An Analysis of the Relationship Between University Strategy Plans and University Rankings

Kim Nolle
Trinity College Dublin
nollek@tcd.ie

Ramya Viswanath
Trinity College Dublin
rviswana@tcd.ie

Benika Yadav
Trinity College Dublin
byadav@tcd.ie

Rupali Singh
Trinity College Dublin
rusingh@tcd.ie

Ashiqur Rahman Habeeb Rahuman
Trinity College Dublin
habeebra@tcd.ie

Nachiketh Janapareddy
Trinity College Dublin
janaparn@tcd.ie

15. April 2024

Abstract

This study analyses the relationship between the strategic plans of universities and their external rankings. Previous work on this topic has only considered highly ranked universities or limited the analysis to institutions from a specific geographical area. Our research aims to broaden the analysis by including universities from across the globe from a wider range of external rankings. We test the hypothesis that similarities in strategic planning correlate with similar rankings. Furthermore, we analyse how different categories of universities could influence the language of the strategy plans, including the geographical location, the age of the institutions or their specialisation in certain fields such as medicine, arts, or engineering. We use an Ensemble of Latent Dirichlet Allocation (E-LDA) to extract key themes from the documents and use a combination of statistical analysis, hierarchical clustering, and network analysis to analyse the contents of the strategy plans. This research not only contributes to the academic discourse on strategic planning in higher education but also offers practical insights for universities striving to enhance their strategic approaches in alignment with their ranking aspirations.

Keywords: *strategic planning, global university rankings, content analysis, ensemble of Latent Dirichlet Allocation, correlation, hierarchical clustering, network analysis*

1 Introduction

”Contemplative, orderly effort to produce fundamental decisions and actions that shape and guide what an organisation is, what it does, and why” is how we define strategic planning (Birnbaum and Edelson, 1989). Strategic planning is a process by which an organisation differentiates itself from its direct or emergent competition to achieve its mission. Strategic planning has been a critical part of universities' growth and development. It acts as a roadmap to achieve all the goals and activities in their mission and vision statement. This not only ensures growth but also ensures that the universities don't make decisions for short-term goals, but rather take steps to achieve long-term goals hence increasing viability and sustainability over time. Every university aiming to achieve excellence, develops, publishes and revisits its strategic planning from time to time (Fathi and Wilson, 2009). Strategic planning of universities has been studied for over a decade. The main agenda of research studies is to identify how the strategic planning of a university can affect its ranking worldwide. Birnbaum and Edelson (1989), not talking about strategic planning directly, but it explores the nature of strategic planning, touching upon important factors such as the development of a robust alumni network, the optimization of student-to-faculty ratios, and the essential for universities to engage in meaningful societal endeavours.

Universities leverage insights gained from global university rankings to achieve mid-term and long-term strategic goals (Hou et al., 2012). The question arises whether a university's global ranking is reflected in its strategic planning. In this research, we aim to answer this question. This begins with pre-processing the entire corpus which involves lemmatizing the text files and generating tokens. Subsequently, E-LDA is performed on the corpus to generate topics and feature vectors, which will be used to find the correlation.

1.1 Research Statement

This research aims to find the relationship between the language of strategic plans of universities and their external ranking. This study will analyse a large corpus of strategic plans for universities that display a wide range of characteristics including different geographical locations, old and new universities and different streams such as medicine, arts, engineering etc.. The following research question will be addressed in this study:

> *Is there a relationship between the texts of a university's strategic plans and the ranking of the university? We hypothesise that if such a relationship exists, then universities with similar strategy plans will have similar rankings.*

To measure the content of the strategic plans, we engineer features from the documents such as the frequency of n-grams and the probability of certain topics in the texts. To quantify relationships between the content of strategic planning and the ranking, we calculate correlation coefficients. More details can be found in section 3.

Previous studies analysing strategic plans or mission statements have limited scopes and only consider high-ranking universities (Breznik and Law, 2019; Bayrak, 2020) or only consider specific regions (Özdem, 2011). Our research project extends beyond this by considering strategic plans from a broader array of universities and looking for correlations between different categories of universities with external rankings.

2 Literature Review

This section gives an overview of previous work on strategic planning in higher education and its relation to external university rankings. First, we look at the characteristics of strategic planning and the requirements they should fulfil. We then give an overview of previous work analysing the contents of strategic plans as well as how they relate to external rankings.

2.1 Strategic Planning

Strategic planning in universities establishes a long-term objective and aims to develop a plan to achieve the mission, vision and goals of educational institutions. (Mahardhika and Raharja, 2023) discuss the various stages involved in strategic planning. First, the mission and goals of the schools are aligned. Second, a strategic plan is created based on a SWOT analysis to achieve organisational goals. Finally, the strategic plan is implemented.

Exploring the impact of strategic planning within an educational aspect, Priyambodo and Hasanah (2021) conduct a case study on how strategic planning in academics can significantly improve the quality of education. The strategic planning involves a comprehensive curriculum, teaching, professional staff, adequate facilities, effective management and support from stakeholders.

In their analysis of strategic planning within universities, Chance and Williams (2009) talk about the need for the development of rubrics for assessing the strategic plans of Universities. They find that the strategic plans of universities are dynamic and require flexibility and constant modification. It is therefore important to assess the quality of strategic plans to ensure that the plans meet educational objectives and respond well to dynamic changes. Chance and Williams advocate for the use of rubrics for assessing the

quality University's strategic plans. The rubrics include various statements of the rationale for education planning that include an introduction, organisation history, executive summary, mission, vision, goals and strategies for refining the plan and evaluating the outcomes and appendices.

Galbraith (1998) challenges the utility of the traditional methods of strategic planning, which may not suffice in the face of modern challenges faced by universities. Instead, he suggests a dynamic approach that incorporates flexibility, systems thinking, and the readiness for change. This perspective is particularly relevant to our investigation into the relationship between the language of strategic plans and university rankings. Galbraith's analysis underscores the necessity for universities to not only articulate their missions and goals but also to embed adaptability and innovation within their strategic frameworks. Such qualities are likely reflected in the linguistic features of their strategic plans, potentially influencing their standings in global rankings.

2.2 Content Analysis of Strategic Plans

There have been various studies that analyse the contents of university strategic plans, often with a focus on the mission statements. In a detailed review of strategic planning across 72 Turkish public universities, Özdem (2011) analyses the mission statements within these strategy plans. The universities are divided into 3 groups on the basis of their foundation year. The study found that these groups emphasised different themes. For example, universities founded before 2006 had a higher focus on research compared to newer universities, which had a higher emphasis on training, education and community service. This suggests that the age of the institution influences its strategic planning.

In a different context, Breznik and Law (2019) explores the thematic content of mission statements of universities ranked as the top 250 in 2016. A combination of network analysis and clustering was used to identify commonalities across these texts. They analysed words associated with the categories education philosophies, strategic orientation, social responsibility and value to stakeholders. The strong linkage of the word "research" suggests that this is a very important value to high-ranking universities, more so than education, and the occurrence of words such as "World", "Global" and "International" in all four clusters suggests that the globalisation of education is an aspect that concerns universities.

