Project proposal

Ram Shiri, Amit Gitin, Lior Skudowitz, Nir Yemini

1. Introduction

The Israeli higher education landscape is shaped by a complex interplay between cooperation and competition. Universities in Israel often join forces in research initiatives, joint academic programs, and national infrastructure projects. At the same time, they compete fiercely over student enrollment, government funding and prestige in international rankings. This duality, is a strategic characteristic of the academic field in Israel.

The central question guiding this study is - How does Israeli universities communicate competition and cooperation, and how does that play out on their websites?

The research investigates how cooperation and competition shapes the strategic positioning of academic institutions. We focus solely on universities, as universities operate within a competitive framework emphasizing prestige, research output, and global visibility — unlike colleges, which often serve regional and vocational roles.

Our aim is to identify whether the language used in official university publications reflects cooperative or competitive strategies and nature.

2. Data

Our preliminary data set consists of 70 collected news articles from the official websites of major Israeli universities, including: Ben-Gurion University, Tel Aviv University, the Technion, the Hebrew University, and Bar-Ilan University. Each article was annotated by the research team based on its tone and content.

For each article, the following variables were recorded:

institution: Name of the publishing university

date: Date of publication

title: Article headline

url: Link to the original article

tone: Overall tone of the article (Cooperative, Competitive, Neutral) sets through what the article mentions.

mentions_cooperation: Whether the article explicitly mentions cooperation.

mentions_competition: Whether the article explicitly mentions competition.

cooperation_type: If applicable, the type of cooperation (Domestic, International, etc.).

thematic_tags: Semicolon-separated content tags.

length_words: Estimated number of words in the article.

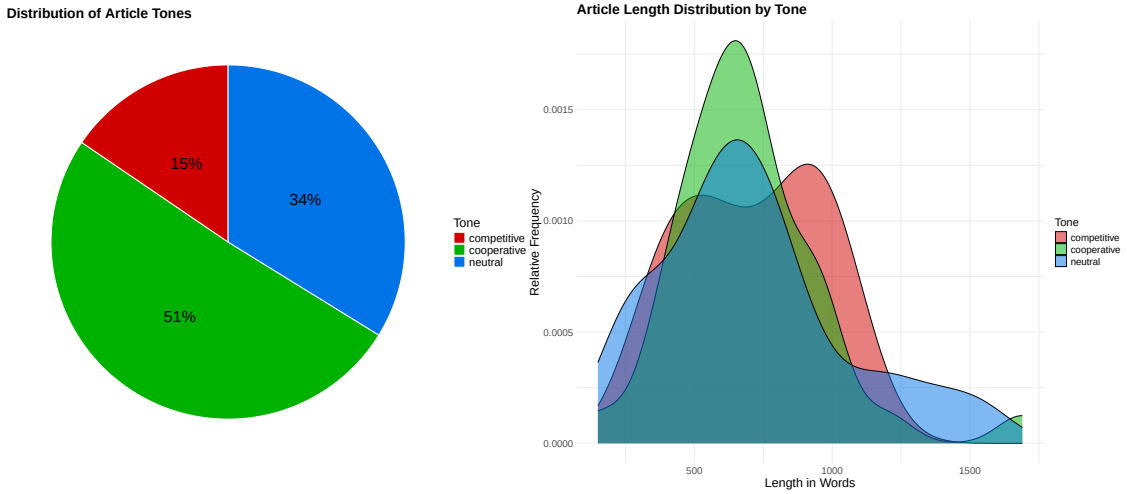
shanghai_ranked: Whether the institution is ranked in the Shanghai ranking (at the time of publishing the article).

