Institute of Architecture of Application Systems

University of Stuttgart
Universitätsstraße 38
D–70569 Stuttgart

Bachelorarbeit

# Ranking Universities based on Academic Reputation and Exploring Potential Correlations to their Success on Social Media

Joshua Schmidt

| | |
|---|---|
| **Course of Study:** | Softwaretechnik |
| **Examiner:** | Prof. Dr. Marco Aiello |
| **Supervisor:** | Prof. Dr. Marco Aiello |
| **Commenced:** | 4th of January 2022 |
| **Completed:** | 4th of July 2022 |

# Abstract

**Purpose -** The idea behind this research is to provide two things: A sophisticated university ranking based on academic reputation for institutions all across the globe, as well as using that ranking to collect relevant data from various Social Media platforms for each university listed.
We hope to find correlations between academic reputation and success on Social Media. If we do, we can use Social Media data to improve upon established ranking systems. The ranking data could then be used to help automate the process of evaluating incoming applications from students all over the world.

**Methodology -** We collect ranking data from four big international rankings for universities in the area of Computer Science: Times Higher Education World University Rankings, QS World University Ranking, ShanghaiRanking Academic Ranking of World Universities and Round University Rankings. The data is normalized and combined into an overall ranking.
To explore potential correlations with success on Social Media, we also collect data concerning each of the universities in our ranking for Facebook, YouTube, Instagram and Twitter and calculate Spearman's rank correlation coefficient for each platform individually.

**Findings -** Analyzing the data, we find positive correlations between how universities rank in the established ranking systems and how well their Social Media accounts are performing. Due to the high number of institutions we look at, we find a high statistical significance as well, leading us to believe that a causal association exists.

**Value -** The overall ranking we create ranks over 1200 universities from all over the globe. The positive correlations between academic reputation and success on Social Media suggests that we can make use of the data collected from both the traditional rankings as well as the Social Media metrics to provide assistance when it comes to evaluating universities in the future. This study also provides a Graphical User Interface to easily work with this data going forward.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

When upcoming students choose the universities they apply to, they have plenty of help in order to make an educated decision; they can simply pick universities close to their home, which according to a study conducted in Ontario, Canada is the most important factor for most [DM06].
Offering one of their desired subjects as well as the reputation of the university are also key criteria on which information is easily obtainable online.
For universities on the other hand, deciding which applicants to accept and which to decline is a much tougher process and currently requires a lot of manual effort, especially when it comes to students applying from other universities all over the globe.
Having to compare different grading systems, various levels of difficulty and matching completed courses with the universities' curriculum are some of the most common problems concerning the applications.
The base of this thesis is a Bachelor Research Project conducted in 2021 at the Institute of Architecture of Application Systems, in which a group of students (including the author of this thesis) already looked into collecting some ranking data on universities and combining them into one general ranking to use as a baseline. Said project includes rankings from three major ranking websites.
The goal is to expand on that work and collect more data about as many universities as possible to provide a sophisticated ranking in order to help automate that process, potentially saving many hours of manual work in the future.
Furthermore we want to take a look at how the more traditional ranking approaches for universities compare to their success on various big Social Media platforms (namely Facebook, Twitter, YouTube and Instagram).
If we find positive correlations between the online rankings for universities and some Social Media metrics, we can make use of the collected Social Media data to improve the decision making process even further.
Also, another advantage of Social Media as a way to judge an institution is the fact that this data is openly available at any time, without relying on annual third party publications, as it is the case with the ranking websites. We can scrape data on Social Media on-demand whenever we need to, with almost any institution being present on most platforms.

An additional inquiry for this thesis comes from the colleagues over at the Master Program INFOTECH of the University of Stuttgart. They manage their own list of universities and want the data to be comparable, which is why a matching between the INFOTECH data and the data obtained over the course of this project will be part of this work.
For the sake of accessibility and reusability, all parts of this work should also be implemented in the form of a Graphical User Interface (GUI).

# 2 State of the Art

## 2.1 Related Work

Universities are very present on Social Media. However the effort they put into creating a successful presence on different platforms seems to be "lacking". At least that is what an article published by Voss and Kumar in 2013 suggests [VK13].
In said article the authors try to understand how universities use Social Media and which value it provides to them. It reaches the conclusion that Social Media accounts are becoming the prevalent form of communication and should be utilized by universities to engage their audiences.

In a study released very recently in May 2022, Kuzmin et al. [KFRZ22] find "a noticeable relationship between the size of the audience in a social network and the dynamics of the brand's revenue".
Nowadays Social Media networks certainly have found their way into almost every students life. This is why we want to consider them as a relevant metric not just for brands, but also for universities. By evaluating the success of universities on Social Media in 2022, especially in comparison to more traditional ranking methods, we will hopefully be able to answer the question whether universities know how to make use of the incredibly popular communication channels.

When it comes to creating rankings for universities based on their academic reputation, there seems to be a decent acceptance of the established ranking systems out there, as there appear to be little to no efforts at creating new or similar ranking concepts.
However, a paper published in 2018 by McCoy et al. [MNW18] looks at ways to create such a university ranking without making use of the traditional ranking metrics commonly used in big university rankings online.
Instead, they use Twitter data to rank universities and later try to compare the results to the more established approaches.
The metric in use is what is called the University Twitter Engagement (UTE) score, which is the total number of primary and secondary followers affiliated with a university.
The research reaches the conclusion that there certainly is a significant, positive rank correlation between said UTE and how universities rank in more traditional, aggregate reputation rankings.

The study is limited to universities in the United States only, which is something we definitely want to expand. However, the approach of comparing Social Media data to established rankings is promising.
We certainly still want to include data from traditional rankings though, provided the methodology is transparent.

## 2.2 Previous Work - Bachelor Research Project

As previously mentioned, a Bachelor Research Project conducted during summer 2021 by students at the Institute of Architecture of Application Systems acts as the groundwork for this project.
The Research Project provides a first attempt at combining three big ranking websites to create an overall ranking. The rankings used are the Times Higher Education World University Rankings, QS World University Rankings and ShanghaiRanking Academic Ranking of World Universities.
All of these will also be considered for this project, as well as potentially including additional candidates.
One of the problems encountered over the course of the Research Project is the matching of university names, as they often appear with slightly different naming conventions across different rankings.
It also touches on the idea of comparing the ranking data to Social Media statistics, but the chosen method of getting said data is by collecting Social Media links from each university's own website, which only works for universities with a well managed web presence.
We will expand on this project, try to include even more ranking data and aim for a definite answer on the question whether there are correlations to success on Social Media.

# 3 Finding Data Sources

The initial question to be answered before starting data acquisition is which sources to use. Both for the traditional university rankings as well as Social Media platforms we need to define what a good data source should offer and then find sources which provide it.

## 3.1 University Ranking Websites

The first major criteria for a ranking to be applicable is size. In order to rank as many universities as possible and have a higher chance of any given university appearing in multiple rankings for a more meaningful result, we want the rankings we use as a baseline to rank at least 500 institutions.
For the sake of comparability of the collected data, we only look at rankings which include universities from all over the world, not just from one specific country. Otherwise we would not be able to tell how they compare to each other when combining the rankings.
Furthermore, since we aim to provide this ranking to faculties in the fields of Informatics, Computer Science and Technology, we specifically look for websites with dedicated rankings for these fields. Another important criteria for including a website in this combined ranking is transparency in methodology.

With those criteria in mind, the rankings we include are the Times Higher Education (THE) World University Rankings, the QS World University Rankings, the ShanghaiRanking Academic Ranking of World Universities (ARWU) and the Round University Ranking (RUR).

### 3.1.1 Times Higher Education World University Rankings

THE is a magazine owned by Inflexion Pvt. Equity Partners LLP providing coverage on topics related to higher education. They also maintain a ranking of universities all over

the world, both in a general fashion as well as subject-specific. The one applicable for this project is the "World University Ranking by subject: computer science" [THE22]. While their general ranking looks at over 1200 institutions, the computer science one still ranks over 600.
They also provide a transparent overview over their ranking indicators and weightings. A final score is calculated using 13 ranking indicators clustered into the five overarching categories "Teaching", "Research", "Citations", "International outlook" and "Industry income" [THEM22]. Weightings are being applied as follows:

- Teaching - 30%
- Research - 30%
- Citations - 27.5%
- International outlook - 7.5%
- Industry income - 5%

The THE ranking is available for the years 2017 to 2022 at the time of this research.