Bayrak (2020)'s research into the mission statements of top-ranked universities highlighted the usage of advanced text analytics tools, such as SAS Enterprise Miner for comprehensive text analysis to explore textual similarities within university mission statements. Additionally, the study used Cluster Hierarchy and Concept Link Diagrams to find thematic clusters within mission statements which presented a promising method for our project's analysis of textual similarities within university strategic plans.

The aforementioned studies suggest that themes could be useful in determining the similarity between strategy plans. One method to extract themes from texts is Ensemble of Latent Dirichlet Allocation (E-LDA), proposed by Srinivas and Rajendran (2019). Srinivas and Rajendran successfully used this methodology to reveal key themes in online student feedback. This helped identify strengths, weaknesses, opportunities, and threats (SWOT) as perceived by students, which in turn could be used to inform strategic university planning.

2.3 Strategy Plans and University Rankings

University rankings use a range of metrics that aim to quantify the quality of research and teaching in higher education institutions (Hazelkorn, 2007). This is used by a wide range of audiences, including students, academic institutions and employers, to compare different higher education institutions. These rankings affect the public perception of universities. High rankings can increase a university's reputation, which in turn helps it attract students, gain funding, and form collaborations with other institutions. As a result, rankings play a crucial role in strategic planning (Hazelkorn, 2007). The ranking results are used to identify weaknesses and problems that are then addressed by the strategic planning.

This is supported by the case study of a Chilean University by Véliz and Marshall (2022), which examines the influence of global university rankings on the strategic planning of the university. It reveals a strong correlation between the global ranking and academic excellence perception of the administrators

of the university, and how the university strategically utilised its global ranking to boost international visibility, attract international students and faculty and foster research collaborations.

Similarly, Dowsett (2020) explores how global rankings impact strategic planning in institutions. The article demonstrates how institutions adjust their tactics to meet worldwide ranking standards, especially the Academic Ranking of World Universities (ARWU). This study highlights that colleges do not only respond to rankings but actively plan strategies to influence them. Dowsett's study demonstrates the interplay between university strategic planning and worldwide rankings, providing insights into how institutions manage and possibly impact their global positions.

These previous works show that external university rankings are an integral resource for strategic planning because they highlight the scope of where universities can improve. Understanding this relationship can provide insights for university administrators and policymakers to make more effective strategic decisions.

In this regard, there are previous works that aim to study the relationships between strategic planning and rankings. For example, Hou et al. (2012) analyse the metrics that contribute most to four main global rankings (Academic Ranking of World Universities, QS Ranking, Webometrics Ranking of World Universities, and Performance Ranking of Scientific Papers for World Universities). Based on this, they propose a model for strategic planning with different actions depending on whether the university aims to stay in the top 30, enter the top 100 or increase its rank. We infer from these findings that similarly ranked universities, who are likely to have similar goals concerning rankings, will propose similar actions in their strategy plans.

Hall and Lulich (2021) study the strategy plans of 52 American universities that were ranked highly in 2020, with a focus on the topic of innovation. They also seek to relate their analysis to the ranking of the university. However, here they consider different categories of rankings (research, undergraduate education and innovation rankings) and analyse the similarities and differences between universities ranked highly for innovation compared to universities ranked highly for research.

To the best of our knowledge, there is no previous research that explicitly quantifies the correlation between the rank of a university and its strategy plans. This literature review also shows that many previous works limited the scope of their analysis either to certain countries or to a narrow range of university ranks. Our work differs in that we aim to analyse the strategy plans of a wider range of universities and quantify the relationship between strategic planning and university ranking.

3 Research Methodology

This section details our approach to answering our research question. Our approach can be divided into three stages: data collection, data pre-processing and data analysis.

3.1 Data Collection

We gathered strategic plans from a diverse array of universities, ensuring comprehensive coverage across various dimensions. Leveraging the QS World Rankings 2024 (QS Quacquarelli Symonds Limited, 2023b) as our primary reference, we focused on the top 600 universities listed.

We chose QS World Rankings for our project due to its comprehensive and globally recognized evaluation criteria, which encompass academic and employer reputation, faculty/student ratio, citations per faculty, and internationalisation metrics (QS Quacquarelli Symonds Limited, 2023a). This broad assessment framework aligns with our objective to explore the multifaceted impact of strategic planning on university performance. Additionally, QS World Rankings' structure, where bunching up of ranks starts at higher numbers, provides a more granular and differentiated view of university standings at the top levels, allowing for precise analysis. QS World Rankings' credibility and widespread acceptance make it an ideal benchmark, enabling us to conduct a meaningful analysis of how strategic planning elements correlate with universities' global standings. This approach not only ensures the relevance and

applicability of our findings but also contributes to the broader discourse on effective strategic planning in higher education.

To achieve a balanced representation, we divided these universities into six sets, each comprising approx. 40 institutions across the ranking ranges 1–100, 101–200, 201–300, 301–400, 401–500, and 501–600. This results in a dataset of strategic plans from 256 universities. When selecting the 40 universities per ranking range, we also consider the following selection criteria to ensure a diverse set of universities:

1. **Continent:** Every set included at least one university from each region (America, Africa, Europe, Asia, Oceania) to ensure global representation.

2. **Public and Private:** To capture institutional diversity, we included both public and private universities.

3. **Type:** We categorised institutions based on their academic focus, incorporating Arts, Engineering, Medical, and Science institutions.

4. **Age of Institution:** A mix of old and new institutions is considered to reflect the evolving landscape of strategic planning practices.

The strategic plans were sourced directly from the official websites of the selected universities and manually downloaded. We ensured the inclusion of the latest available strategic plan for each university to maintain currency. The text content of each plan was then extracted from the strategic plan PDFs and saved into individual text files, preserving the original formatting and structure. The additional details for each university like the type of university (all streams offering university, technical university, medical school or business school), whether it's public or private, year of establishment, and location were also sourced directly from the university website. This dataset of strategic plans collected and all universities' details organized in a CSV file are available for future use on our GitHub repository[1] for reproducibility as well as future research work.

3.2 Data Pre-Processing

Before conducting our analysis of the textual data extracted from the strategic plans of universities, we pre-process the data to ensure consistency and remove noise. First, we remove all punctuation marks and convert the entire corpus to lowercase to standardize the text and facilitate uniform analysis. In this step, we ensure that only alphabetic characters are retained, to exclude irrelevant data components, such as page numbers.

