data set, the data may be manipulated. We can also investigate the reputation and social status of the publisher (such as numbers of fans and subscriptions) to see if there is manipulated data. Moreover, in the future, artificial intelligence (AI) can help us identify potentially manipulated data.

Chapter 8: Conclusion

The major purpose of the study is to use text analytics to explore how communication on social media affects university branding. After we extracted the content relevant to the top seven universities in the C9 League from Twitter and followed regulated steps with these unstructured data to accomplish the analysis, we arrived at the insights into the relationship between social media and university branding. We now present the conclusions for the following three aspects: (1) the effectiveness of text analytics; (2) the effect of social media on university branding; (3) room for improvement.

- **The effectiveness of text analytics**. In our study, we mainly used the systematic text-analytics approach. Text analytics could help people to gain insight from social media users' comments and communication with others. The deep values of the text content would be found in these unstructured data, so that after digesting the useful information, universities could understand what they should do in order to improve their branding.
- **The effect of social media on university branding.** Based on the results in chapter 7, users on Twitter prefer to communicate detailed information about these universities with each other. In addition, people pay more and more attention to concepts beyond the academic, and the interactions between posters and commenters are becoming more and more important. Universities should note the value of posting a variety of content and making connections with users to lead them to give 'Likes' for the posts and enable more users to give 'Comments' or 'Retweets'. Thus, universities could improve their branding image by posting useful and appealing information in order to attract users to share the content on the social media platform.
- **Room for improvement.** At present, we have finished the whole study. However, we still need to make some adjustments for the study. Since this was the first demonstration in which we used the text analytics method for research, we still lack some knowledge

and skills. In further studies, we would be able to become more skilful by absorbing more academic knowledge about text analytics and use it to make improvements to our methods.

Reference

Book and Journal articles

Ali-Choudhury, R., Bennett, R., & Savani, S. (2009). University marketing directors' views on the components of a university brand. *International Review on Public and Nonprofit Marketing*, 6(1), 11.

Anctil, E. J. (2008). Recommendations for selling higher education. *ASHE Higher Education Report*, 34(2), 89–98.

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

Chauhan, K., & Pillai, A. (2013). Role of content strategy in social media brand communities: a case of higher education institutes in India. *Journal of Product & Brand Management*, 22(1), 40-51.

Constantinides, E., & Zinck Stagno, M. C. (2011). Potential of the social media as instruments of higher education marketing: A segmentation study. *Journal of marketing for higher education*, 21(1), 7-24.

Diebold, F. X. (2012). On the Origin (s) and Development of the Term "Big Data". Penn Institute for Economic Research Working Paper 12-037.

Edgington, D. (2011). Conditionals, causation, and decision. *Analytic Philosophy*, 52(2), 75-87.

Farshid, M., Pitt, L., & Botha, E. (2011). How sociable? An exploratory study of university brand visibility in social media. *South African Journal of Business Management*, 42(2), 43-51.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.

Gray, B. J., Fam, K. S., & Llanes, V. A. (2003). Branding universities in Asian markets. *Journal of product & brand management*, 12(2), 108-120.

Grimes, S. (2008). Text technologies in the mainstream: Text analytics solutions, applications and trends. *A white paper prepared for Text Analytics Summit*, *2008*.

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, 1(1), 60-76.

Hayhoe, R., Zha, Q., & Fengqiao, Y. (2012). Peking University–Icon of cultural leadership. In *Portraits of 21st century Chinese universities* (pp. 95-130). Springer, Dordrecht.

higher education courses: A conjoint approach. *Marketing Intelligence & Planning*, 19(3), 179–187.

Huynh, T., & Nguyen, N. (2015). Communicating university brand via social media: case: communicating Lahti University of Applied Sciences brand to international students using social media.

Judson, K. M., Aurand, T. W., Gorchels, L., & Gordon, G. L. (2009). Building a university brand from within: University administrators' perspectives of internal branding. *Services Marketing Quarterly*, 30(1), 54–68.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401-415.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, 34(1), 111-117.

Keller, K.L. (2008). *Strategic Brand Management: Building, Measuring, and Managing Brand Equity*. New Jersey: Pearson Prentice Hall.

Laroche, M., Habibi, M. R., & Richard, M. O. (2013). To be or not to be in social media: How brand loyalty is affected by social media?. *International Journal of Information Management*, 33(1), 76-82.

Laroche, M., Habibi, M. R., Richard, M. O., & Sankaranarayanan, R. (2012). The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty. *Computers in Human Behavior*, 28(5), 1755-1767.