### 3.1.2 Quacquarelli Symonds (QS) World University Rankings

QS is a dedicated website managed by Quacquarelli Symonds Limited with the purpose of providing various rankings on universities across the globe.
Unfiltered, the site ranks over 1200 universities. For the subject "Computer Science and Information Technology", which is the one we use, the number of institutions still exceeds 600 [QS22].
QS also provides insight into the methodology for creating the ranking. It consists of five main components according to their documentation [QSM22]:

- Academic reputation
- Employer reputation
- Research citations per paper
- H-index (a way of measuring both the productivity and impact of the published work of a scientist or scholar)
- International Research Network (by broad-faculty area)

We are able to collect data from the QS ranking for the years 2019 to 2022.

### 3.1.3 ShanghaiRanking Academic Ranking of World Universities

ShanghaiRanking's ARWU is published by the Shanghai Ranking Consultancy and features universities from all over the globe.
Their generic ranking containts roughly 1000 institutions while the one for "Computer Science & Engineering" looks at about 500 [ARWU21].
As far as methodology goes, ARWU defines five key indicators on their website [ARWUM21]:

- Q1 (number of papers published by an institution in an Academic Subject)
- CNCI (ratio of citations of papers published to the average citations of papers in the same category)
- IC (International collaboration)
- Top (number of papers published in Top Journals)
- Award (total number of the staff of an institution winning a significant award)

As of conducting this study, a 2022 ARWU ranking is not available yet, so we can only collect data from 2017 to 2021.

### 3.1.4 Round University Ranking

The RUR Ranking Agency publishes the Round University Ranking in partnership with the international company Clarivate Analytics annually, ranking over 1000 universities worldwide. For "Technical Sciences", their Subject Rankings list over 700 institutions, making this the biggest ranking included in this project [RR21].
The Round University Ranking is designed to evaluate productivity of research and makes use of the following five criteria according to their methodology page [RRM21]:

- Normalized citation impact
- Citation per papers
- Papers per academic and research staff
- International research reputation
- Share of research publications written in international co-authorship

The RUR data can be scraped for the years 2014 to 2021.

This leaves us with the following data sources for our ranking:

| Ranking Name | Ranking specific for Computer Science? | Number of institutions included | Available years |
|---|---|---|---|
| THE | Yes | 600+ | 2017-2022 |
| ARWU | Yes | 600+ | 2017-2021 |
| QS | Yes | 500 | 2019-2022 |
| RUR | Yes | 700+ | 2014-2021 |

**Table 3.1:** University Rankings included - Overview

## 3.2 Social Media Platforms

When it comes to Social Media, there are a lot of potential platforms to look at. We want to include the most used ones to get the most meaningful results. Also, we do prioritize platforms that allow institutions to create public accounts that allow users to interact in a way we can quantify. That excludes private messenger applications like WhatsApp, WeChat or Snapchat.
A report published by DataReportal in partnership with We Are Social and Hootsuite and reported by Statista (see Fig. 3.1) helps shed some light on the question which networks are the most popular at the time of conducting this research [SM22].

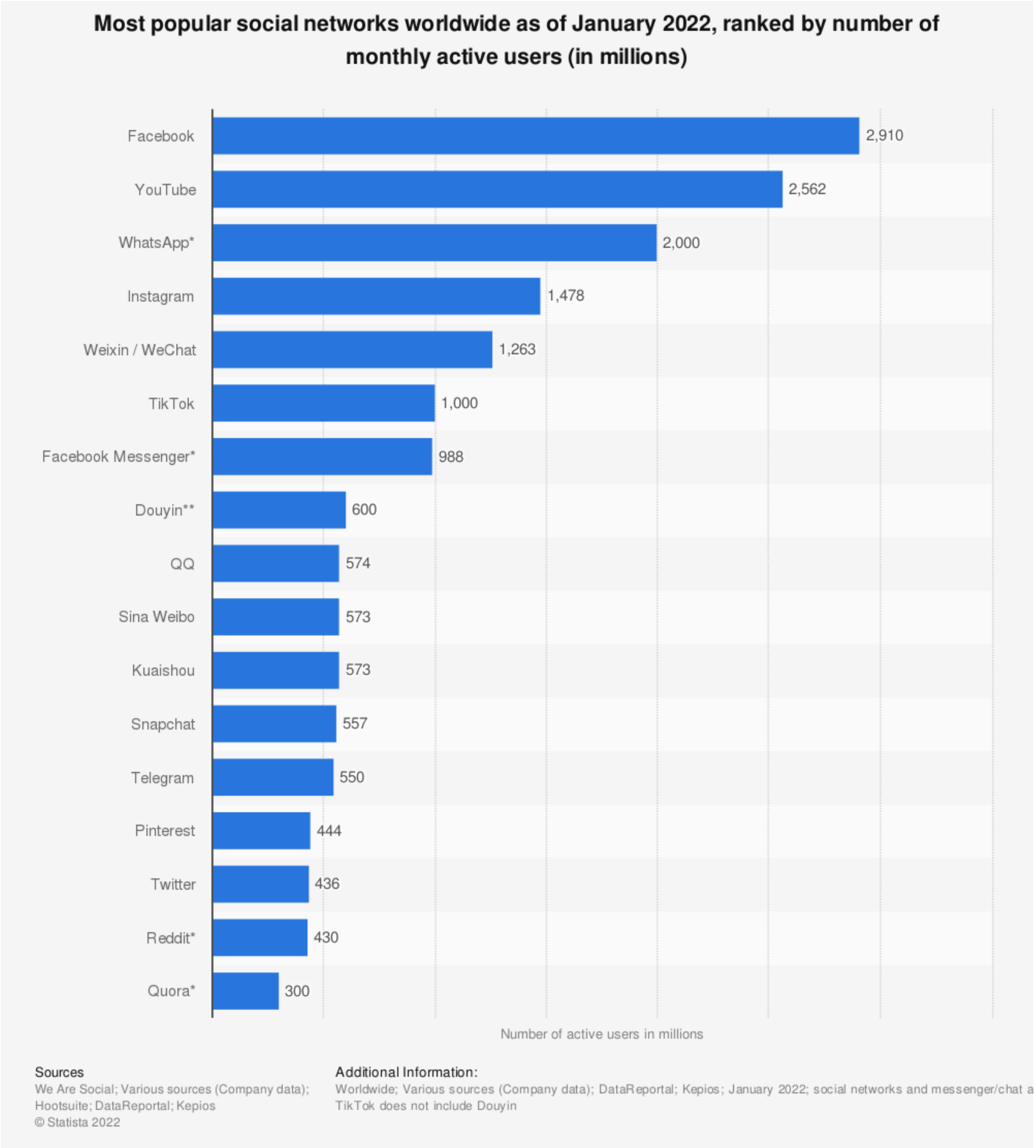



**Figure 3.1:** Most popular Social Networks (January 2022) - Source: DataReportal, reported by Statista [SM22]

According to that study, Facebook and YouTube are the number one and two most popular social networks worldwide as of January 2022, with Facebook pulling almost 3 billion users onto their site in just a single month, closely followed by YouTube's

impressive monthly user count of over 2.5 billion. Because of this, those two are the most relevant for this project.
Third on the list provided by Statista is WhatsApp, however as mentioned previously, messenger applications are not interesting to us in the scope of this work.
The next big candidate we want to include is Instagram. With almost 1.5 billion monthly active users in January 2022 the platform ranks fourth.
At least in the western hemisphere, Twitter is also among the top Social Media platforms (ranking 15th worldwide), still coming in at about 436 million active users, certainly enough reason to take a closer look at the network. It seems as to eastern countries use different Social Media platforms for the most part, so we are going to focus on the ones we have easy access to in Europe.
Ignoring messenger applications and platforms almost exclusively used in Asia, this brings us to these Social Media platforms: YouTube, Facebook, Twitter and Instagram.

| Platform | Monthly active users |
|---|---|
| Facebook | 2.91 billion |
| YouTube | 2.56 billion |
| Instagram | 1.48 billion |
| Twitter | 436 million |

**Table 3.2:** Social Media platforms included - Overview

## 3.3 Defining the Statistical Approach

Before collecting any data from any of the designated Social Media platforms, we need to define which statistics are relevant to us when it comes to each individual platform.

### 3.3.1 Facebook - Follower

When it comes to Facebook, the solution is rather straight forward. The most commonly used practice of interacting with the public page of a brand, company or university is by liking and following.
Liking a page on Facebook is seen as a way of supporting it, while following ensures you see a page's posts on your timeline. The two usually go hand in hand with only small gaps between the two statistics on most Facebook pages. The number of users who currently like and follow a Facebook page is openly displayed and can be collected.

### 3.3.2 YouTube - Subscribers

In the case of YouTube, the desired metric is the amount of Subscribers a university's channel has. A Subscription on YouTube is free and makes sure the user is informed about any new video a content creator uploads. It is publicly available on any channel page.

