Final Report: Coopetition in Israeli Academia

Ram Shiri, Amit Gitin, Lior Skudowitz, Nir Yemini

1. Introduction

The Israeli higher education landscape is shaped by a complex interplay of cooperation and competition. Universities in Israel often collaborate on research initiatives, joint academic programs, and national infrastructure projects. Simultaneously, they compete vigorously for student enrollment, government funding, and international prestige. This dual dynamic is a defining strategic feature of the academic field in Israel. This study explores a central question: **How do Israeli universities communicate both competition and cooperation, and how is this reflected on their websites?**

The project builds on the foundational work of Nir Yemini, who analyzed the strategic behavior of Israeli academic institutions. While his research included both universities and colleges, our focus is solely on universities. This decision stems from the distinct nature of universities, which operate in a prestige-driven environment that prioritizes research excellence and global recognition , unlike colleges, which primarily serve regional and vocational functions.

By examining how universities present themselves through their websites and official publications, this research aims to uncover how strategies of coopetition (cooperation + competition) are articulated in their communication. Ultimately, the study seeks to determine whether the language used reflects a predominantly cooperative or competitive orientation.

2. Data

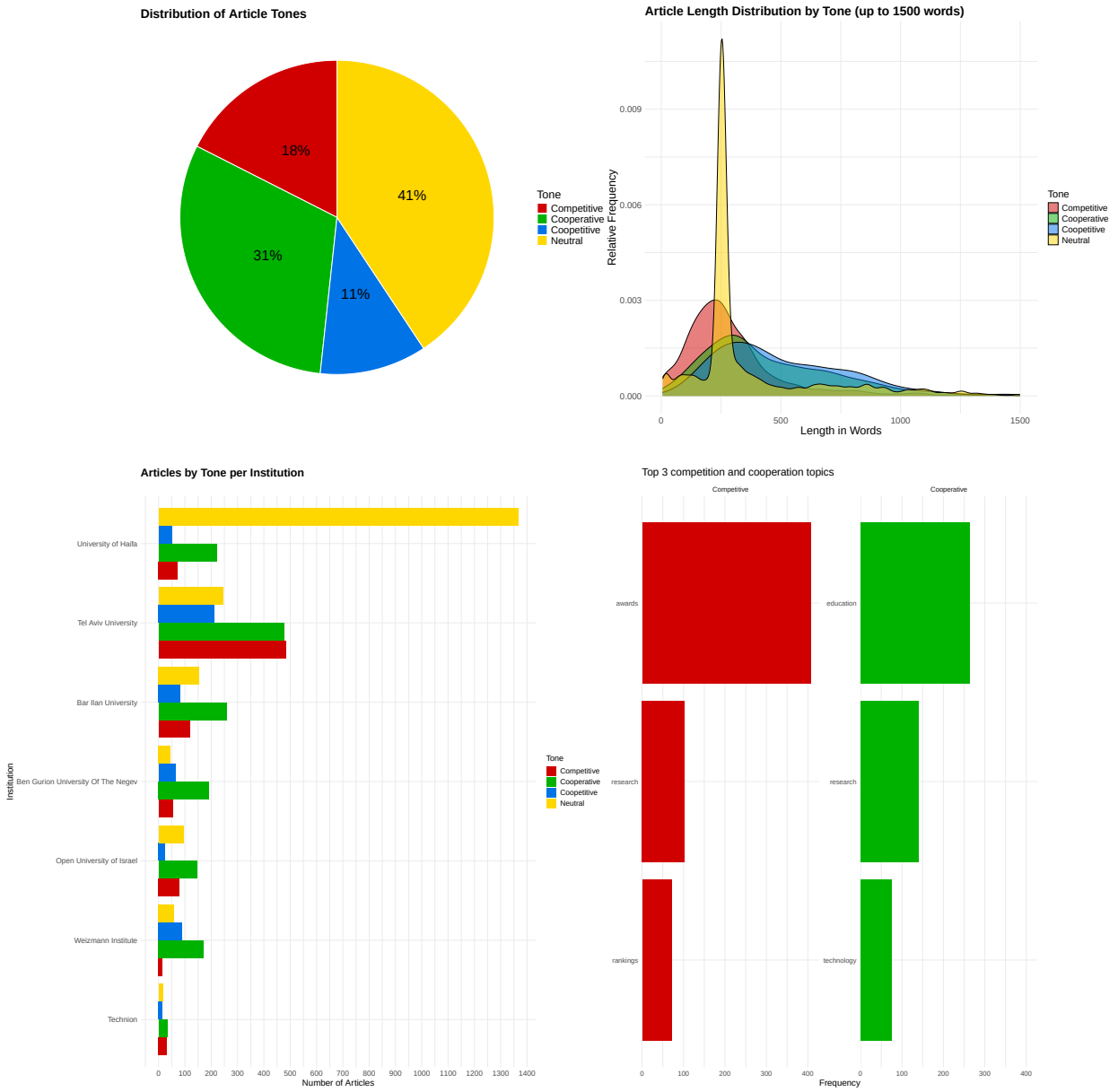
In Israel, there are 10 universities. Three universities' websites didn't respond so we couldn't scrape the data. We scraped seven websites with over 3,000 news articles from Israeli universities, including: Ben-Gurion University, Tel Aviv University, the Technion, the Hebrew University, and Bar-Ilan University. Each article was annotated by ChatGPT 4o1 API based on its tone: cooperative, competitive, coopetitive, and neutral. We used reasoning questions about keywords and structural metadata to make sure that every article is classified correctly. Here is the dataset summary:

```
Articles: 4,000+

Universities: 7 (Not including Ariel Raichman and Huji)

Features: tone, cooperation type, thematic tags, article length, etc.

Storage: CSV file with 11 structured columns; full description in /data/README.md
```