Li, M., & Chen, Q. (2011). Globalization, internationalization and the world-class university movement: The China experience. *A handbook of globalization and higher education*, 241-255.

Mael, F., & Ashforth, B.E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational*

Behavior, 13(2), 103–123.

María Cubillo, J., Sánchez, J., & Cerviño, J. (2006). International students' decision-making process. *International Journal of Educational Management*, 20(2), 101-115.

Mattson, E., & Barnes, N. G. (2009). Social media and college admissions: The first longitudinal study. Retrieved July 11, 2010.

Moogan, Y. J., Baron, S., & Bainbridge, S. (2001). Timings and trade-offs in the marketing of higher education courses: A conjoint approach. *Marketing Intelligence & Planning*, 19(3), 179–187.

Narayana, C. L., & Markin, R. J. (1975). Consumer behavior and product performance: An alternative conceptualization. *Journal of Marketing*, 39(4), 1–6.

Pringle, J., & Fritz, S. (2019). The university brand and social media: using data analytics to assess brand authenticity. *Journal of Marketing for Higher Education*, 29(1), 19-44.

Qi, B., & Mackie, L. (2014). Utilising Social Media Technology to Raise Brand Awareness in Higher Education. *In WEBIST* (1) (pp. 400-405).

Rutter, R., Roper, S., & Lettice, F. (2016). Social media interaction, the university brand and recruitment performance. *Journal of Business Research*, 69(8), 3096-3104.

Singh, S., & Sonnenburg, S. (2012). Brand performances in social media. *Journal of interactive marketing*, 26(4), 189-197.

Sinha, A. (2007). *Sweet Spot: How to Maximize Marketing for Business Growth*. New Jersey: Wiley and Sons, Inc.

Stephenson, A. L., Heckert, A., & Yerger, D. B. (2016). College choice and the university brand: exploring the consumer decision framework. *Higher Education*, 71(4), 489-503.

Veletsianos, G. (2012). Higher education scholars' participation and practices on Twitter. *Journal of Computer Assisted Learning*, *28*(4), 336-349.

Verma, J. P., Agrawal, S., Patel, B., & Patel, A. (2016). Big data analytics: Challenges and applications for text, audio, video, and social media data. *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI)*, 5(1).

Waeraas, A., & Solbakk, M. N. (2008). Defining the essence of a university: Lessons from higher education branding. *Higher Education*, 57, 449–462.

Wijaya, B. S., & Putri, D. M. (2013). Is Social Media Impactful for University's Brand Image?. *Jurnal Manajemen Teknologi*, 13, 278-297.

Zun, H. E. (2002). University Spirit and University Brand [J]. *Journal of Wuhan University of Science and Technology*, 1.

Websites

American Marketing Association. (2019). Definitions of Marketing. Retrieved from https://www.ama.org/resources/Pages/Dictionary.aspx?dLetter=B&dLetter=B

Clement, J. (2019, Aug 14). Number of monthly active Twitter users worldwide from 1st quarter 2010 to 1st quarter 2019 (in millions). Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Cloudflare. (2019). What Is a Web Crawler? How Web Spiders Work. Retrieved from https://www.cloudflare.com/learning/bots/what-is-a-web-crawler/

Economist. (2018, Nov 17). Tsinghua University may soon top the world league in science research. Retrieved from https://www.economist.com/china/2018/11/17/tsinghua-university-may-soon-top-the-world-league-in-science-research

Ministry of Education, PRC. (2011). Introduction to "985 Project". Retrieved from http://www.moe.gov.cn/s78/A22/xwb_left/moe_843/201112/t20111230_128828.html

Ministry of Education, PRC. (2016). Introduction to "211 Project". Retrieved from http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/moe_1985/200804/9084.html

Ministry of Education, PRC. (2019). International students in China, 2018. Retrieved from http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201904/t20190412_377692.html

Ortiz-Ospina, E. (2019, Sep 18). The rise of social media. Retrieved from https://ourworldindata.org/rise-of-social-media

Pro Web Scraping. (2016). Web Scraping Vs Web Crawling. Retrieved from http://prowebscraping.com/web-scraping-vs-web-crawling

State Council, PRC. (2010, Sep 28). Studying in China. Retrieved from http://www.gov.cn/zwgk/2010-09/28/content_1711971.htm

State Council, PRC. (2019, Feb 23). China Education Modernization 2035. Retrieved from http://www.gov.cn/zhengce/2019-02/23/content_5367987.htm