### 3.3.3 Instagram - Follower

On Instagram, users follow an account if they like the content being provided. The number of Followers is by far the best way of measuring the success on Instagram, as it is also displayed on top of any Instagram page.

### 3.3.4 Twitter - Follower

Similar to Instagram, the number one way of showing support to a Twitter account is by following. Twitter constructs the feed it shows to its users based on their follows combined with some suggestions. On Twitter, we can easily extract the amount of Followers an account has using web scraping.

### 3.3.5 Preparing the Analysis

The main point of interest when it comes to the statistical analysis after collecting the data will be exploring potential correlations. If we find that the ranking we create correlates with the success of universities on Social Media, that can help increase the amount of useful data for future decision making. We would no longer just make use of the traditional ranking systems, but also keep track of how universities do on Social Media to judge them better.
In order to also find potential differences between the Social Media platforms, we will look at each platform and it's potential correlations separately.

# 4 Realisation

Programming language of choice is Python. All code files, as well as any input or output files are stored in the repository provided by the Institute of Architecture of Application Systems.
To ensure easy usability in the future we keep the project structure as modular as possible. Each step along the way has its own directory with each individual step producing the input files for the subsequent one. This way it is possible to repeat either the entire process or only specific steps, which comes in handy when providing the GUI application.
Since most of our desired data is not available through Application Programming Interfaces (APIs), we collect it ourselves via web scraping, using Selenium. Selenium is an open source module for Python and offers various tools for controlling web browsers as well as performing browser automation.
The actual data acquisition is split into multiple steps:

- Obtaining Ranking Data
- Data Matching
- Generating our own ranking
- Obtaining Social Media statistics

## 4.1 Obtaining Ranking Data

As previously discussed, we prefer to use a web scraping approach to obtain the data in this research. For some rankings, the data is available through third-party APIs, however we prefer to not be reliant on those if not necessary, especially since they are not even provided by the ranking publisher directly.
Since all the data is available on the websites, collecting it ourselves can be done quite easily by using the Selenium library.
We provide a separate file containing the web scraping functionality for each of the ranking pages. The process to scrape the data we want from each individual ranking is quite similar however. The program opens the corresponding link with Selenium and

then extracts the data from the HTML source code.
Depending on the layout of the page, we can either directly obtain everything or need to perform certain actions like clicking the button to move to the next page until we are done.
The fastest out of all the websites to scrape is by far the Round Ranking, because it presents the entire ranking on the first load of the page. All other rankings are designed so that we need to click through pages to get to all the data, which makes the program run a little longer, but still not by a huge amount.
All steps of the scraping process were executed on a regular home computer with an AMD Ryzen 5 3600 6-Core Processor (3.59 GHz), which is what the run-time estimates in Table 4.1 are based on.

| Ranking Name | Corresponding File | Estimated Run-Time |
|---|---|---|
| THE | times_scraper.py | <10 minutes |
| ARWU | arwu_scraper.py | <15 minutes |
| QS | qs_scraper.py | <15 minutes |
| RUR | round_scraper.py | <5 minutes |

**Table 4.1:** Web Scraping Process for Ranking Data - Run-Time Estimate

Before using the ranking data obtained in this step as the input files for the overall ranking, we do some cleanup to make sure the data is easily comparable.
Most importantly, we need to figure out a way to unify the names, as not every ranking uses the same naming conventions for the institutions. Of course, we want to combine entries from two separate rankings describing the same university but with a slightly different name into one entry in our ranking.

## 4.2 Data Matching

One of the main issues when it comes to combining the collected ranking data is the fact that university names for the same institution are not consistent throughout the different rankings.
In an attempt to match as many of the names as possible, we use a centralized registry to match each name. We attempt to create a unique identifier this way.
The registry we choose here is the Research Organization Registry (ROR). "ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organization in the world." [ROR19]

The project aims to provide exactly what we need - unique identifiers for as many institutions all across the globe as possible. The registry does not only include universities, but also all other kinds of organizations, however the universities are what is useful to us. Most importantly, it comes with a very easy-to-use API, allowing us to query any university name, returning it's name in the registry, which we refer to as the "ROR name"; our unique identifier.
We add the filter "Education" to reduce the number of hits and possible mismatches and look at the first ten returned results to make sure we don't mistakenly accept just the first result while the correct match was returned later in the list for some reason.
A sample GET request for the University of Stuttgart looks like this:

https://api.ror.org/organizations?query=University%of%Stuttgart&filter=types:Education

and leads to the response in Fig. 4.1.


```
time_taken:            16
▼ items:
  ▼ 0:
      id:              "https://ror.org/04vnq7t77"
      name:            "University of Stuttgart"
      email_address:   null
      ip_addresses:    []
      established:     1829
    ▶ types:           […]
    ▶ relationships:   […]
    ▶ addresses:       […]
    ▶ links:           […]
      aliases:         []
      acronyms:        []
      status:          "active"
    ▶ wikipedia_url:   "http://en.wikipedia.org/…/University_of_Stuttgart"
    ▶ labels:          […]
    ▶ country:         {…}
    ▶ external_ids:    {…}
```


**Figure 4.1:** Example Response from the ROR API for the Query "University of Stuttgart"

The API offers a lot of info on every entry, however for our purpose we only extract the returned values for name and id.
By running the output files from each ranking websites against the ROR API, we retrieve

the ROR name for each university which we can use as our unique identifier to make sure we match the right universities when combining the rankings.

## 4.3 Generating our own Ranking

To ensure comparability when combining the collected ranking data into one overall ranking, we need to unify the scores. As mentioned previously, not all websites use the same scale for their scores.
Also, some of the rankings do not provide a numerical value all the way through. QS for example only provides an exact score until rank 50, afterwards universities are still ranked but without a publicly available value indicating their score. Because of that we will determine a university's score in our ranking by measuring its relative position in all the rankings it appears in.
Hence our overall score will be the top percentage a university appears in across all the rankings. For example, the Massachusetts Institute of Technology comes in at the highest rank in our ranking for 2021 with an overall score of 99.84%, indicating it ranks in the top 0.16% across all four rankings on average.
Also, to give some perspective on how meaningful each individual entry in the ranking is, we add a confidence score indicating in how many underlying rankings the university appeared in, ranging from a minimum value of 1 to a maximum of 4, since we collect data from four rankings.
We do collect data for every year if possible, however for 2022 not every ranking is available as of conducting this study unfortunately, so the 2021 data sets create the most recent, complete ranking.
Since we have the data for all other available years, we do create the combined rankings for them as well, but the missing rankings lead to lower confidence scores in those. For 2021, we are able to provide a ranking consisting of over 1200 universities. Table 4.2 shows the Top 5 of that ranking to demonstrate.

| Rank | ROR name | Overall Score[%] | Times Score | ARWU Score | QS Score | Round Score | Conf. |
|---|---|---|---|---|---|---|---|
| 0 | Massachusetts Institute of Technology | 99.84 | 91.1 | 355.3 | 93.7 | 97.3 | 4.0 |
| 1 | Stanford University | 99.46 | 91.7 | 315.6 | 93.1 | 94.2 | 4.0 |
| 2 | University of Oxford | 99.4 | 92.9 | 234.2 | 89.1 | 100.0 | 4.0 |
| 3 | Carnegie Mellon University | 99.39 | 89.8 | 279.9 | 93.0 | 95.2 | 4.0 |
| 4 | ETH Zurich | 99.19 | 90.5 | 264.0 | 85.6 | 97.1 | 4.0 |

**Table 4.2:** Top 5 of the Combined Ranking for 2021

## 4.4 Obtaining Social Media Statistics

With our own ranking now generated, we can begin collecting Social Media data for all universities listed. We do so using the same tools we already use for scraping ranking data. Our main helper is once again the Selenium library. We prefer scraping the data ourselves wherever possible here as well, in order to not rely on any third party applications. The process varies depending on the Social Media platform.

### 4.4.1 Facebook

Facebook does not require the user to be logged in in order to see most of its content. However, using the search function to find the page of a university does require a Login. Since we do not want this project to be attached to any personal account, we need to circumvent this.
Luckily, searching for a university's ROR name in combination with the term "Facebook" on Google yields the desired result very consistently, provided the university is present on the platform. An alternative to this method would be using the university website to find Social Links. This is the approach presented in the Bachelor Research Project, however this depends a lot on the layout of each individual website.
After finding the page link via Google, it is simply a matter of visiting the page with Selenium, finding the element containing the amount of Followers and storing the data.

Currently, the success rate when looking for Facebook data this way is about 84%, meaning we find Facebook pages for about 1000 universities.

### 4.4.2 YouTube

To obtain the number of Subscribers of a university, we do not need a YouTube account. That data can be looked up by anyone without logging into the platform.
This means we can use the YouTube search bar. Our scraper simply searches for the universities ROR name on the page, filters by channels and checks whether a matching result is presented.
If so, it takes the number of Subscribers and stores it alongside the channel name. We find YouTube data for roughly 80% of universities in our ranking, leading to well over 900 values.