Following this, we remove common stopwords, focusing our analysis on the substantive content of the text rather than its stylistic elements (Omar et al., 2015). This step is crucial for reducing the dataset's noise and dimensionality. We use WordNet lemmatisation, an algorithm leveraging the WordNet lexical database, for a more nuanced reduction of words to their base or dictionary form. Lemmatisation, unlike stemming, takes into account the morphological analysis of the words, leading to higher-quality textual data for analysis.(Srinivas and Rajendran, 2019)

To capture meaningful patterns in the text, we consider the frequency counts of words (uni-grams) and sequences of two words (bi-grams). Bi-grams are particularly useful for disambiguating word sense (Pedersen, 2001). After calculating the absolute frequency counts, we normalise these to relative frequency counts to account for the variability in document lengths. For the correlation analysis, we only consider the top 20 uni-grams and top 20 bi-grams per document to reduce data dimensionality whilst keeping the most important information.

In addition to these traditional pre-processing steps, we employ phrase detection algorithms to identify and preserve valuable multi-word expressions that could be lost in individual word analysis. Using

[1] https://github.com/knolle1/CS7IS4-Group-7

the Gensim library, we detect phrases in the corpus, allowing us to treat significant bi-grams as single units in subsequent analysis steps. This approach is particularly beneficial for capturing complex concepts within strategic plans.

3.3 Data Analysis

3.3.1 E-LDA Analysis

To capture more high-level concepts in the strategy plans, we generate features that represent the themes in the documents using Ensemble of Latent Dirichlet Allocation (E-LDA), a combination of multiple LDA models which was proposed by Srinivas and Rajendran (2019).

Latent Dirichlet Allocation LDA (Blei et al., 2003), which is a statistical approach, posits that documents are mixtures of topics, where each topic is characterised by a distribution over words. Starting with random topic assignments for each word, LDA iteratively refines these assignments. This refinement process involves evaluating each word's topic based on how prevalent the topic is across documents and the frequency of the word within those topics. Through this iterative refinement, LDA adjusts topic assignments to maximise the likelihood of observing the words given the topics, leading to a distribution of topics per document and a distribution of words for each topic.

To perform E-LDA, we first train multiple LDA models of different topic sizes to identify the topics within the strategic plans. Next, the coherence of each model was evaluated using the 'c_v' coherence measure. This measure is a robust indicator of a model's ability to produce meaningful and cohesive topics. Models that demonstrate a high degree of coherence are selected for further analysis, ensuring that our thematic investigation is grounded in models that accurately reflect the underlying structures of the strategic plans. The selected high-coherence models are then combined by calculating the average of their outputs.

We used Model Averaging as our strategy to ensemble the different LDA models. This method involves training multiple LDA models, each with different configurations (like varying numbers of topics), and then averaging their topic-word distributions. It aims to stabilize the topics across models, making our findings more robust.

E-LDA enables us to uncover the underlying thematic structures in the strategic plans, facilitating a deeper understanding of the strategic focuses articulated by universities. These topic features are also used in our further analysis (Shini and Srividhya, 2023).

3.3.2 Correlation Analysis

After generating feature vectors consisting of uni-grams, bi-grams and the topic features, the Euclidean distance between these feature vectors of all university pairs is calculated. This distance describes the similarity between the strategy plans. The difference between university rankings for all university pairs is also calculated. This distance describes the similarity in external university rankings. We hypothesise that if a relationship between strategy plans and external university rankings exists, then universities with similar strategy plans will have similar rankings. We therefore expect these two distances to be dependent on each other.

To test this, we conduct a two-sided hypothesis test using the Spearman rank correlation coefficient. This test statistic gives the correlation between the rank of two variables rather than the correlation between their values (Daniel, 1990). With X and Y being the distance between feature vectors and the distance in the ranking respectively, the null- and alternative hypotheses are as follows:

$$H_0 : X \text{ and } Y \text{ are independent}$$
$$H_1 : X \text{ and } Y \text{ are either directly or inversely related}$$

The Spearman rank correlation coefficient provides a measure of the strength of the relationship between two variables, even if the relationship is not linear (Daniel, 1990), unlike Pearson correlation,

which only captures linear relationships (Runkler, 2020). This is useful for our analysis because we are interested in the relationship between X and Y regardless of whether it is linear or not.

If the two-sided hypothesis test using the Spearman rank correlation coefficient rejects the null hypothesis, then this confirms our research hypothesis that similarities in strategic planning correlate with similar rankings. On the other hand, if we fail to reject the null hypothesis, then there could be two implications: The first is that there is no quantifiable relationship between strategic planning and global rankings. The second implication other patterns in strategic planning and rankings are not captured by this hypothesis.

Because of this, we also analyse the direct correlation between the topic features produced by the E-LDA and the university rankings. This would allow us to investigate the relationship between topics and rankings independently from other universities. In addition, we consider topic similarity and hierarchical clustering as described in the next section.

3.3.3 Topic Similarity and Hierarchical Clustering

We first compute the cosine similarity, a measure used to quantify the closeness or similarity between vectors representing topics or documents. Cosine similarity calculates the cosine of the angle between two vectors, defined by the formula:

$$\text{cosine similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} \quad \text{where } \mathbf{A} \text{ and } \mathbf{B} \text{ are feature vectors}$$

This metric highlights how similar two items are irrespective of their size. The computed similarities are used to understand and analyze the relationships among topics or documents, offering insights into the thematic structure and affinities within the dataset (Popat et al., 2017).

Following the calculation of cosine similarities, the similarity matrix is converted into a distance matrix. This transformation is critical for the subsequent clustering analysis and is achieved by subtracting the similarity scores from one:

$$\text{distance} = 1 - \text{similarity}$$

The distance matrix effectively represents the dissimilarity between each pair of vectors, preparing the data for hierarchical clustering.

Following the correlation analysis, we utilize hierarchical clustering to visually represent the thematic similarities between universities. The Agglomerative Hierarchical clustering method, (Vijaya et al., 2019) is a bottom-up approach, where each data point or university is treated as a cluster itself initially and merged gradually based on their similarities. By converting our cosine similarity matrix (derived from the normalized feature vectors) into a distance matrix, we apply Ward's method to cluster the universities. minimize the sum of squared differences within all clusters, which effectively means it groups the data points into clusters in a way that the total within-cluster variance is minimized at each step of clustering. The resulting Dendrogram visualizes how universities group together based on their strategic thematic content, providing a bird's-eye view of the strategic landscape across institutions (Huang et al., 2013).

3.3.4 Network Analysis

Expanding upon our E-LDA and statistical correlation methods in sections 3.3.1 and 3.3.2, we employed a procedure where network analysis was used to represent the connections among the themes that emerged from strategic plans. It is through network analysis that the interconnection of strategic concepts can be elaborated in terms of thematic linkages.

Network analysis, commonly represented as networks with nodes and edges, is a systematic way of studying structural relationships within large complex datasets. Nodes in these networks typically represent discrete elements such as themes, phrases, or concepts whereas edges are indicative of connections or interactions between them. In particular, this method allows for the assessment of entities' relative

importance and strength of their links and is very effective at revealing hidden patterns plus clusters within data sets (Drieger, 2013). We use network analysis to gain insight into the underlying objectives and linkages that make up the universities' strategic vision's intricate network.