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. Methods and Results

We analyzed how Israeli universities frame their messaging. Given the subjective, categorical nature of tone and lack of labeled data, we used exploratory visualizations rather than predictive modeling. All visualizations in this report were generated using R and ggplot2.



3.1 Distribution of Article Tones - Most articles use a neutral tone. In terms of competition and cooperation, more articles have a cooperative tone than a competitive one, and a smaller proportion is cooperative.

3.2 Article Length by Tone - A plot that shows that neutral articles are typically shorter, while cooperative and coopetitive articles tend to be longer.

3.3 Articles by Tone per Institution - A grouped bar chart reveals that most universities use a cooperative tone, while Tel Aviv University uses more competitive framing.

3.4 Universities' most discussed topics - Competitive: awards, research, and ranking. Cooperative: education, research, and technology.

4. Limitations and Future Work

This study is subject to several limitations that should be acknowledged. First, we were unable to collect data from all university websites, as some sites were either inaccessible or malfunctioning during the data collection period. As a result, the dataset does not fully represent the entire university landscape in Israel.

Second, the analysis focused exclusively on universities and did not include colleges. Since universities and colleges serve different functions and operate within distinct institutional frameworks, this focus limits the generalizability of the findings. Future research could incorporate both types of institutions to explore potential differences in the communication of cooperation and competition strategies.

Third, the timing of website publications varied significantly across institutions. While some universities published recent updates, others had only older materials available. This temporal discrepancy makes it difficult to compare institutions within the same time frame and limits our ability to identify concurrent trends or strategic shifts.

Finally, the analysis relied entirely on the ChatGPT language model.

Dependence on a single model introduces potential bias or inaccuracies in content interpretation and tagging. Future studies may benefit from integrating additional language models or combining automated tagging with manual annotation to improve reliability and deepen the analysis.

Addressing these limitations in future research will contribute to a more comprehensive understanding of Israel's higher education landscape and provide a clearer view of how cooperation and competition strategies are articulated across the sector.

Code & Data available at: **GitHub**

Appendix

Variables in `ALL_articles.csv`

Variable	Description
`institution`	Name of the university that published the article
`date`	Publication date of the article (in day/month/year format)
`title`	The article's headline
`url`	Link to the original article
`tone`	Categorical tone label: Cooperative, Competitive, Neutral, or Coopetitive
`mentions_cooperation`	TRUE/FALSE – Whether cooperation is explicitly mentioned
`mentions_competition`	TRUE/FALSE – Whether competition is explicitly mentioned
`cooperation_type`	Type of cooperation (e.g., Domestic, International, Interdisciplinary)
`thematic_tags`	Main themes or topics mentioned, separated by semicolon (;)
`justification`	Free-text field justifying the tone classification (optional)
`length_words`	Number of words in the article (numeric approximation)

final-project/ README.md # Project overview and team info proposal.Rmd # Research proposal file final_report.Rmd # Full analysis and visualizations final_report.pdf # Knitted output data/ ALL_articles.csv # Cleaned dataset (annotated) shanghai.csv # Shanghai ranking data code/ plotting_code.R # Reusable ggplot functions scrape/ scrape_code.R # Scraping scripts (if relevant)