### 4.4.3 Instagram

Instagram is the strictest out of all the platforms we look at in terms of privacy. Without logging in, we are unable to obtain any data. Even attempts of logging in with a dedicated user account for the purpose of this project only works a set amount of times until Instagram detects non-human behaviour and stops the scraper from logging in.
Since we are checking over 1200 university entries we need to come up with a different approach.
Google is once again the answer; when searching for a ROR name in combination with "Instagram", Google will display the amount of Followers in the preview text of the search results, making a login obsolete. This approach provides us with the data we need without even having to visit Instagram. It results in data for about 85% of institutions.

### 4.4.4 Twitter

Collecting data from Twitter is much easier than Instagram. Similar to YouTube, Twitter allows us to view pages without being logged in and use the built-in search function as well. We make use of this and simply search for the ROR name on Twitter, filter by accounts, find the matching Twitter page, visit it and store the number of Followers together with the page name. In this case, we find data for over 950 universities on Twitter, which equals about 80% of our ranking.

### 4.4.5 Estimated Run-Times

We iterate over every university in our combined ranking for every platform. Since our ranking includes over 1200 institutions, this process may take a very long time. The estimated run-time for the Social Media data collection process is significantly higher than for scraping the ranking data previously, as shown in Table 4.3.

| Platform | Corresponding File | Estimated Run-Time (in minutes) |
|---|---|---|
| Facebook | collect_facebook_data.py | 180+ |
| YouTube | collect_youtube_data.py | 120+ |
| Instagram | collect_instagram_data.py | 120+ |
| Twitter | collect_twitter_data.py | 120+ |

**Table 4.3:** Web Scraping Process for Social Media Data - Run-Time Estimates

This concludes the data acquisition process. We now have data sets from four major ranking websites, resulting in a combined ranking of over 1200 universities. For each of those we have access to the amount of Facebook Followers, YouTube Subscribers, Instagram Followers and Twitter Followers, provided they are present on each platform.

## 4.5 INFOTECH Ranking

The colleagues over at the Master Program INFOTECH of the University of Stuttgart maintain their own ranking list of universities. In order to draw a comparison and provide additional information to their ranking, we match their data against our ranking results. We once again use the ROR registry as a way to uniquely identify similar entries in the two rankings by running the entire INFOTECH list against the ROR API. After that we use the ROR names obtained to match every university appearing in the INFOTECH list to entries in our ranking.
The full list can be reviewed in the GUI application so that the institute can make use of the results.

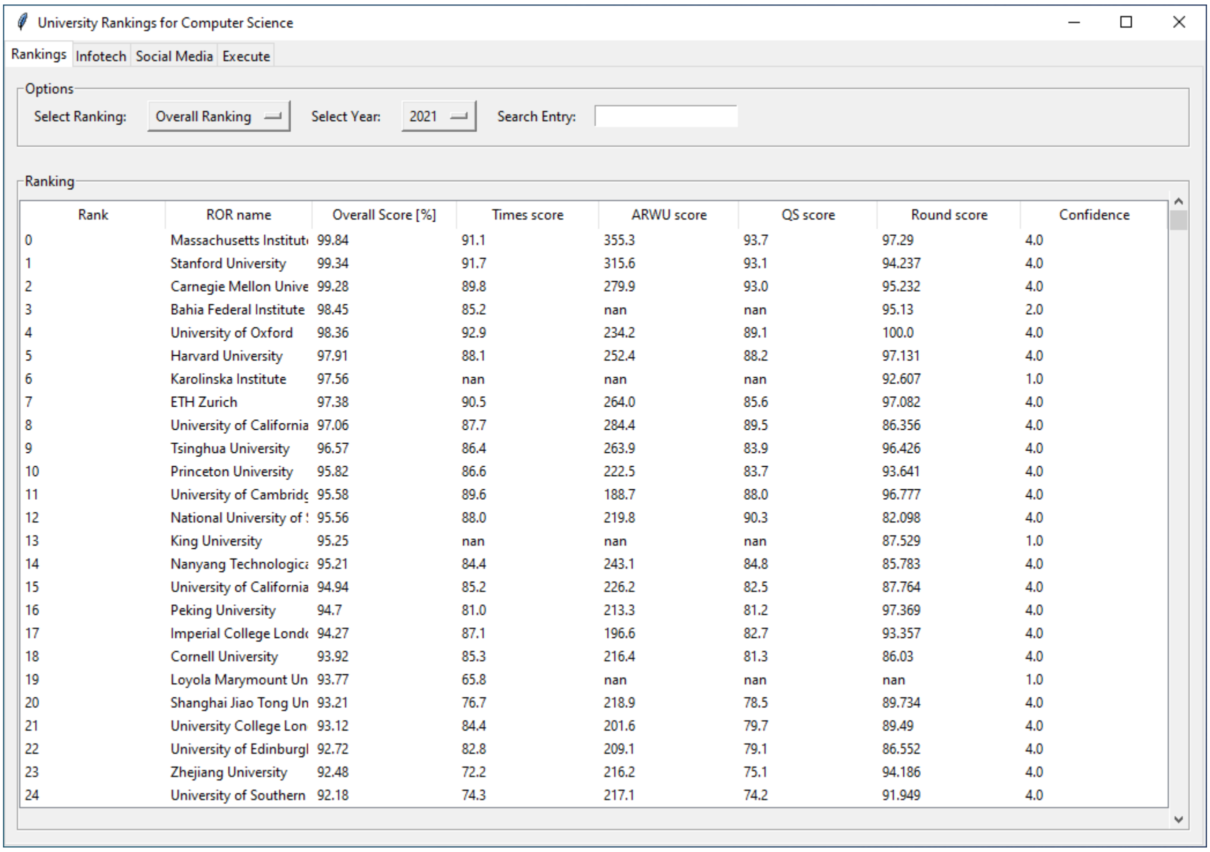
## 4.6 Graphical User Interface

As one of the main goals of this work is to help facilitate the process when evaluating incoming applications from students all over the world, we need to provide some easy

way of accessing and using the tools and data we put together.
All the functionality created during this research is accessible through an application providing a GUI. It comes with the following features:

- View and filter through collected ranking data as well as the combined ranking
- Access scraped Social Media statistics
- Show and filter the matched INFOTECH list
- Execute and repeat each step of the process separately if necessary

A sample screenshot of the application is shown in Fig. 4.2. Demo videos on how to use each of the features are provided in the repository as well.




| Rank | ROR name | Overall Score [%] | Times score | ARWU score | QS score | Round score | Confidence |
|---|---|---|---|---|---|---|---|
| 0 | Massachusetts Institut | 99.84 | 91.1 | 355.3 | 93.7 | 97.29 | 4.0 |
| 1 | Stanford University | 99.34 | 91.7 | 315.6 | 93.1 | 94.237 | 4.0 |
| 2 | Carnegie Mellon Unive | 99.28 | 89.8 | 279.9 | 93.0 | 95.232 | 4.0 |
| 3 | Bahia Federal Institute | 98.45 | 85.2 | nan | nan | 95.13 | 2.0 |
| 4 | University of Oxford | 98.36 | 92.9 | 234.2 | 89.1 | 100.0 | 4.0 |
| 5 | Harvard University | 97.91 | 88.1 | 252.4 | 88.2 | 97.131 | 4.0 |
| 6 | Karolinska Institute | 97.56 | nan | nan | nan | 92.607 | 1.0 |
| 7 | ETH Zurich | 97.38 | 90.5 | 264.0 | 85.6 | 97.082 | 4.0 |
| 8 | University of California | 97.06 | 87.7 | 284.4 | 89.5 | 86.356 | 4.0 |
| 9 | Tsinghua University | 96.57 | 86.4 | 263.9 | 83.9 | 96.426 | 4.0 |
| 10 | Princeton University | 95.82 | 86.6 | 222.5 | 83.7 | 93.641 | 4.0 |
| 11 | University of Cambridg | 95.58 | 89.6 | 188.7 | 88.0 | 96.777 | 4.0 |
| 12 | National University of ! | 95.56 | 88.0 | 219.8 | 90.3 | 82.098 | 4.0 |
| 13 | King University | 95.25 | nan | nan | nan | 87.529 | 1.0 |
| 14 | Nanyang Technologica | 95.21 | 84.4 | 243.1 | 84.8 | 85.783 | 4.0 |
| 15 | University of California | 94.94 | 85.2 | 226.2 | 82.5 | 87.764 | 4.0 |
| 16 | Peking University | 94.7 | 81.0 | 213.3 | 81.2 | 97.369 | 4.0 |
| 17 | Imperial College Londo | 94.27 | 87.1 | 196.6 | 82.7 | 93.357 | 4.0 |
| 18 | Cornell University | 93.92 | 85.3 | 216.4 | 81.3 | 86.03 | 4.0 |
| 19 | Loyola Marymount Un | 93.77 | 65.8 | nan | nan | nan | 1.0 |
| 20 | Shanghai Jiao Tong Un | 93.21 | 76.7 | 218.9 | 78.5 | 89.734 | 4.0 |
| 21 | University College Lon | 93.12 | 84.4 | 201.6 | 79.7 | 89.49 | 4.0 |
| 22 | University of Edinburgl | 92.72 | 82.8 | 209.1 | 79.1 | 86.552 | 4.0 |
| 23 | Zhejiang University | 92.48 | 72.2 | 216.2 | 75.1 | 94.186 | 4.0 |
| 24 | University of Southern | 92.18 | 74.3 | 217.1 | 74.2 | 91.949 | 4.0 |

**Figure 4.2:** Graphical User Interface - Screenshot

This concludes the data acquisition portion of the project. What is left is analyzing the data we obtained and check for potential correlations with statistical significance.