Each university is considered as a node labelled with their 'global rank', and the similarity of their strategic plans became an edge. The strength of an edge or connection is obtained by calculating the cosine similarity between the feature vectors of the universities 3.3.3.

Louvain's algorithm also made it possible to build communities, based on the weight (Ghosh et al., 2018) that can be visually differentiated by colour. The matrix facilitated the generation of a network plot by grouping universities into communities based on the thematic similarity of their strategic plans, hence forming clusters. To distinguish among institutions that have similar rankings they are named as rank-count. This reorganization aims at making nodes group together depending on how dense or tightly related they are to one another.

4 Results

This section presents and interprets the results of our different analyses.

4.1 E-LDA Topics

Our E-LDA model produced 10 different topics. The words contained in the distribution of each topic are visualised in Figure 1 as word clouds. The size of the words in the word cloud is proportional to their importance within the word distribution. This helped us interpret the topics, which can also be seen in Figure 1.

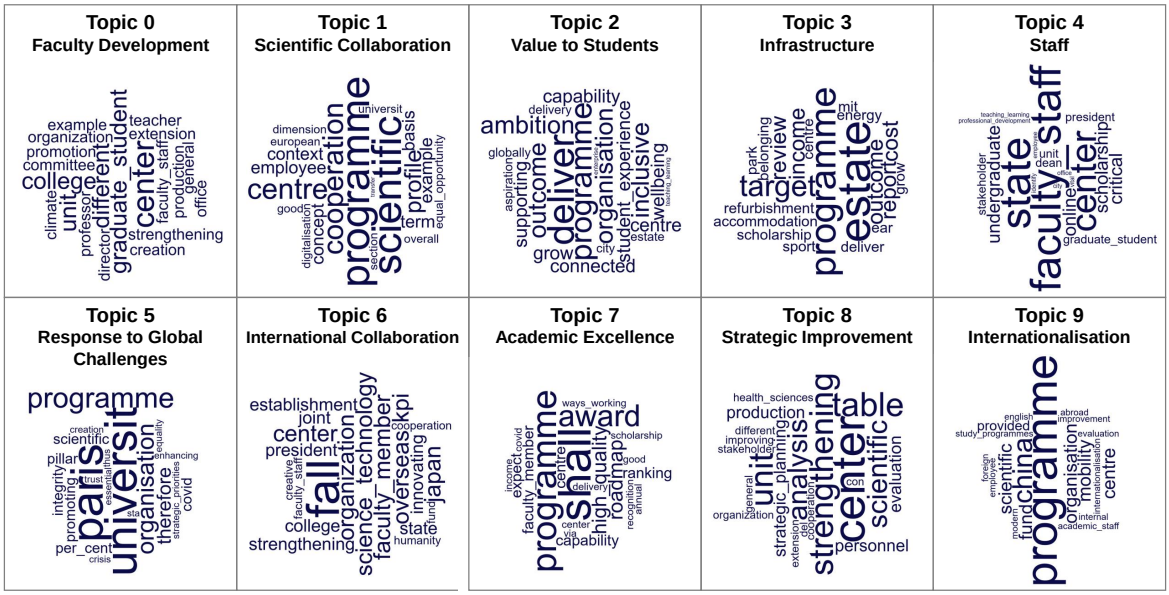

Figure 1: Word clouds of the topics extracted from E-LDA

To evaluate the performance of the E-LDA, we calculated a topic similarity matrix, which can be seen in Figure 2. The bright red squares on the diagonal that each topic matches well with itself, which is expected. If there are other red spots away from the middle diagonal, it suggests that some topics look a lot like each other. The dark blue areas in the rest of the matrix are good because they show that most topics are different from each other, which is what we want. If we see lighter blue, it means those topics

are somewhat similar and might be grouped into bigger themes. This matrix indicates that our model is working right by showing us if our topics are well separated.

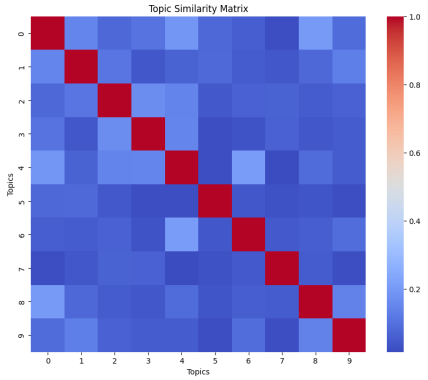

Figure 2: Heatmap of inter-topic similarity scores from E-LDA model

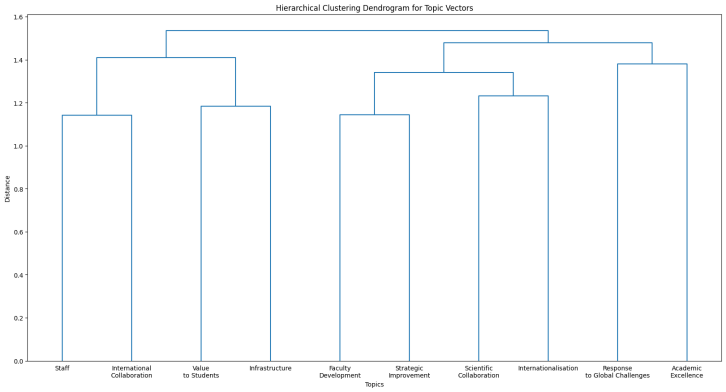

Figure 3: Thematic clustering of university strategy topics

The dendrogram Figure 3, displayed above provides valuable insights into the thematic relationships within the strategic plans of universities. It reveals how closely related certain topics are, based on the clustering patterns observed. For instance, the proximity between 'Faculty Development' and 'Strategic Improvement' suggests a potential overlap in content, indicating that strategic improvements within these institutions may be closely linked to the development of their faculty. On the other hand, 'Internationalisation' appears as a distinct and separate entity, highlighting its unique position in the strategic planning discourse. The clustering of 'Value to Students' and 'Academic Excellence' at a shorter distance signifies a strong connection, perhaps underscoring a common focus on enhancing student experiences and outcomes as a pathway to achieving excellence. This visual representation enables us to pinpoint areas of strategic convergence and divergence among the universities' plans, thereby facilitating a nuanced understanding of the strategic emphases laid out in the documents.