Source code

```
library(knitr)
library(tidyverse)
library(broom)
library(htmltools)
library(ggplot2)
library(patchwork)
library(lubridate)
library(forcats)
library(tidytext)




# Reading
df <- read_csv("C:/Avoda/submission/ALL_articles_final_corrected.csv", show_col_types = FALSE)
  mutate(
    tone = tolower(trimws(tone)),
    tone = case_when(
      tone %in% c("cooperative", "collaborative") ~ "cooperative",
      tone == "competitive" ~ "competitive",
      tone == "neutral"     ~ "neutral",
      tone == "Coopetitive"       ~ "Coopetitive",  #
```

```
      TRUE ~ NA_character_
    ),
    length_words = suppressWarnings(as.numeric(length_words)),
    Date = suppressWarnings(dmy(Date))
  ) %>%
  filter(!is.na(tone), !is.na(institution))


library(tidyverse)
library(patchwork)
library(readr)
library(lubridate)

# === Load and clean the dataset ===

df <- read_csv("C:/Avoda/submission/ALL_articles_final_corrected.csv", show_col_types = FALSE) %>%
  mutate(
    tone = tolower(trimws(tone)),
    tone = case_when(
      tone %in% c("cooperative", "collaborative") ~ "Cooperative",
      tone == "competitive" ~ "Competitive",
      tone == "neutral"     ~ "Neutral",
      tone == "coopetitive" ~ "Coopetitive",
      tone == "mixed"       ~ "Coopetitive",
      TRUE ~ NA_character_
    ),
    length_words = suppressWarnings(as.numeric(length_words)),
    Date = suppressWarnings(dmy(Date))
  ) %>%
  filter(!is.na(tone), !is.na(institution))

# === Defining Colors ===
tone_colors <- c(
  "Cooperative" = "#00B200",   #
  "Competitive" = "#D00000",   #
  "Neutral"     = "#FFD700",   #
  "Coopetitive" = "#0073E6"    #
)

# === Universities Names  ===
institution_labels <- c(
  "BGU"     = "Ben Gurion University Of The Negev",
  "BIU"     = "Bar Ilan University",
  "HAIFA"   = "University of Haifa",
  "OpenU"   = "Open University of Israel",
  "TECH"    = "Technion",
  "TLV"     = "Tel Aviv University",
  "Weizmann Institut" = "Weizmann Institute"
)
```

```r
# ===count table by Uni and Tone ===
df_counts <- df %>%
  count(institution, tone) %>%
  complete(institution, tone = names(tone_colors), fill = list(n = 0)) %>%
  mutate(institution = recode(as.character(institution), !!!institution_labels))

institution_totals <- df_counts %>%
  group_by(institution) %>%
  summarise(total = sum(n), .groups = "drop")

df_counts_ordered <- df_counts %>%
  left_join(institution_totals, by = "institution") %>%
  mutate(institution = fct_reorder(institution, total))

# === Pie chart ===
pie_chart <- df %>%
  count(tone) %>%
  mutate(
    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 %>%
  filter(length_words < 1500) %>%
  ggplot(aes(x = length_words, fill = tone)) +
  geom_density(alpha = 0.5) +
  labs(title = "Article Length Distribution by Tone (up to 1500 words)",
       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)
```

```
  )

# === Bar Plot per Institution ===
coopetition_bar <- df_counts_ordered %>%
  ggplot(aes(x = n, y = institution, fill = tone)) +
  geom_col(position = "dodge", height = 0.7) +
  labs(title = "Articles by Tone per Institution",
       x = "Number of Articles", y = "Institution", fill = "Tone") +
  scale_x_continuous(breaks = seq(0, 1500, by = 100)) +  #every 1000
  scale_fill_manual(values = tone_colors) +
  theme_minimal() +
  theme(
    plot.title = element_text(size = 18, face = "bold"),
    axis.text.y = element_text(size = 12),
    axis.text.x = element_text(size = 12),
    axis.title = element_text(size = 14),
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 13)
  )

# === Top 3 competition and cooperation topics ===

custom_stopwords <- c("article", "articles", "highlights", "news", "press", "media", "update",

tags_df <- df %>%
  filter(tone %in% c("Cooperative", "Competitive"), !is.na(thematic_tags)) %>%
  separate_rows(thematic_tags, sep = ";|,") %>%
  mutate(thematic_tags = tolower(trimws(thematic_tags))) %>%
  filter(!thematic_tags %in% custom_stopwords)

#sort by tone
top_tags <- tags_df %>%
  count(tone, thematic_tags, sort = TRUE) %>%
  group_by(tone) %>%
  slice_max(n, n = 3, with_ties = FALSE) %>%
  ungroup()

top_tone_tags_plot <- top_tags %>%
  mutate(thematic_tags = reorder_within(thematic_tags, n, tone)) %>%
  ggplot(aes(x = thematic_tags, y = n, fill = tone)) +
  geom_col(show.legend = FALSE) +
  coord_flip() +
  facet_wrap(~tone, scales = "free_y") +
  labs(
    title = "Top 3 competition and cooperation topics",
    x = NULL,
    y = "Frequency"
  ) +
  scale_fill_manual(values = c("Competitive" = "#D00000", "Cooperative" = "#00B200")) +
  scale_y_continuous(expand = expansion(mult = c(0, 0.05))) +
```

```
  scale_x_reordered() +
  theme_minimal(base_size = 14)

# === connect ===
(pie_chart + length_density) /
  (coopetition_bar + top_tone_tags_plot) +
  plot_layout(heights = c(1, 1.3))
```