# 5 Analysis

For the kind of data we have, Spearman's rank correlation coefficient is the most fitting way of analyzing potential correlations. It determines how well a monotonic function is able to describe the relationship between two variables. The Spearman correlation coefficient is used to check for similar or identical behaviour of ranked variables.
First, we will determine the Spearman coefficient between our generated ranking and each individual platform, describe the observations briefly and finally try to reach conclusions by comparing the results.
A Spearman coefficent between 0 and 1 implies a positive correlation, while a result between 0 and -1 suggests a negative association, with 1 or -1 being a perfect positive or negative correlation. Exactly 0 would imply no connection at all.
We also calculate the so called p-value, a value describing the significance of the test. As with most correlation coefficients, we look to reject or accept a Hypothesis $H_0$ being that there is no correlation between our sets of variables. $H_1$ is the opposite, meaning that there is indeed a statistically significant correlation
We can reject $H_0$ if we find p-values smaller than 0.05, in which case we will believe $H_1$. The lower our p-value, the stronger that assumption will be. Otherwise, we will be accepting $H_0$ and assume no correlation.
We collect data for multiple years, but two of the four rankings we scrape data from do not have their annual ranking published for 2022 as of the time conducting this research. Since the Social Media data we collect is the current state, we want to compare it to the newest, complete ranking data we have which is 2021.
We will briefly look at the results for each Social Media platform individually before drawing conclusions about general correlations between success on Social Media and academic reputation.

## 5.1 Facebook

The first Social Media platform we look at is Facebook. A scatter plot of the complete set of variable is shown in Fig. 5.1. We definitely observe a trend toward higher scores towards the top of the ranking, possibly suggesting correlation.

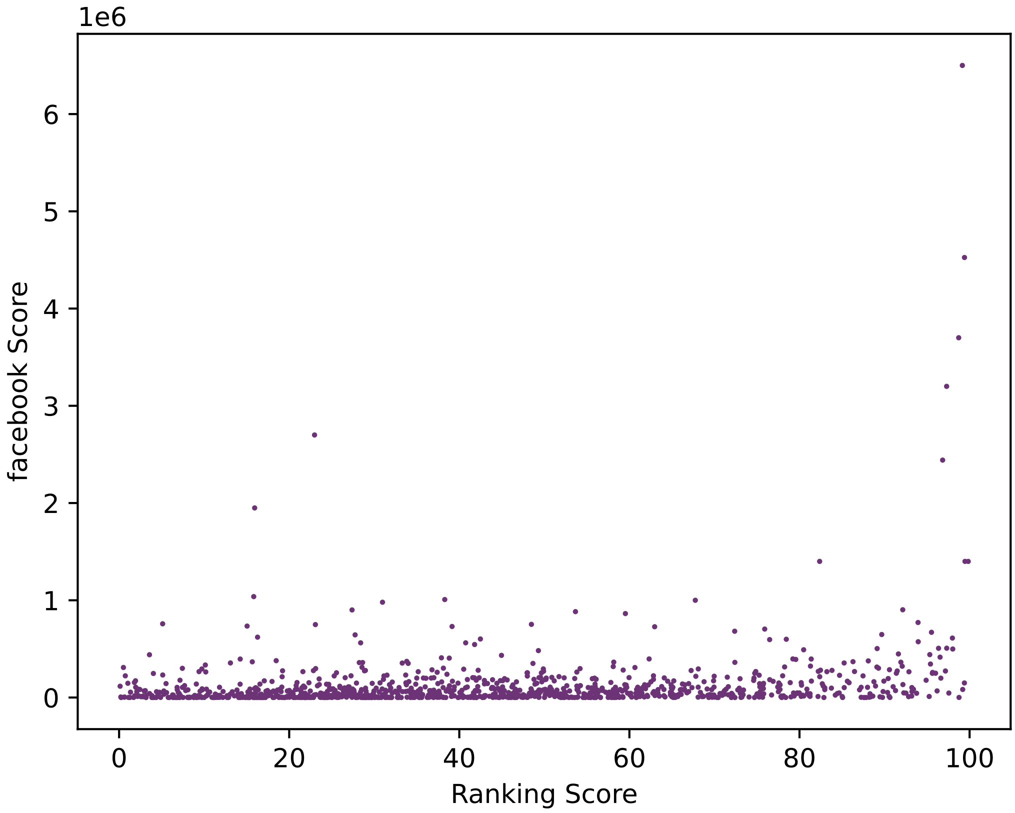



**Figure 5.1:** Scatter Plot - Facebook

The assumption is somewhat confirmed as we calculate a slightly positive correlation of 0.26, displayed in Fig. 5.2. Since we found data on the platform for over 1000 universities, the p-value is very low (<0.001), indicating that the correlation is statistically significant.

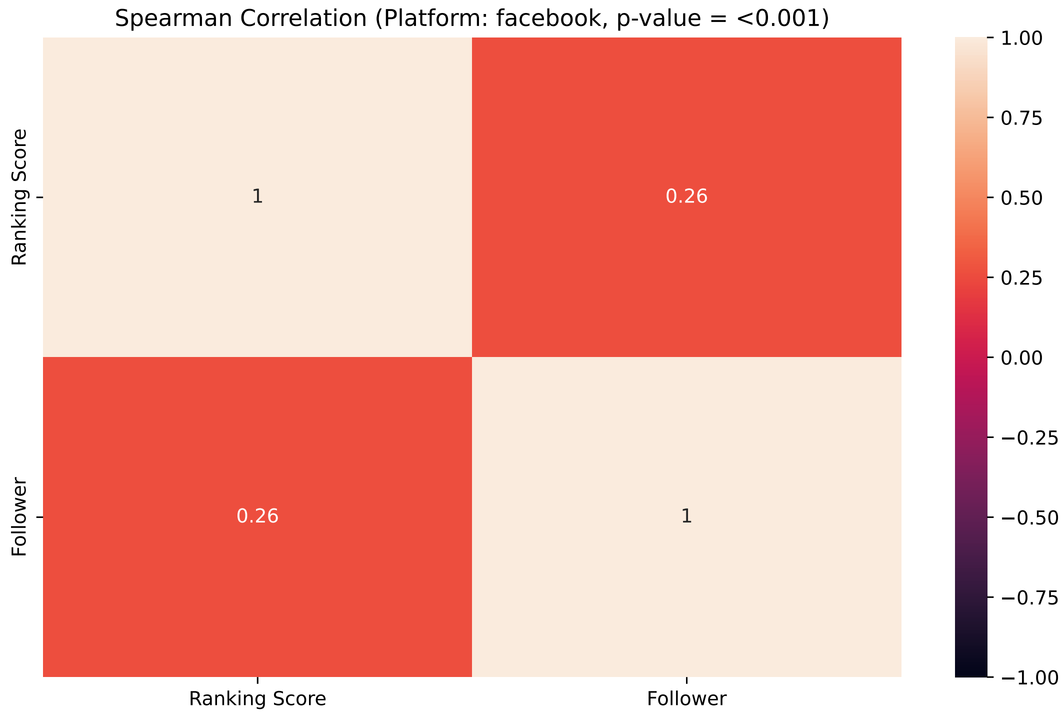



**Figure 5.2:** Spearman Correlation - Facebook

## 5.2 YouTube

In the case of YouTube, we are seeking correlations between our overall ranking and almost 1000 collected data points for YouTube subscribers. Pictured below in Fig. 5.3 is the plot of all the value pairs for YouTube. We do see some extremely high scores towards the top of the ranking once again.