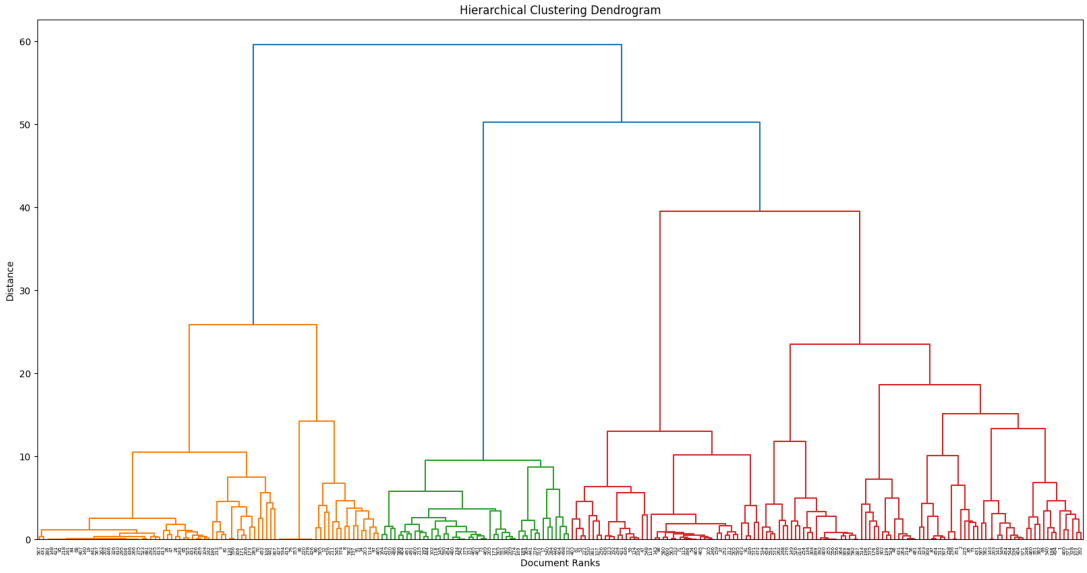

Figure 4: Thematic clustering of university strategy topics

In the plot Figure 4, the large, distinct clusters marked by different colours suggest that universities can be broadly categorized into groups with significant differences in the underlying metrics used for ranking. Within these major clusters, further subclusters hint at more subtle relationships, perhaps reflecting specialized characteristics or niche areas of excellence. Close linkages near the base of the

dendrogram highlight universities with very similar profiles, which could indicate comparable academic, research, or cultural environments. These insights could be invaluable for prospective students, faculty recruitment, or research collaboration efforts, as they illustrate not just the rank but the contextual closeness of universities within the broader educational landscape.

4.2 Correlation Analysis

As described in section 3.3.2, we first analyse the correlation between the similarity of strategy plans and similarity in external rankings. To calculate the similarity of strategy plans we considered different combinations of features. Table 1 shows a summary of the calculated Spearman rank correlation coefficients. It can be seen that there are only two cases where $p < 0.05$: when only bi-gram features are used and when uni- and bi-gram features are used. This means that the null hypothesis (the similarity in these feature vectors and the similarity in rankings are independent) is rejected, suggesting that there is a significant correlation in these cases. However, when looking at the correlation coefficients for these feature combinations (-0.009995 and -0.02255 respectively), it can be seen that these are close to 0, which suggests that the relationship is very weak. We, therefore, conclude that there is no quantifiable relationship between the similarity of strategy plans and similarity in external ranking. Our hypothesis that similarly ranked universities have similar strategy plans is therefore falsified.

Document Features	**Correlation**	**p-value**
Uni-gram features	-0.0005806	0.8832
Bi-gram features	-0.009995	0.01145 *
Topic features	0.003484	0.3781
Uni-gram and bi-gram features	-0.02255	1.162e-08 *
Uni-gram, bi-gram and topic features	-0.004676	0.2368

Table 1: Spearman correlation between the similarity of strategy plans and similarity in external rankings (* where p $<$ 0.05)

Our correlation analysis is extended by calculating the correlation between topic features and the external rankings directly. Table 2 presents the results of the correlation calculations. These are also presented visually in figure 5 to be able to see differences in correlations more easily. The results show that for the topics *Value to Students*, *Strategic Improvement* and *Internationalisation*, the correlations have a p-value < 0.05, which means that these topics are related to the rankings of the universities. For the other topics, the null hypothesis that the topics and rankings are independent cannot be rejected and so no relationship could be measured.

Idx	**Topic**	**Correlation**	**p-value**
0	Faculty Development	0.01889	0.765
1	Scientific Collaboration	-0.0655	0.2994
2	Value to Students	-0.1533	0.0146 *
3	Infrastructure	-0.0727	0.2493
4	Staff	0.0429	0.4964
5	Response to Global Challenges	0.0860	0.1725
6	International Collaboration	0.00297	0.9624
7	Academic Excellence	0.0926	0.1418
8	Strategic Improvement	0.1665	0.0079 *
9	Internationalisation	0.2232	0.0003 *

Table 2: Spearman correlation between the topic features and external rankings (* where p $<$ 0.05)

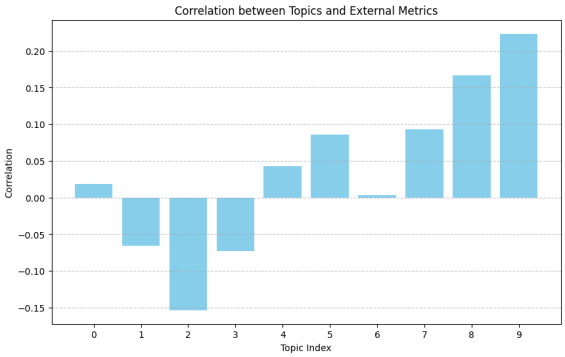

Figure 5: Barchart of correlations between topic features and rankings

The topic *Value to Students* is negatively correlated with external rankings, which suggests that lower-ranked universities place more emphasis on providing value to their students to distinguish themselves

from other universities. The topics *Strategic Improvement* and *Internationalisation* are positively correlated, which shows that highly ranked universities focus on their international reputation and place emphasis on strategic planning to improve and meet stakeholders' expectations.

4.3 Network Analysis

The network plot obtained, as mentioned in 3.3.4shows the formation of around ten clusters, each of which is denoted by a different colour. The clusters represent groups of universities with strategic plans that have similar themes/topics. For example, the presence of one cluster that is densely connected like the dark blue one could mean several highly ranked universities that focus more on certain thematic areas such as Internationalization or Strategic Improvement, according to our E-LDA analysis 4.1.

The graph shows that no one cluster dominates ranks higher than others, which means that leading institutions may not necessarily agree on a single set of strategic themes. Besides, various universities with different strategies are located in different parts of the ranking spectrum as seen in Figure 7 and Figure 8; indicating that the universities with the same level of ranking focus on different ranges of strategic themes