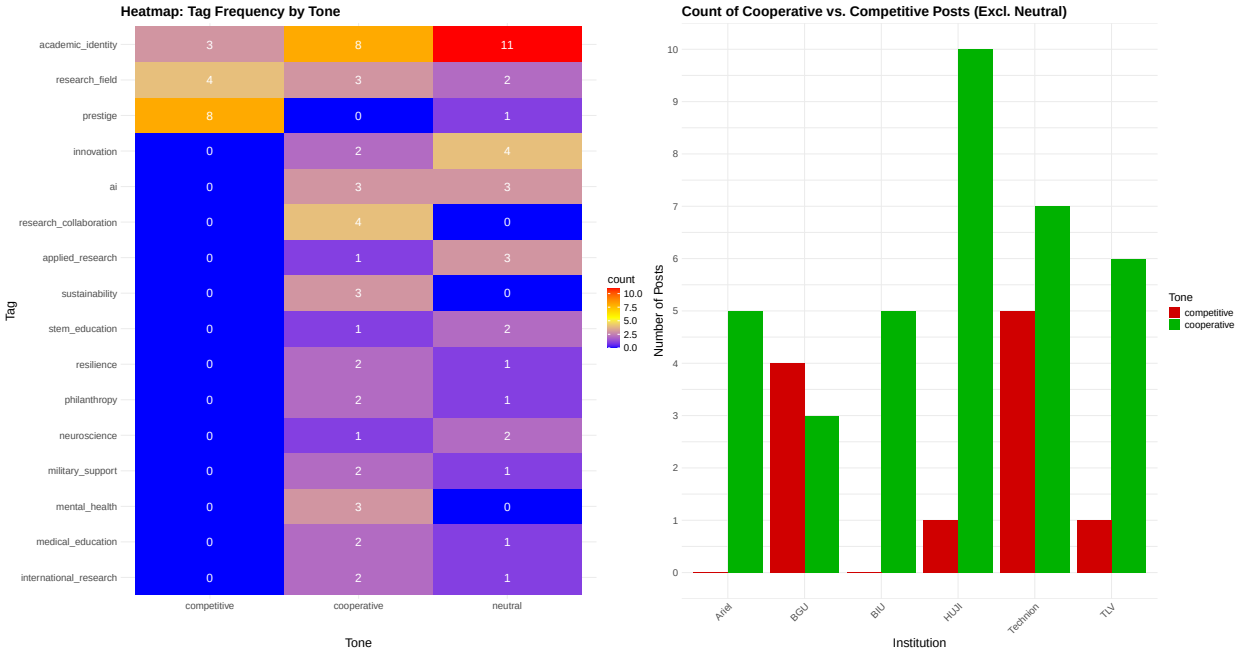
All data has been saved as a CSV file in the /data folder, accompanied by a README.md file that describes each field. Text content has been cleaned, and categorical fields have been normalized. This structured dataset enables us to run visualizations and exploratory analyses.

3. Preliminary results

As part of this study, we conducted an initial qualitative review of the official news sections published on the websites of several Israeli universities. The goal of this stage was to identify emerging patterns in how universities communicate cooperation and competition and position themselves in the higher education landscape. A preliminary analysis reveals that universities emphasize cooperation over competition.

shanghai rank for Israel's Institution: https://www.shanghairanking.com/institution?name=&r=Israel





4. Data analysis plan

Variables:

Outcome (Y): Tone of the article (categorical), or binary indicators of cooperation/competition mentions

Predictors (X): Institution, Shanghai ranking, mentions_competition, mentions_cooperation, cooperation_type, thematic_tags, length_words

Group Comparisons:

Ranked vs. unranked universities (Shanghai). Institutions with high vs. low competition frequency. Institutions with high vs. low cooperation frequency. different Institutions agains in comparison to one another.

Methods:

Data visualization using ggplot2

Correlation and regression analyses to identify relationships between cooperation framing and institutional prestige/metrics

Results to Extract:

Does Shanghai ranking correlate with tone or cooperation mentions?

Are certain tags more associated with competitive or cooperative tones?

creation of a new parameter called "coopetition" which includes both cooperation and competition as factors and see if there is correlations to any of the tags we created especially the Shanghai rank.

Teamwork:

Nir Yemini: Provided the research foundation and theoretical framing; ensures project alignment with dissertation scope.

Ram Shiri: Led manual data collection, writes code for visualizations, contributes to content writing, and planning of regression. strategy

Amit Gitin: Assisting with article collection and tag annotation; supports visualization and helps refine exploratory analysis.

Lior Skudowitz: Focuses on data validation and cleaning, contributes to visualization and writing, and supports regression modeling and result interpretation.

Appendix

Data README

```
# README - Dataset Description

This dataset contains manually collected articles from Israeli university websites, annotated for tone a

## Columns Description

- **institution**: Name of the university that published the article.
- **date**: Date of publication of the article.
- **title**: Title of the article.
- **url**: URL of the article.
- **tone**: Overall tone of the article: Collaborative, Competitive, or Neutral.
- **mentions_cooperation**: Boolean. Whether the article mentions cooperation.
- **mentions_competition**: Boolean. Whether the article mentions competition.
- **cooperation_type**: Type of cooperation mentioned, if any (e.g., Domestic, International).
- **thematic_tags**: Semicolon-separated thematic tags indicating article content (e.g., research_field
- **length_words**: Approximate length of the article in number of words.
- **shanghai_ranked**: Boolean. Whether the publishing university is ranked in the Shanghai ranking.
```