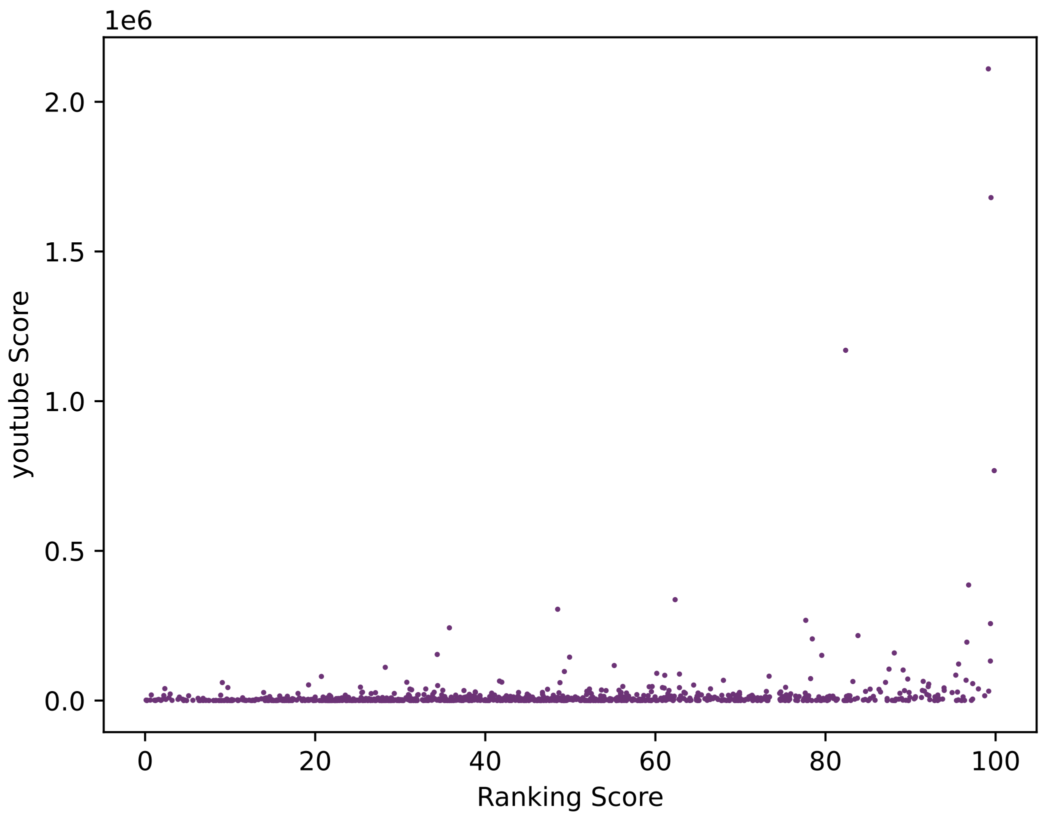



**Figure 5.3:** Scatter Plot - YouTube

We find a Spearman Correlation coefficient of 0.29 with a p-value of <0.001 (see Fig. 5.4). The significance is not surprising once again, as we are looking at over 900 value pairs. Once again, we observe a slightly positive correlation between the amount of people that subscribe to or follow a university's account on YouTube and how that university does in our ranking.

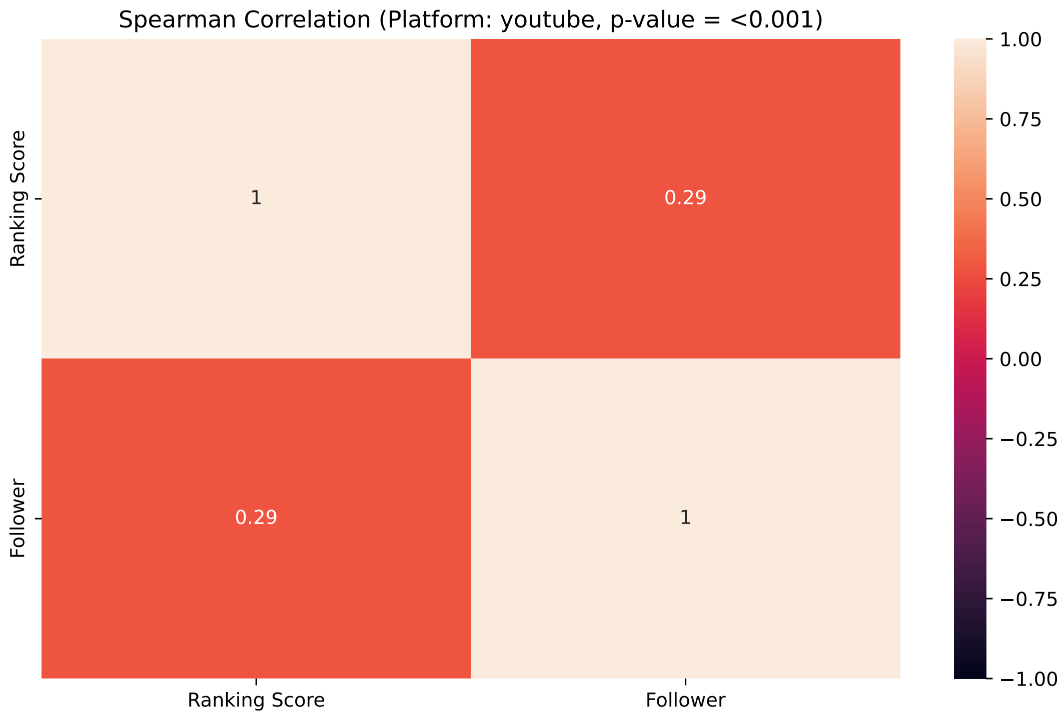



**Figure 5.4:** Spearman Correlation - YouTube

## 5.3 Instagram

We found even more data points for Instagram than for YouTube, this time over 1000 accounts we could match to the universities in our ranking, see the Instagram Plot in Fig. 5.5 for reference.

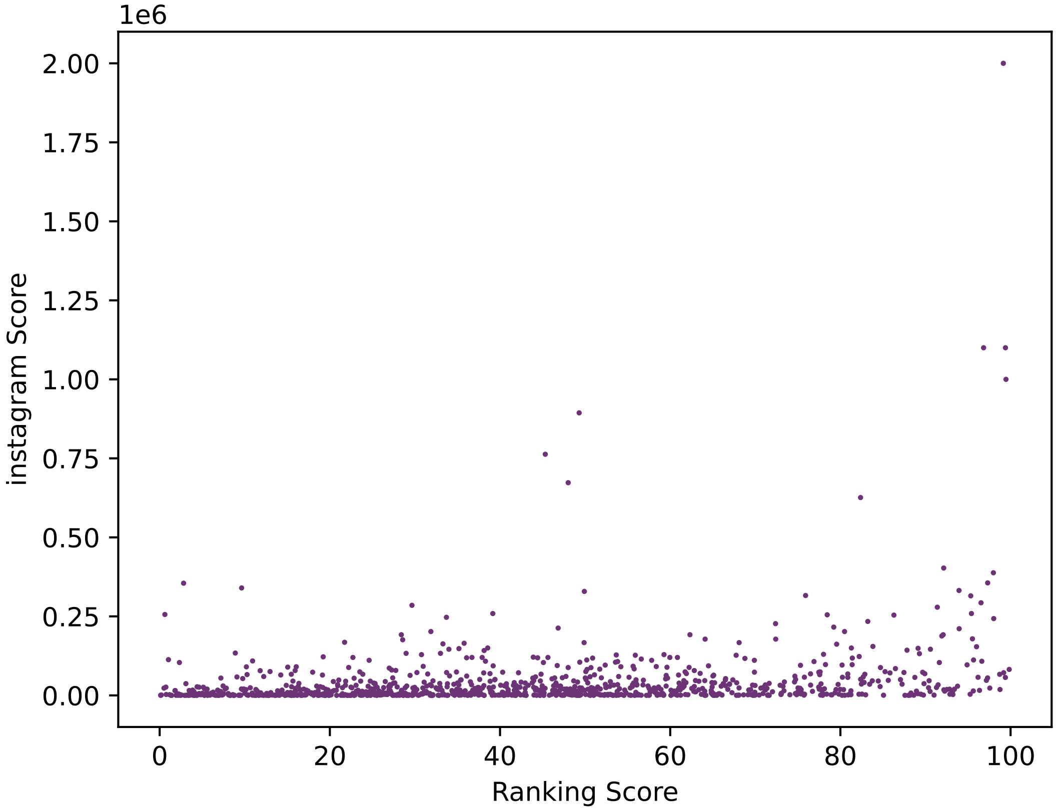



**Figure 5.5:** Scatter Plot - Instagram

The Spearman Correlation coefficient and p-value can be seen in Fig. 5.6 below. With a Spearman coefficient of 0.32, we observe another positive correlation, even slightly stronger than the previous ones, even if still relatively weak.