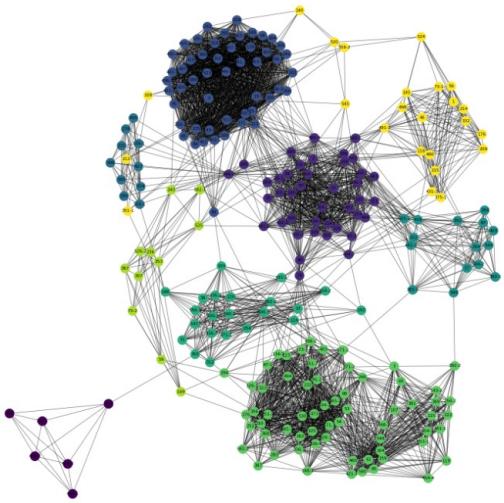

Figure 6: Network Analysis with Clusters using Louvain's Algorithm

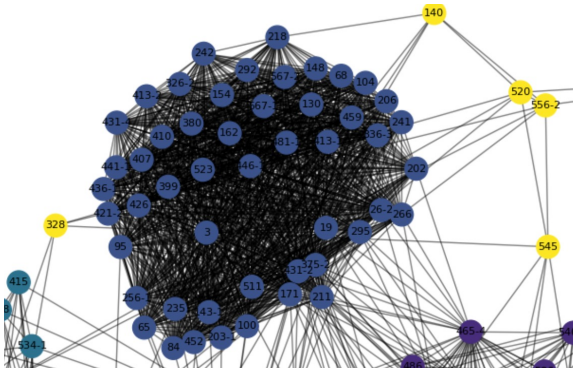

Figure 7: Intra Cluster Network with varied Global Ranks

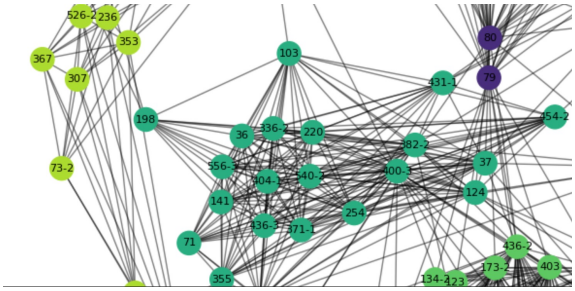

Figure 8: Inter-cluster Universities with Similar Global Ranks

The less-dense inter-cluster connections than intra-cluster ones represent over-arching strategies that run through several theme-based groups. These connections denote shared goals or challenges common to all universities regardless of their primary strategic orientations.

Our statistical analysis shown earlier is supported by the missing pronounced clustering in rankings despite interconnectivity. The network graph does not confirm our initial hypothesis that is, universities that are similarly ranked should possess related strategic plans. This diversity is exemplified by the network graph. It may be observed that universities have some common strategic themes but these do not necessarily correspond to their ranking outcomes.

5 Further Analysis

Our correlation analysis found no relationship between the similarity in strategic plans and the similarity in ranking. We therefore decided to extend the analysis to see if there were stronger relationships between the strategic plan and other characteristics of universities.

To do this, we used the top n-grams that represent the textual content in strategic plans to cluster universities based on how similar they were. To perform the text clustering, the top 40 Uni-grams and 40 Bi-grams were extracted from the pre-processed strategic plans instead of 20, to capture more detailed linguistic patterns with a wider range of vocabulary for a comprehensive representation of the textual content. The extracted n-grams from the textual content serve as the features for clustering. First, we tried and tested Density-Based Spatial Clustering of Applications with Noise (DBSCAN). DBSCAN (Deng, 2020) is a clustering algorithm that groups the data based on density and can identify arbitrary-shaped clusters making it ideal for complex datasets where the number of clusters is unknown. Additionally, DBSCAN clusters all the points it detects as noise that do not belong to any specific density cluster, and groups them in a cluster labelled as -1. Figure 9 shows the clusters formed using DBSCAN, in 2D. Each point on the scatter plot represents an individual data point or university that has been transformed using t-SNE (t-Distributed Stochastic Neighbor Embedding), a technique for dimensionality reduction, to visualize our high dimensional data in 2 dimensions. From Figure 9, it is evident that DBSCAN algorithm didn't work for our dataset as it labelled a majority of universities as noise which might be because of the diverse range of n-grams in our dataset and no specific density-based clusters were identified.

We then tried the Agglomerative Hierarchical clustering, (Vijaya et al., 2019) for a nuanced analysis of how the strategic plans are interconnected and to determine the varying levels of similarity between them and understand the hierarchy of clusters, which is visualised using a dendrogram. We used the Term Frequency-Inverse Document Frequency matrix (TF-IDF) to first convert the bag of n-grams into numerical representation for creating the dendrogram and further clustering. We then visually interpreted the dendrograms to determine the optimal number of clusters. Each university is assigned to a cluster using Agglomerative Hierarchical clustering. The composition of each cluster is examined to understand which universities were grouped, by analysing the distribution of categorical attributes - continent, and type, and the distribution of university QS rankings in each cluster.

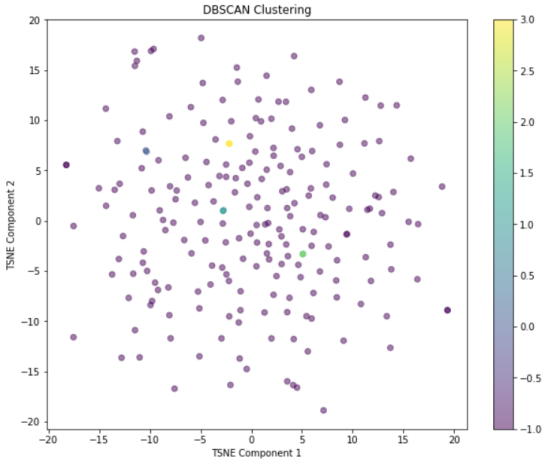

Figure 9: DBSCAN clusters in 2D

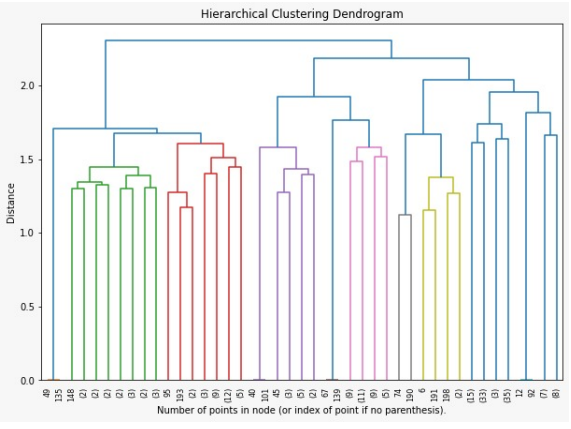

Figure 10: Dendrogram of Agglomerative Hierarchical Clustering

Agglomerative_Cluster	continent	
0	Africa	6.122449
	Asia	16.326531
	Europe	73.469388
	South America	4.081633
1	Africa	3.488372
	Asia	37.209302
	Australia	1.162791
	Europe	36.046512
	North America	18.604651
	Oceania	3.488372
2	Asia	11.764706
	North America	88.235294
3	Asia	14.285714
	Europe	71.428571
	North America	14.285714
4	Asia	19.230769
	Australia	13.461538
	Europe	46.153846
	North America	7.692308
	Oceania	13.461538

dtype: float64

Figure 11: Continent composition of Agglomerative Hierarchical Clusters with percentages

The n-grams clustering analysis yielded diverse clusters of universities based on the content of their strategic plans. The dendrogram Figure 10 indicated the presence of distinct branches that suggested the formation of five clusters. Hence, the number of clusters used to perform Agglomerative Hierarchical clustering was 5. The resulting clusters showed that their composition varied significantly. Despite the presence of distinct clusters based on n-grams, no strong correlation was observed between cluster assignments and attributes such as university rankings, type, or year established. But analysing the composition of clusters in terms of continents or location, as presented in the Figure 11, revealed that some clusters were dominated by universities from specific continents, while others had a more diverse representation.