State Council, PRC. (2019, Jul 17). Twelve policies to facilitate immigration to China. Retrieved from http://www.gov.cn/xinwen/2019-07/17/content_5410623.htm

The Python Software Foundation. (2019, Nov 25). The Python Tutorial. Retrieved from https://docs.python.org/3/tutorial/index.html

World Bank. (2019a). GDP Ranking. Retrieved from https://data.worldbank.org/data catalog/GDP-ranking-table

World Bank. (2019b). GDP Ranking PPP Based. Retrieved from https://datacatalog.worldbank.org/dataset/gdp-ranking-ppp-based

Wikipedia References

Belt and Road Initiative

Wikipedia (2019). Belt and Road Initiative. Retrieved from https://en.wikipedia.org/wiki/Belt_and_Road_Initiative

C9 League

Wikipedia (2019). C9 League. Retrieved from https://en.wikipedia.org/wiki/C9_League

Correlation coefficient

Wikipedia contributors. (2020, May 16). Correlation coefficient. Retrieved from https://en.wikipedia.org/w/index.php?title=Correlation_coefficient&oldid=957049160

China's military power

Wikipedia (2019). List of countries by military expenditures. Retrieved from https://en.wikipedia.org/w/index.php?title=List_of_countries_by_military_expen ditures&oldid=922499416

International student

Wikipedia (2019). International student. Retrieved from https://en.wikipedia.org/wiki/International_student

Peking University

Wikipedia (2019). Peking University. Retrieved from https://en.wikipedia.org/wiki/Peking_University

Project 985

Wikipedia (2019). Project 985. Retrieved from https://en.wikipedia.org/wiki/Project_985

Tsinghua University

Wikipedia (2019). Tsinghua University. Retrieved from https://en.wikipedia.org/wiki/Tsinghua_University

University Alliance of the Silk Road

Wikipedia (2019). Belt and Road Initiative. Retrieved from https://en.wikipedia.org/wiki/Belt_and_Road_Initiative#University_Alliance_of_the_Silk_Road

Web crawler

Wikipedia (2019). Web crawler. Retrieved from https://en.wikipedia.org/w/index.php?title=Web_crawler&oldid=926598987

Appendix 1

Leximancer concept-mapping algorithm

Extracted from:
Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. Behavior research methods, 38(2), 262-279.

(Salton, 1989), which is known to perform well as a text classifier (Dumais, Platt, Heckerman, & Sahami, 1998). This metric, derived from Bayesian decision theory, takes into consideration not only how frequently two words co-occur, but also how often they occur apart; this is similar to a log odds, or two-way contingency statistic. This metric gives a tighter binding of relevant terms to concepts that is suitable for extracting discriminating attributes of entities or concepts. For example, consider a document in which the occupational hazards of postal workers are discussed. To characterize the identity of a concept such as *dog* in this text, terms such as *bark*, *kennel*, and *tail* may be diagnostic, in that those terms may appear frequently alongside *dog* and infrequently elsewhere. Note that in *other* documents, *bark* could be diagnostic of trees. However, the term *postman*, although it may appear in relational encounters with *dog*, will occur more often elsewhere in other relationships. Thus, it seems appropriate to consider *postman* and *dog* as separate categories in this text, with the category of *dog* being discriminated by such words as *bark*, *kennel*, and *tail*.

The second stage, relational extraction, begins with the classification, or coding, of text segments, using the learned semantic classifiers. This is an implementation of naive Bayesian accumulation of evidence, using the term weights. After this process, the following statistics are available: concept count, concept co-occurrence count and relative concept co-occurrence frequency, and word count within each text segment classified within a concept. In addition, the text episodes classified within each concept and each pair of concepts can be retrieved and inspected.

There are many forms of statistical, data mining, and network analyses that could be performed on the concept statistics. It must be noted that the concepts show an approximate power law distribution of decreasing frequency within most data sets. As a result, co-occurrence information will lead to asymmetric attachment between concepts if the frequency of each concept is considered. In concrete terms, the relative co-occurrence frequency between two concepts will change, in general, depending on which concept the frequency is relative to. The resulting information can be expressed as an asymmetric concept co-occurrence matrix containing relative co-occurrence frequencies. Equivalently, this can be viewed as a concept network with directed weighted arcs. Relative co-occurrence frequency can also be considered as a frequentist approximation to the conditional probability of finding a second concept, given the first.