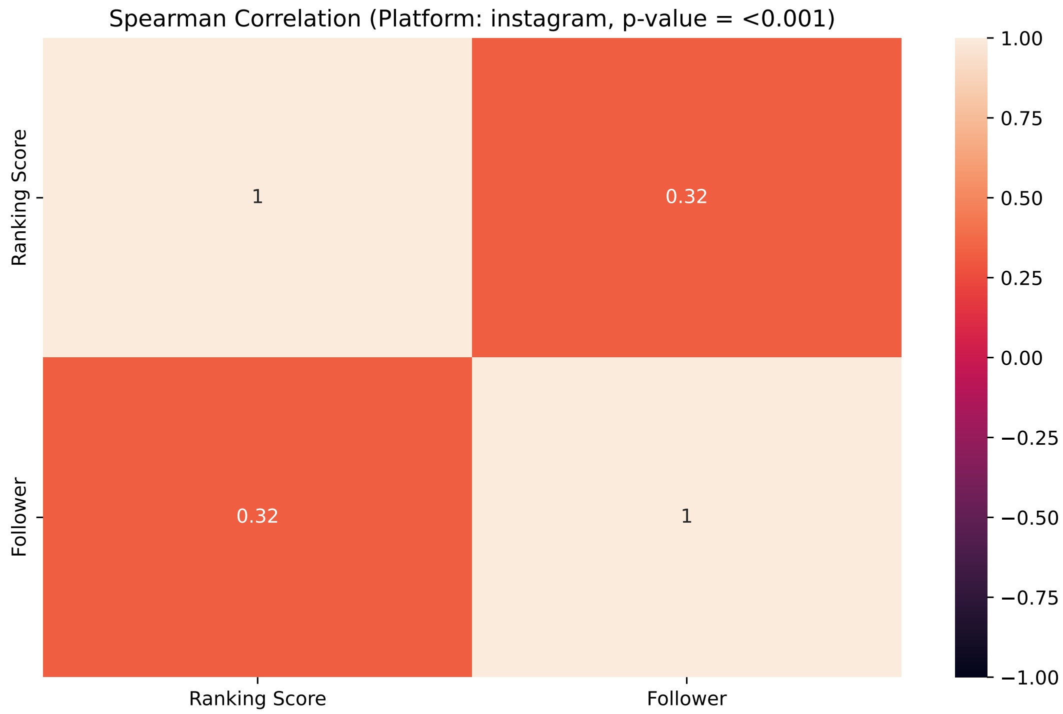



**Figure 5.6:** Spearman Correlation - Instagram

## 5.4 Twitter

For Twitter, the amount of universities we found data for is close to 1000 once again. The Plot is shown in Fig. 5.7. Universities scoring about 80% and upwards in our ranking seem to have higher success on the platform.

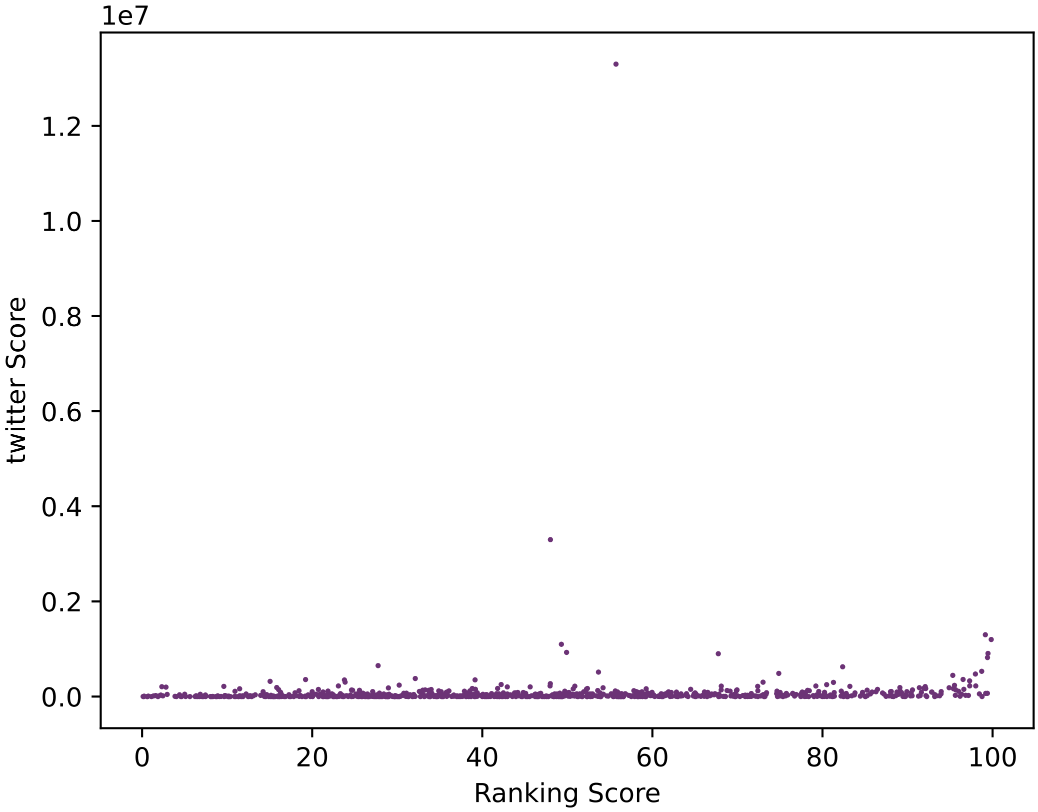



**Figure 5.7:** Scatter Plot - Twitter

We observe a correlation coefficient of 0.29, showing, once again, a slightly positive correlation with a p-value of <0.001 (as seen in Fig. 5.8). After all the other platforms, this comes to no surprise and is in line with the previous results.

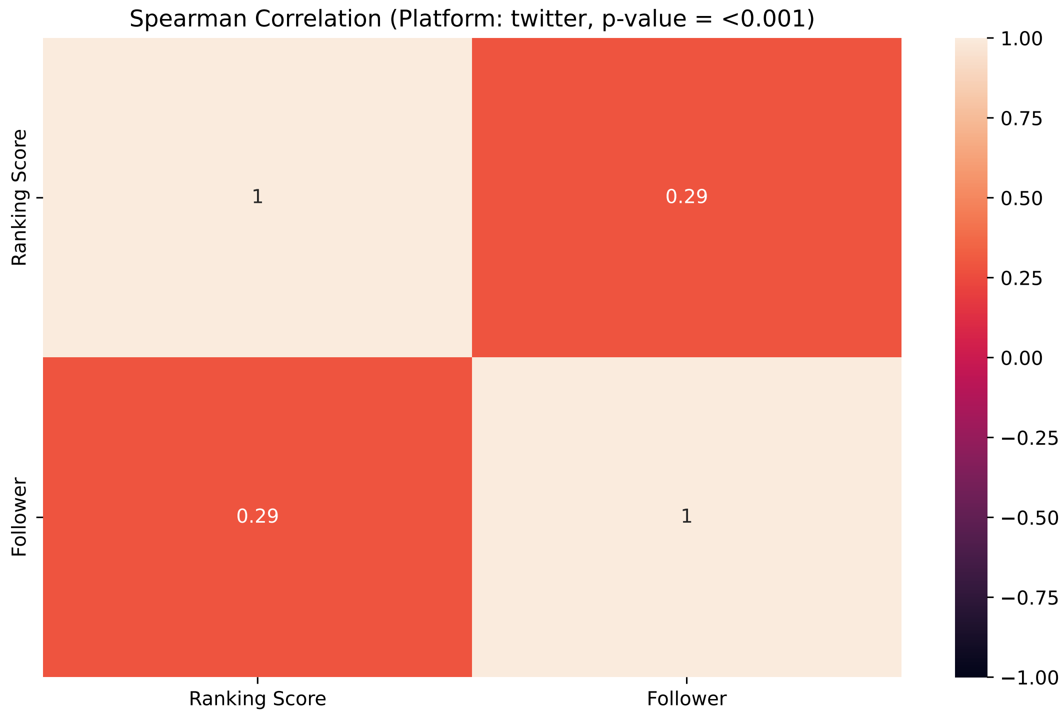



**Figure 5.8:** Spearman Correlation - Twitter

With all of the individual platforms out of the way, we have a very good amount of data to make statements about potential correlations, especially by looking at recurring trends that are consistent between all the platforms.

