**Project Title: IDENTIFICATION OF CRITICAL RELATED SUB-COMPONENTS IN ONLINE CONVERSATIONS**

**Team members:**

| Full name | ASU ID |
|---|---|
| Dhruva Patel | 1230583512 |
| Priyesha Jethi | 1229734729 |
| Sagar Rajesh Kumar Sinha | 1229611034 |
| Swati Kashyap | 1225466398 |

**Problem statement:**

Online conversations typically exist in two primary forms: collaborative, characterized by rational engagement, and canceling, marked by less rational discourse, often leading to someone exiting mid-discussion. The internal dynamics of both of these forms consist of several related sub-components such as hate speech, microaggressions, trolls, politeness, alignment, deception, etc. Hence, the project attempts to decipher some of these latent components (specifically hate speech, trolls and deception) in discussions, which in turn would help in better estimation of the overall effect. The results of this study could potentially aid researchers to conglomerate ideas from seemingly independent domains.

**Related work:**


1. Mueller, T.S. 2021. Blame, then shame? Psychological predictors in cancel culture behavior. The Social Science Journal. (Jul. 2021), 1–14. DOI:https://doi.org/10.1080/03623319.2021.1949552.
2. D. Clark, M. 2020. DRAG THEM: A brief etymology of so-called “cancel culture.” Communication and the Public. 5, 3–4 (Sep. 2020), 88–92. DOI:https://doi.org/10.1177/2057047320961562.
3. Chen, K. and Sanchez, L.A.D. 2019. Conversation Structure and Quality of Political Discourse: Evidence from Reddit. SSRN Electronic Journal. (2019). DOI:https://doi.org/10.2139/ssrn.3362008.
4. Zidraşco, T. and Bobicev, V. Agreement: How to Reach it? Defining Language Features Leading to Agreement in Dialogue.
5. Ali, O. et al. 2020. Automated Detection of Racial Microaggressions using Machine Learning. 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (Canberra, ACT, Australia, Dec. 2020), 2477–2484.

6. Addawood, A. et al. 2019. Linguistic Cues to Deception: Identifying Political Trolls on Social Media. Proceedings of the International AAAI Conference on Web and Social Media. 13, (Jul. 2019), 15–25. DOI:https://doi.org/10.1609/icwsm.v13i01.3205.
7. Cao, S. et al. 2015. GraRep: Learning Graph Representations with Global Structural Information. Proceedings of the 24th ACM International on Conference Information and Knowledge Management (Melbourne Australia, Oct. 2015), 891–900.
8. Grover, A. and Leskovec, J. 2016. node2vec: Scalable Feature Learning for Networks. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Francisco California USA, Aug. 2016), 855–864.
9. Perozzi, B. et al. 2014. DeepWalk: online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (New York New York USA, Aug. 2014), 701–710.
10. Li, C. et al. 2021. Knowledge Graph Analysis of Russian Trolls: Proceedings of the 10th International Conference on Data Science, Technology and Applications (Online Streaming, --- Select a Country ---, 2021), 335–342.

**Initial hypothesis:**
Our research endeavors to explore the following question: Can we computationally model human social behavioral patterns in online discussions through a collective study of various aforementioned sub-components? Consequently, our ultimate objective is to establish a baseline accuracy using graph-based deep learning approaches on a generalized testbed that addresses this endeavor.

**Dataset(s):**

| | |
|---|---|
| Dataset source (link and reference) | Archive of Reddit Conversations: https://convokit.cornell.edu/documentation/subreddit.html<br><br>**Provide the proper ACM reference format**<br><br>The corpora collection has been derived from PushShift's reddit data archives. |
| Number of instances | Archive consists data from 948,169 subreddits till Oct 2018 |
| Number of features | 14 |
| Class distribution (# instances in each class, if applicable) | None currently. We aim to perform label annotation and achieve a balanced class distribution amongst the problems under consideration. Though in practicality, we might probably get an imbalanced dataset. |

| | |
|---|---|
| Dataset splits | Not applicable now. Depends on the collected dataset's label distribution after annotation. |
| Preprocessing steps | Since the conversations are hosted on the Convokit platform, they are stored in a particular format. Hence we utilize the following preprocessing steps:<br>1. Convert the data hosted on the Convokit platform into dataframes to standardize the format for analysis<br>2. Remove bot comments and entries lacking speaker information to ensure data quality and relevance<br>3. Exclude conversations that have only a single level of depth to focus on better quality exchanges<br>4. Label annotation on the preprocessed dataset(s)<br>5. Transform the processed data into a structure compatible with Graph Neural Networks (GNNs), ensuring the data can be efficiently processed and analyzed by these models<br>6. Stratified data split<br>7. Treat data imbalance. Implement strategies to ensure that subproblems with lesser representation are adequately accounted for, preventing bias in model performance. |

**Method(s):**
In our project, we will address the problem of nodes, edges and graph detection using a combination of multiple algorithms, each chosen for its unique strengths and capabilities, and perform a comprehensive empirical analysis.

For node classification, we have planned to utilize the following algorithms

- DeepWalk
- BertGCN
- GraRep

For link prediction, our plan is to utilize

- GNAE: Graph Neural Auto-Encoders
- Graph InfoClust
- Variational Graph AutoEncoders

We plan to apply these algorithms through widely-used deep learning frameworks like PyTorch, Keras, and TensorFlow. Opting for these libraries guarantees access to effective and well-maintained tools, facilitating the development and experimentation of these algorithms.

The differentiation lies in the fact that a combination of these detection approaches would enable us to deal with a complex social problem, such as modeling of cancel culture across a greater breadth, instead of focusing on a specific problem and ignoring the related sub-components.

To the best of our knowledge, currently no State of the Art (SOTA) method has been developed and we will try establishing the baseline for the proposed problem statement.

**Evaluation:**
We employ a comprehensive approach to dataset and algorithm evaluation, utilizing both quantitative and qualitative metrics. This enables a thorough empirical analysis of both of our tasks, allowing us to assess the effectiveness of our heuristics in dataset creation and accuracy of task-specific algorithms.

- Inter-Annotator Agreement (IAA): Usage of metrics such as Cohen's Kappa or Fleiss' Kappa between different annotators

- Intra Annotator Agreement**:** Measure of consistency within the annotations provided by a single annotator over multiple instances

- Cross Validation and Splitting Strategies
    - Stratified Cross Validation
        - Since our annotation procedure might result in class imbalance, we will use stratified cross validation to ensure that each of the validation split(s) maintains the same class distribution as the original dataset

- Node, Edge and Graph Classification (measurement against other baseline models in related work(s))
    - Accuracy
    - Precision, Recall and F1 Score
    - AUC Curve

For qualitative analysis, we will be using the following metric(s)
- Node Coloring
- t-SNE (t-Distributed Stochastic Neighbor Embedding)

**Management plan:**

| | |
|---|---|
| Dataset Creation | Swati, Dhruva |
| Data Preprocessing | Priyesha, Sagar |
| Model Training and Development | Priyesha, Sagar |
| Evaluation and Optimization | Swati, Dhruva |
| Future Advancements | All members |
| Final Report Documentation | All members |

Assessing each member's strengths, skills, and interests, we allocated subparts of the project accordingly, ensuring everyone is working on aspects they are proficient in and passionate about.

To hold each other accountable, we'll establish clear deadlines and milestones for each task or deliverable. Regular check-ins will be scheduled to review progress, address any challenges, and provide support if needed. We plan to hold weekly in-person meetings to work together and integrate, and communicate through Zoom and WhatsApp for any updates.