6 Discussion

Our results showed that there is no correlation between the similarity in strategic plans and the similarity in external ranking. This falsifies our hypothesis that similarly ranked universities have similar strategic plans. These finding were supported by our hierarchical clustering and network analysis, which used the E-LDA topics to cluster the universities. Both the hierarchical clustering and network analysis showed that there was a wide range of external rankings within each cluster, again suggesting no relationship between the similarity of strategy plans and university ranking.

However, we were able to find correlations between some topics and the rankings directly. This shows that while there is no relationship between the similarity of strategic plans and the rankings, there may be a relationship between higher-level concepts in the strategy plans (represented by the E-LDA topics) and external rankings.

The topics *Strategic Improvement* and *Internationalisation* were positively correlated with rankings and the topic *Value to Students* was negatively correlated. This shows some alignment between the strategic planning of high-ranking universities and the criteria for the QS World Rankings, which can be seen in Table 3. The positive correlation of the topic *Internationalisation* reflects the QS indicators *International Faculty Ratio*, *International Student Ratio* and *International Research Network*. On the other hand, the QS ranking does not take the student experience and well-being into account directly. This is reflected in the fact that *Value to Students* is less of a focus of high-ranking universities, which is evidenced by the negative correlation.

The topic *Strategic Improvement* is not an indicator in the QS World Ranking. However, the positive correlation between *Strategic Improvement* and the university's ranking suggests that a focus on strategic

Indicator	Weight
Academic Reputation	30%
Employer Reputation	15%
Faculty Student Ratio	10%
Citations per Faculty	20%
International Faculty Ratio	5%
International Student Ratio	5%
International Research Network	5%
Employment Outcomes	5%
Sustainability	5%

Table 3: QS World Ranking 2024 indicators and their weighting (QS Quacquarelli Symonds Limited, 2023a)

planning is an effective way to achieve high rankings. This supports Dowsett (2020)'s findings that taking certain strategic directions can actively improve the rank of a university.

In our further analysis, we investigated if other characteristics such as geographical location, and the age of universities or fields of study are better reflected in strategy planning. In this analysis, we uncovered patterns in strategic planning and geographical location. The reason for this might be cultural. Breznik and Law (2019) found that culture, politics and economic conditions influence the mission statements of higher education institutions. Similar to mission statements, strategic plans reflect the values and goals of organisations (Hyde-Clarke, 2023) and so it stands to reason that cultural, political and economic factors influence strategic planning as well. To the best of our knowledge, no research has been done to analyse similarities and differences in attitudes towards higher education or policy-making in higher education in different countries across different continents. Future research in this area could help provide more insights into what drives similarities in strategic planning.

Another aspect that could be studied in future research is a closer analysis of the linguistic similarities within and differences between clusters of strategic planning documents. Such an analysis could help identify the linguistic features that drive the formation of these clusters. This in turn can aid in the interpretation of these clusters as well as give more in-depth insights into the strategic priorities of universities.

7 Conclusion

Using a combination of E-LDA, correlation analysis, and hierarchical clustering analysis, we investigated the relationship between the strategic planning of higher education institutions and external university rankings. We also tested the hypothesis that similarly ranked universities have similar strategic plans. Our results falsified this hypothesis, as no significant correlation could be found. However, we were able to identify that higher-ranked universities placed a higher emphasis on certain topics, namely *Internationalisation* and *Strategic Improvement*. We also found that some clusters of universities with similar strategic plans were dominated by universities from the same continent, suggesting a link between strategic planning and geographical location. These insights can help further the understanding of strategic directions that guide institutions and how strategic planning is related to external perceptions of success and excellence.

References

Bayrak, T. (2020). A content analysis of top-ranked universities' mission statements from five global regions. *International Journal of Educational Development 72*, 102130.

Birnbaum, R. and P. J. Edelson (1989). How colleges work: The cybernetics of academic organization and leadership.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Breznik, K. and K. M. Law (2019). What do mission statements reveal about the values of top universities in the world? *International Journal of Organizational Analysis 27*(5), 1362–1375.

Chance, S. and B. T. Williams (2009). Assessing university strategic plans: A tool for consideration. *Educational Planning 18*(1), 38–54.

Daniel, W. W. (1990). *Applied Nonparametric Statistics* (2. ed.). PWS-KENT.

Delamaire, A., M. Juganaru-Mathieu, and M. Beigbeder (2019, May). Correlation between textual similarity and quality of lda topic model results. In *2019 13th International Conference on Research Challenges in Information Science (RCIS)*, pp. 1–6. IEEE.

Deng, D. (2020, 9). DBSCAN Clustering Algorithm based on density. *2020 7th International Forum on Electrical Engineering and Automation (IFEEA)*.

Dowsett, L. (2020). Global university rankings and strategic planning: a case study of australian institutional performance. *Journal of Higher Education Policy and Management 42*(4), 478–494.

Drieger, P. (2013). Network analysis as a method for visual text analytics. *Procedia - Social and Behavioral Sciences*, 4–17. 9th Conference on Applications of Social Network Analysis (ASNA).

Fathi, M. and L. Wilson (2009). Strategic planning in colleges and universities. *Business Renaissance Quarterly 4*(1).

Galbraith, P. L. (1998). When strategic plans are not enough: Challenges in university management. *Systems Dynamice 10*(1–2), 55–84.

Ghosh, S., M. Halappanavar, A. Tumeo, A. Kalyanaraman, H. Lu, D. Chavarrià-Miranda, A. Khan, and A. Gebremedhin (2018). Distributed louvain algorithm for graph community detection. In *2018 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 885–895.

Hall, R. and J. Lulich (2021). University strategic plans: What they say about innovation. *Innovative Higher Education 46*(3), 261–284.

Hazelkorn, E. (2007). The impact of league tables and ranking systems on higher education decision making. *Higher Education Management and Policy 19*(2).

Hou, A. Y., R. Morse, and C.-L. Chiang (2012). An analysis of mobility in global rankings: Making institutional strategic plans and positioning for building world-class universities. *Higher Education Research Development 31*(6), 841–857.

Huang, X., X. Zhang, Y. Ye, S. Deng, and X. Li (2013). A topic detection approach through hierarchical clustering on concept graph. *Applied Mathematics Information Sciences 7*(6), 2285.

Hyde-Clarke, N. (2023). Academic leadership and leadership styles in strategic plans: a study of five top-ranked public universities in south africa. *Cogent Social Sciences 9*(1), 2220533.