## 5.5 Interpretation

The data paints a very clear picture here. Although we only reach correlation values of up to 0.32, all of the platforms provide positive correlations with very low p-values highly suggesting a statistically significant, causal relationship.
The sheer amount of universities we look at makes it very unlikely these observations happen by pure chance across multiple Social Media platforms. As a result, we conclude with quite certainty that there is a positive correlation between success on Social Media and academic reputation in the more traditional rankings for universities.
As to why the correlation is not higher, we consider one major factor to be a relatively low activity level of universities on Social Media, especially smaller ones. While almost all of them have accounts on most of the platforms, a large portion of them is not maintaining their online presence incredibly well, as discovered by Voss et al. [VK13] as well.
Also, as mentioned previously, we look at Social Media networks primarily used by western countries, yet our ranking also includes many universities from the eastern hemisphere. Lower activity levels by those institutions are to be expected when it comes to platforms less relevant to them.
Another reason might be that universities with good academic reputation are more incentivized to manage an active Social Media account than lesser known institutions. We will take a small look at the efforts that two universities put into their Social Media presence and how they pay off in a brief case study. The two institutions we compare will be the University of Stuttgart, sitting in the middle of our ranking at rank 419, as well as the best ranking university, the Massachusetts Institute of Technology (MIT).

## 5.6 Case Study: Comparing the University of Stuttgart and the MIT

We start by comparing the raw numbers for these two institutions in our collected data, as well as adding some quantification as to how much effort they put into their Social Media accounts. We will look at the amount of Tweets and YouTube videos and compare them to the success on these two platforms.
The comparison regarding YouTube can be found in Table 5.1 and the one regarding Twitter in Table 5.2.

| University | YouTube Videos | Videos per Year | YouTube Subscribers | Subscribers per Video |
|---|---|---|---|---|
| Uni Stuttgart | 519 | 51 | 23.500 | 45 |
| MIT | 633 | 48 | 777.000 | 1227 |

**Table 5.1:** Comparison of University of Stuttgart's and the MIT's success on YouTube

| University | Tweets | Tweets per Year | Twitter Follower | Follower per Tweet |
|---|---|---|---|---|
| Uni Stuttgart | 11.388 | 876 | 11.802 | 1 |
| MIT | 21.518 | 1537 | 1.2 million | 56 |

**Table 5.2:** Comparison of University of Stuttgart's and the MIT's success on Twitter

The data hear speaks a pretty clear language. Even though we can not determine the quality of the YouTube videos or Tweets posted by the institutions, the quantity compared to the success on the platforms is very different for the two universities.
To make sure this is not caused by significant differences in time and effort these two put into managing their presences on the platforms, we also look at the time frame. We do so by calculating the amount of Videos per year as well as Tweets per year.
The MIT's success on both platforms is overwhelming in comparison to the University of Stuttgart's. It received about 1227 Subscribers per video and 56 Follower per Tweet, while the University of Stuttgart only scores about 45 Subscribers per video or 1 Follower per Tweet.
The effort both institutions put into YouTube seems to be about similar, while the MIT posts about twice as many Tweets each year, which still stands in no proportion to the 56 Followers per Tweet opposed to the 1 Follower per Tweet for Stuttgart.

There definitely seems to be an advantage for renowned universities when it comes to Social Media, as the difference in quality (if any) probably does not justify the huge difference in success here. This makes a lot of sense considering all of our scatter plots analyzed earlier show a rise in success for universities with scores of about 80% and more.
Of course, these are only two institutions out of all the ones we ranked, but we do see a trend here of a renowned university seeing much higher success on Social Media then a lesser known one. We already observe a positive correlation between the traditional ranking systems and success on Social Media, which is supported even more by these observations; a university ranking high seems to correlate with a successful Social Media presence, even without putting significantly more effort in.
Now, we can't say for sure whether a good reputation causes success on Social Media or if they behave the other way around. It seems like a bit of both is the case, which could be a reason the effects are quite strong and disproportionate towards the top of the ranking.
Either way a correlation seems very likely, meaning we can suggest including Social Media statistics in future decision-making processes when it comes to judging universities. In an attempt to improve it's reputation, a university should also probably invest into their Social Media presence and activity, as it promotes the institution and seems to be a great indicator to check whether the reputation is actually growing or not.

# 6 Conclusions and Outlook

With this research we provide a feasible way of combining four of the biggest university rankings world wide into one ranking, including over 1200 institutions from the field of Computer Science. This ranking can be used for faculties to judge incoming applications and better understand the differences in quality of other institutions.
Also, we collect a lot of data about each of these institutions when it comes to their presence on various Social Media platforms.
Overall, we definitely see evidence strongly suggesting there to be a positive correlation between success on Social Media and academic reputation.
Notably, the Social Media platforms we collect data for are all more relevant to the western hemisphere. For eastern universities, we could probably find more meaningful data by looking at the biggest platforms used by their students. Of course that data is harder to obtain due to language and privacy barriers, however it is worth noting for future work.
A great way to build on the findings even further as well would be to look into both the quality and the quantity of the efforts all of these institutions put into their Social Media presences in general.
The small case study suggests that there certainly is a correlation between a university's reputation and the ratio between input and output when it comes to Social Media.
Looking into ways to judge the quality of postings, Tweets or videos on different platforms would be a good way of confirming or denying this suspicion.
For now, we strongly suggest making use of all the data openly available on the web to automate parts of the application process at universities. Academic reputation as well as Social Media metrics are great ways of judging a university and could save a lot of manual work.

# Bibliography


[ARWU21] ShanghaiRanking. *2021 Global Ranking of Academic Subjects - Computer Science Engineering*. 2021. URL: https://www.shanghairanking.com/rankings/gras/2021/RS0210 (cit. on p. 17).

[ARWUM21] ShanghaiRanking. *ShanghaiRanking's Global Ranking of Academic Subjects Methodology 2021*. 2021. URL: https://www.shanghairanking.com/methodology/gras/2021 (cit. on p. 17).

[DM06] T. Drewes, C. Michael. "How do students choose a university?: an analysis of applications to universities in Ontario, Canada." In: *Research in Higher Education* 47.1 (Nov. 2006), pp. 781–800. DOI: https://doi.org/10.1007/s11162-006-9015-6 (cit. on p. 11).

[KFRZ22] E. A. Kuzmin, M. V. Faminskaya, E. A. Rodionova, E. R. Zinatullina. "Digital Effect: Relationship between Brand Media Image and Company Performance." In: *Jurnal Ilmiah Peuradeum* 10.2 (June 2022), pp. 491–508. DOI: https://doi.org/10.26811/peuradeun.v10i2.785 (cit. on p. 13).

[MNW18] C. G. McCoy, M. L. Nelson, M. C. Weigle. "Mining the Web to approximate university rankings." In: *Information Discovery and Delivery* 46.3 (2018), pp. 173–183. DOI: https://doi.org/10.1108/IDD-05-2018-0014 (cit. on p. 13).

[QS22] QS TopUniversities. *QS World University Rankings by Subject 2022: Computer Science and Information Systems*. 2022. URL: https://www.topuniversities.com/university-rankings/university-subject-rankings/2022/computer-science-information-systems (cit. on p. 16).

[QSM22] QS TopUniversities. *Subject Rankings Methodology*. 2022. URL: https://www.topuniversities.com/subject-rankings/methodology (cit. on p. 16).

[ROR19] California Digital Library. *Research Organization Registry*. 2019. URL: https://ror.org/ (cit. on p. 24).

[RR21] Round University Ranking. *Subject Rankings - Technical Sciences*. 2021. URL: https://roundranking.com/ranking/subject-rankings.html#world-technical-2021 (cit. on p. 17).

[RRM21] Round University Ranking. *RUR Academic World University Rankings Methodology*. 2021. URL: https://roundranking.com/methodology/academic-ranking.html (cit. on p. 17).

[SM22] Statista. *Most popular social networks worldwide as of January 2022, ranked by number of monthly active users*. 2022. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ (cit. on pp. 18, 19).

[THE22] Times Higher Education. *World University Rankings 2022 by subject: computer science*. 2022. URL: https://www.timeshighereducation.com/world-university-rankings/2022/subject-ranking/computer-science#!/page/0/length/25/sort_by/rank/sort_order/asc/cols/stats (cit. on p. 16).

[THEM22] Times Higher Education. *World University Rankings 2022 by subject: computer science methodology*. 2022. URL: https://www.timeshighereducation.com/world-university-rankings-2022-subject-computer-science-methodology (cit. on p. 16).

[VK13] K. A. Voss, A. Kumar. "The value of social media: are universities successfully engaging their audience?" In: *Journal of Applied Research in Higher Education* 5.2 (Sept. 2013), pp. 156–172. DOI: https://doi.org/10.1108/JARHE-11-2012-0060 (cit. on pp. 13, 40).

All links were last followed on July 4, 2022.