The choice of relative co-occurrence frequency as the measure of concept co-occurrence was influenced by two factors. This measure is much less tightly binding than a two-way contingency measure, and this is desirable because we now want to measure incidental interactions between concepts, such as those between *dog* and *postman*. Second, it was felt that throwing away all the asymmetric attachment information, which is endemic to natural language (see, e.g., Nelson, McEvoy, & Pointer, 2003), was not justified. In very many instances in which word or document similarity measures are required, including many analyses of results from HAL and LSA, the vector cosine measure is used. However, vector cosine is a symmetric measure. Neither is it equivalent to symmetrizing the matrix by pairwise averaging of link values. Finally, it is noted that Nelson and colleagues have used the relative frequency of word free association to calculate their free association norms (e.g., Nelson et al., 2003).

As a result of this choice of a real-valued asymmetric measure, many analytical tools are not applicable. Multidimensional scaling (MDS), factor analysis, and the vast majority of social network and graph theory measures either do not incorporate both directions of an asymmetric link or do not deal with real-valued links. In addition, the Leximancer method seeks to discover implicit, indirect relationships between concepts. This facility can allow discovery of previously unknown relationships. As a result of these requirements, the techniques of complex systems simulations and emergent behavior were examined as approaches for calculating a concept map.

The Leximancer concept-mapping algorithm is based on a variant of the spring-force model for the many-body problem (e.g., Chalmers & Chitson, 1992). The method used in Leximancer simulates forces between the concepts. It is a highly dissipative iterative numerical model and comes under the definition of a complex network system. The map is an indicative visualization that presents concept frequency (brightness), total concept connectedness (hierarchical order of appearance), direct interconcept relative co-occurrence frequency (ray intensity), and total (direct and indirect) interconcept co-occurrence (proximity). The formation of groups of directly and indirectly related concepts displays emergent behavior—that is, exhibits information that was not apparent by inspection of the input concept co-occurrence matrix. For this reason, it is not appropriate to demand that the final concept map should explain as much of the initial variance as possible. If that were the case, concepts that were initially unrelated by the direct co-occurrence measure should be unrelated on the map, which in turn would not identify indirect relationships.

The emergent concept groups are normally referred to as *themes*. Identification of themes by the observer is greatly facilitated by employing the hierarchy of concept connectedness. Each highly connected concept is a parent of a thematic region and can be used to characterize that region. It is noted that the problem of matching structure between different concept networks is made much harder by the variation in names of equivalent concept nodes between the networks. The comparative maps of the Holy Bible in French and in English, which will be presented later, provide an extreme example of this; none of the concept names are identical.

It must be emphasized that as with most algorithms, there are parameters that must be set, and these choices will be expected to influence the results. The most critical parameter is the length of the text segment. This is

Appendix 2

Reliability statistic for factors

Reliability Statistics			
F1	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.654	.713	4

Reliability Statistics			
F2	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.579	.649	5

Reliability Statistics			
F3	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.434	.499	3

Reliability Statistics			
F4	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.614	.660	3

Reliability Statistics			
F5	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.769	.837	2

Reliability Statistics			
F6	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.372	.393	3

Reliability Statistics			
F7	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.711	.713	2

Reliability Statistics			
F8	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.374	.451	3

Reliability Statistics			
F9	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.419	.420	2

Reliability Statistics			
F10	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.329	.395	3

Reliability Statistics			
F11	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.372	.373	2

Reliability Statistics			
F12	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.280	.340	2

Reliability Statistics			
F13	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.250	.353	3

Reliability Statistics			
F14	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.380	.386	2

Reliability Statistics			
F15	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.464	.466	2

Reliability Statistics			
F16	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.479	.486	2

Reliability Statistics			
F17	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.232	.241	2

Reliability Statistics			
F18	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.264	.315	3

Reliability Statistics			
F19	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.358	.359	2

Reliability Statistics			
F20	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.275	.279	2

Reliability Statistics			
F21	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.345	.354	2

Reliability Statistics			
F22	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.258	.264	2

Reliability Statistics			
F23	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.271	.324	2

Reliability Statistics			
F24	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.254	.281	2

Reliability Statistics			
F25	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.202	.233	2

Reliability Statistics			
F26	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.244	.246	2

Reliability Statistics			
F27	Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
	.205	.206	2

Appendix 3

Scale Statistic for the 7 dimensions of the university branding of China

Dimension 1: Ranking

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.37	.778	.882	9

Dimension 2: Degree

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.26	.712	.844	10

Dimension 3: Exchange

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.17	.291	.539	10

Dimension 4: Administration

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.31	.437	.661	11

Dimension 5: Discipline

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.29	.393	.627	10

Dimension 6: Technology

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.12	.214	.463	7

Dimension 7: Research

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
.18	.283	.532	9