Source code

```
library(knitr)
library(tidyverse)
library(broom)
library(htmltools)
library(ggplot2)
library(patchwork)

opts_chunk$set(echo=FALSE) # hide source code in the document
  df <- read_csv("data/cooperation_articles.csv")
  df <- df %>%
    mutate(
      mentions_cooperation = tolower(as.character(mentions_cooperation)) == "true",
      mentions_competition = tolower(as.character(mentions_competition)) == "true"
    )

#prepare the table
coopetition_table <- df %>%
  filter(!is.na(tone)) %>%
  mutate(tone = tolower(trimws(tone))) %>%
  filter(tone %in% c("collaborative", "competitive", "neutral")) %>%
  count(institution, tone) %>%
  pivot_wider(names_from = tone, values_from = n, values_fill = 0) %>%
  mutate(total = collaborative + competitive + neutral)


tone_colors <- c(
  "cooperative" = "#00B200",
  "competitive" = "#D00000",
```

```
  "neutral"     = "#0073E6"
)



# Pie chart
pie_chart <- df %>%
  count(tone) %>%
  mutate(tone = tolower(tone),
         tone = ifelse(tone == "collaborative", "cooperative", tone),  # rename tone
         perc = n / sum(n),
         label = paste0(round(perc * 100), "%")) %>%
  ggplot(aes(x = "", y = perc, fill = tone)) +
  geom_col(width = 1, color = "white") +
  coord_polar(theta = "y") +
  labs(title = "Distribution of Article Tones", fill = "Tone") +
  geom_text(aes(label = label), position = position_stack(vjust = 0.5), size = 8) +
  scale_fill_manual(values = tone_colors) +
  theme_void() +
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    legend.title = element_text(size = 18),
    legend.text = element_text(size = 16)
  )

# Density plot
length_density <- df %>%
  mutate(tone = tolower(tone),
         tone = ifelse(tone == "collaborative", "cooperative", tone)) %>%
  ggplot(aes(x = length_words, fill = tone)) +
  geom_density(alpha = 0.5) +
  labs(title = "Article Length Distribution by Tone",
       x = "Length in Words", y = "Relative Frequency", fill = "Tone") +
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    axis.title = element_text(size = 18),
    axis.text = element_text(size = 14),
    legend.title = element_text(size = 16),
    legend.text = element_text(size = 14)
  )

# Heatmap
tag_heatmap_df <- df %>%
  filter(!is.na(thematic_tags), !is.na(tone)) %>%
  separate_rows(thematic_tags, sep = ";") %>%
  mutate(thematic_tags = str_trim(tolower(thematic_tags)),
         tone = tolower(tone),
         tone = ifelse(tone == "collaborative", "cooperative", tone)) %>%
  count(thematic_tags, tone) %>%
  pivot_wider(names_from = tone, values_from = n, values_fill = 0) %>%
  filter(rowSums(across(-thematic_tags)) >= 3)
```

```r
tag_heatmap_long <- tag_heatmap_df %>%
  pivot_longer(-thematic_tags, names_to = "tone", values_to = "count")

tag_heatmap <- ggplot(tag_heatmap_long, aes(x = tone, y = reorder(thematic_tags, count))) +
  geom_tile(aes(fill = count)) +
  geom_text(aes(label = count), color = "white", size = 6) +
  scale_fill_gradientn(colors = c("blue", "yellow", "red")) +
  labs(title = "Heatmap: Tag Frequency by Tone", x = "Tone", y = "Tag") +
  guides(fill = guide_colorbar(title.theme = element_text(size = 16),
                               label.theme = element_text(size = 14))) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    axis.title = element_text(size = 18),
    axis.text = element_text(size = 14)
  )

# Bar plot
coopetition_absolute <- coopetition_table %>%
  select(institution, collaborative, competitive) %>%
  filter((collaborative + competitive) > 0) %>%
  pivot_longer(cols = c(collaborative, competitive),
               names_to = "tone", values_to = "count") %>%
  mutate(tone = ifelse(tone == "collaborative", "cooperative", tone))

coopetition_bar <- ggplot(coopetition_absolute, aes(x = institution, y = count, fill = tone)) +
  geom_col(position = "dodge") +
  scale_y_continuous(breaks = 0:max(coopetition_absolute$count)) +
  labs(title = "Count of Cooperative vs. Competitive Posts (Excl. Neutral)",
       x = "Institution", y = "Number of Posts", fill = "Tone") +
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 20, face = "bold"),
    axis.title = element_text(size = 18),
    axis.text = element_text(size = 14),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.title = element_text(size = 16),
    legend.text = element_text(size = 14)
  )

# Combine plots
(pie_chart + length_density) /
  plot_spacer() /
  (tag_heatmap + coopetition_bar) +
  plot_layout(heights = c(1, 0.2, 1.3))

cat(readLines('data/README.md'), sep = '\n')
```