Mahardhika, B. and S. Raharja (2023). The importance of strategic planning with modern trends in education. *AL-ISHLAH: Jurnal Pendidikan 15*, 1807–1820.

Omar, M., B. W. On, I. Lee, and G. S. Choi (2015). Lda topics: Representation and evaluation. *Journal of Information Science 41*(5), 662–675.

Pedersen, T. (2001). A decision tree of bigrams is an accurate predictor of word sense. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.

Popat, S. K., P. B. Deshmukh, and V. A. Metre (2017). Hierarchical document clustering based on cosine similarity measure. *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)* (Oct), 153–159.

Priyambodo, P. and E. Hasanah (2021). Strategic planning in increasing quality of education. *Nidhomul Haq: Jurnal Manajemen Pendidikan Islam 6*(1), 109–126.

QS Quacquarelli Symonds Limited (2023a). QS World University Rankings. `https://support.qs.com/hc/en-gb/articles/4405955370898-QS-World-University-Rankings` (Acessed 12. Apr. 2024).

QS Quacquarelli Symonds Limited (2023b). QS world university rankings 2024: Top global universities. `https://www.topuniversities.com/world-university-rankings` (Acessed 27. Feb. 2024).

Runkler, T. A. (2020). *Data Analytics: Models and Algorithms for Intelligent Data Analysis*. Springer Fachmedien.

Shini, G. and V. Srividhya (2023). Implicit aspect-based sentiment analysis for restaurant review using lda topic modeling and ensemble approach. *International Journal of Advanced Technology and Engineering Exploration 10*(102), 554.

Srinivas, S. and S. Rajendran (2019). Topic-based knowledge mining of online student reviews for strategic planning in universities. *Computers Industrial Engineering 128*, 974–984.

Vijaya, V., S. Sharma, and N. Batra (2019, 2). Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*.

Véliz, D. and P. Marshall (2022). Influence of global rankings on strategic planning from the perspective of decision-makers: A case study of a chilean research university. *Higher Education Quarterly 76*(3), 638–652.

Özdem, G. (2011). An analysis of the mission and vision statements on the strategic plans of higher education institutions. *Educational Sciences: Theory Practice 11*(4), 1887–1894.

Authors' Contributions

Name	Contributions
Kim Nolle	**Project role (Chair):** Created and shared the agendas for meetings, led the group meetings and was responsible for communications with the professor. **Reviewed literature:** Hazelkorn (2007); Hou et al. (2012); Breznik and Law (2019); Hall and Lulich (2021) **Methodology design:** Contributed to designing the data collection methodology, the data pre-processing methodology and the data analysis methodology **Implementation / analysis:** Sourced strategy plans for institutions ranked 1–100. Wrote the Python scripts for extracting text from PDFs (extract_text.py) and for calculating the uni- and bi-gram frequencies (counts.py). Conducted the correlation analysis between document similarities and ranking similarities (correlation_analysis.ipynb) **Project report:** Wrote the data analysis methodology section in the mid-term paper. Reviewed the peer feedback. Wrote the correlation analysis sections (3.3.2 and 4.2), the discussion and the conclusion in the final paper. Incorporated peer review feedback into the report. Edited and reviewed the report.
Benika Yadav	**Project role (Recorder):** Recorded and Documented the minutes of all meetings throughout the project **Reviewed literature:** Bayrak (2020); Véliz and Marshall (2022) **Methodology design:** Contributed to designing the data collection methodology, data pre-processing methodology and data analysis methodology **Implementation / analysis:** Sourced strategy plans for institutions ranked 101–200. Contributed to writing the Python script for preprocessing the data, and preprocessed the data using the same. Wrote the Python script for performing clustering and conducted clustering analysis using n-grams to find any correlation between the university strategic plan and other categories. **Project report:** Wrote the data collection section and contributed to writing the data pre-processing section in the mid-term paper. Reviewed and edited the Data Collection, and Preprocessing section. Wrote the Further Analysis section.
Ramya Viswanath	**Project role (Accountant):** Maintained time sheets and recorded individual time spent per week. **Reviewed literature:** Chance and Williams (2009); Mahardhika and Raharja (2023) **Methodology design:** Contributed to designing the data collection methodology the data pre-processing methodology and the data analysis methodology **Implementation / analysis:** Sourced strategy plans for institutions ranked 201–300. Contributed to writing the Python script for preprocessing the data, and preprocessed the data using the same. Contributed to writing the Python script for performing clustering and conducted clustering analysis using n-grams to find any correlation between the university strategic plan and other categories. **Project report:** Wrote the research statement in the mid-term paper. Contributed to Writing the Further Analysis section.
Rupali Singh	**Project role (Verifier):** **Reviewed literature:** Özdem (2011) Ghosh et al. (2018) Drieger (2013) **Methodology design:** Contributed to designing the data collection methodology and the data pre-processing methodology **Implementation / analysis:** Sourced strategy plans for institutions ranked 401–500. Conducted the network Analysis and clustering between the Strategic Planning and the Ranking based on the similarity matrix of the feature vectors. **Project report:** Wrote the introduction in the mid-term paper. Edited the introduction section according to the peer reviews. Also wrote the Network analysis section in the Methodology and Result section.

Name	Contributions
Nachiketh Janpareddy	**Project role (Ambassador):** Attended meetings with other group to expand our knowledge on methodology and analysis. **Reviewed literature:** Galbraith (1998); Srinivas and Rajendran (2019); Huang et al. (2013); Omar et al. (2015); Shini and Srividhya (2023); Delamaire et al. (2019) **Methodology design:** Contributed to designing the data collection methodology and the data pre-processing methodology. Designed and implemented the E-LDA. **Implementation / analysis:** Sourced strategy plans for institutions ranked 501–600. Implemented code for E-LDA Topic analysis and rank correlation with topic distribution **Project report:** Contributed to the abstract, pre-processing and E-LDA topics and correlation analysis of the paper.
Ashiqur Rahman Habeeb Rahuman	**Project role (Monitor):**Maintained handbook to monitor each group member's weekly proceedings and literature reviews relevant to the research. **Reviewed literature:** Dowsett (2020), Drieger (2013) **Methodology design:** Contributed to designing the data collection methodology and data analysis. **Implementation / analysis:** Sourced strategy plans for institutions ranked 301–400. Conducted Network analysis and clustering to represent connection among strategic plans and ranking based on the similarity matrix of the feature vectors. **Project report:** Wrote the section 3.3.4 in methodology and 4.3 in results. Contributed to writing the Introduction and Literature Review. Edited the Literature Review section according to the peer reviews.

Signed: Kim Nolle
(Kim Nolle, 15. April 2024)

Signed: Benika
(Benika Yadav, 15. April 2024)

Signed: V Ramya
(Ramya Viswanath, 15. April 2024)

Signed:
(Ashiqur Rahman Habeeb Rahuman, 15. April 2024)

Signed: Rupali
(Rupali Singh, 15. April 2024)

Signed: Nachiketh J
(Nachiketh Janapareddy, 15. April 2024)