CERIAS

The Center for Education and Research in Information Assurance and Security

Mechanisms of Virality in Public Online Discourse

- **Is it possible to detect the potential for virality within a social media network by analyzing variations in human value features?**

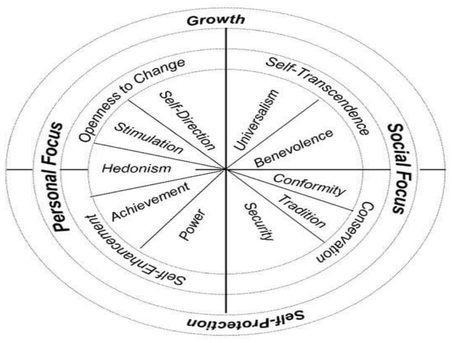
Analyzing Value Variations in Online Communities

- **Focus:** Examining discussions in platform-specific communities.
- **Comparison:** Identifying differences between **high-engagement** and **low-engagement** groups.
- **Time-Based Analysis:** Tracking how value distributions shift over time.
- **Outcome:** Understanding how values influence engagement trends.

Contact Information:
Author: Nicholas Harrell
Email: nharrel@purdue.edu
Affiliated with: CERIAS, PMRI

Classification Process

Step 1. Finetune a DeBERTa LLM to classify Values.



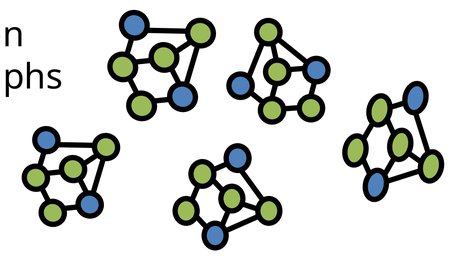
$$E = \begin{bmatrix} e_1^{(1)} & e_2^{(1)} & \dots & e_{768}^{(1)} \\ e_1^{(2)} & e_2^{(2)} & \dots & e_{768}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ e_1^{(10)} & e_2^{(10)} & \dots & e_{768}^{(10)} \end{bmatrix}$$

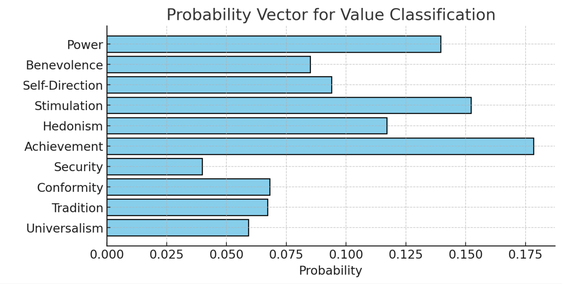
where:

- $E \in \mathbb{R}^{10\times768}$ is the **final embedding matrix**.
- $e_j^{(i)}$ represents the **embedding value** for the j**-th dimension** of the i**-th value**.
- Each row corresponds to one of the **10 human values**.
- Each column represents one of the **768-dimensional embedding features**.

Step 2. Extract Last Layer Embeddings for each training sample.

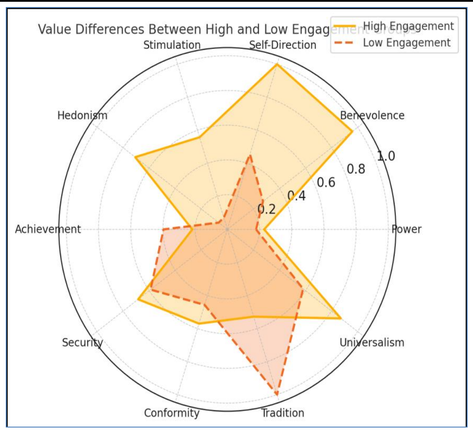
Step 3. Using Graph Convolution Networking, I generate sub-graphs for each value.





Step 4. Generate Probability Embedding using softmax.

Step 5. Analyze the differences in high engagement and low engagement groups within a topic community.



Analysis

Platforms analyzed

gab

MINDS

VK

Telegram

GETTR

4chan

Results

Table 4.2. Percentage of Staticially Significant Topics Found Based on Threshold

Threshold	Significant Results	Significance Rate (%)	Median Tš	Mean Effect Size
50	73	74.49	15.43	0.26
60	70	71.43	15.47	0.27
70	70	71.43	15.36	0.27
80	72	73.47	14.48	0.32
90	62	63.27	12.17	0.39

Note. Percentages consider all significant topics including ones filtered out for low effect scores.

Out of **105 identified topic communities**, **73% exhibited a statistically significant difference** in value profiles between **high- and low-engagement groups**, when segmented by the median engagement rate.

Conclusion: The study demonstrates that it is possible to detect potential of virality based on value changes.

Implications: Strategic message adjustments that align with a community's **value profile** may effectively **influence engagement behavior**, potentially amplifying reach and interaction. This implication could have both positive and negative benefits dependent upon the application.

PURDUE UNIVERSITY

CERIAS