Modeling Persuasion in Reddit Conversations

Abstract

This study examines how conversation dynamics and user characteristics influence persuasion within the Change My View (CMV) subreddit. We employ Graph Neural Networks (GNNs) and fine-tuned BERT embeddings to predict the likelihood of a comment receiving a delta, an indicator of successful persuasion. Our experiments reveal that GNN-based models, particularly the Distance-Weighted GNN, outperform baseline models by effectively leveraging conversation structure. Contrary to prior research emphasizing OP malleability, we find that enriched OP embeddings provide minimal additional benefits. These results suggest that capturing conversational depth and structure is more critical for persuasion modeling than OP-specific linguistic features alone.

Introduction

Persuasion plays a crucial role in shaping public opinion, decision-making, and discourse across various domains, including law, politics, and online discussions. Digital platforms such as Reddit offer a unique opportunity to analyze how persuasion unfolds in real-world debates. The Change My View (CMV) subreddit is particularly suited for this research, as it provides structured discussions where users explicitly indicate when they have been persuaded by awarding deltas to impactful comments. This structured feedback makes CMV an ideal dataset for studying persuasion in online conversations.

Unlike traditional classification tasks, persuasion modeling requires capturing subtle, context-dependent cues that emerge over the course of interactions. These nuances are difficult to detect using standard classifiers or zero-/few-shot learning approaches, as they lack the ability to model complex conversational dependencies and evolving discourse structures. Traditional research on persuasion modeling has primarily relied on linguistic and sentiment-based features to predict argument success. While these methods provide useful insights, they often overlook the conversational structure—how arguments develop over the course of a discussion. Graph-based learning methods, particularly Graph Neural Networks (GNNs), allow for a more nuanced representation of discussion dynamics by incorporating both textual content and hierarchical relationships within conversations.

This study explores whether integrating GNN-based representations with fine-tuned language models improves the ability to predict persuasive success. Specifically, we investigate three key aspects:

- How well do GNN-based models perform compared to traditional language models like BERT?
- What role does the depth of a comment in a discussion tree play in its persuasive success?
- Do enriched OP embeddings contribute significantly to persuasion modeling?

To address these questions, we represent CMV discussions as graphs, where nodes correspond to individual comments, and edges reflect reply relationships. We compare a baseline BERT model with two GNN architectures: a Distance-Weighted GNN, which assigns predefined edge weights based on comment depth, and an Edge-Weighted GNN, which learns edge transformations dynamically. Additionally, we incorporate OP-specific embeddings to capture a user's malleability and assess whether these features enhance predictive performance.

By analyzing thousands of discussions with annotated persuasive comments, this research aims to provide deeper insights into the relationship between conversation structure and persuasion.

Literature Review

Persuasion in online discussions has been extensively studied using data from the *Change My View* (CMV) subreddit, where explicit markers like delta awards indicate persuasive success. Wei et al. (2016) focus on ranking persuasive comments using a combination of surface text features, argumentation-based attributes, and social interaction metrics [2]. Their findings highlight that while surface-level features are less effective, argumentation-related and social interaction features—such as reply depth and interaction patterns—significantly improve persuasion prediction.

Tan et al. (2016) complement this work by examining interaction dynamics and linguistic strategies associated with persuasion [1]. They show that early participation increases persuasive success, and stylistic elements like emotional arousal, concreteness, and the use of personal pronouns are strong indicators. Interestingly, they find that arguments dissimilar in wording from the original post are more likely to persuade, suggesting that introducing novel perspectives is more impactful than linguistic alignment.

Together, these studies underscore the importance of both conversational structure and linguistic style in modeling persuasive success in online discussions.

Dataset Overview

The dataset used in this study is derived from the Change My View (CMV) subreddit, a structured online discussion platform where users engage in debates and award deltas to comments they find persuasive. The dataset consists of hundreds of thousands of comments spanning discussions from January 2013 to August 2015, with explicit delta annotations identifying successful persuasion attempts.

Each conversation is structured as a tree, with the Original Poster (OP) as the root node and subsequent replies forming branches. A comment is considered persuasive if the OP explicitly acknowledges it with a delta. However, delta assignments are relatively sparse, making the dataset highly imbalanced, with only a small fraction of comments receiving a delta.

To facilitate graph-based learning, we preprocess the dataset by converting conversations into structured graphs. Nodes represent individual comments, and edges indicate the hierarchical reply relationships. Each node is enriched with BERT-based embeddings, capturing semantic meaning, while OP embeddings are fine-tuned separately to incorporate user-specific malleability. Additionally, each edge is assigned attributes based on the shortest path distance from the OP, which plays a critical role in modeling the impact of conversation depth on persuasion.

To further understand the structure of persuasive conversations, Figures 1 and 2 provide a visual representation of the dataset. The first figure illustrates an example of a conversation graph, highlighting the hierarchical nature of discussions and the position of delta nodes. The second figure presents a distribution of delta node depths, revealing that most persuasive comments occur within the first level of the conversation tree, suggesting that early engagement in discussions is crucial for persuasion.

1.

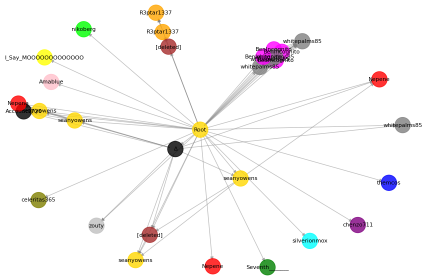
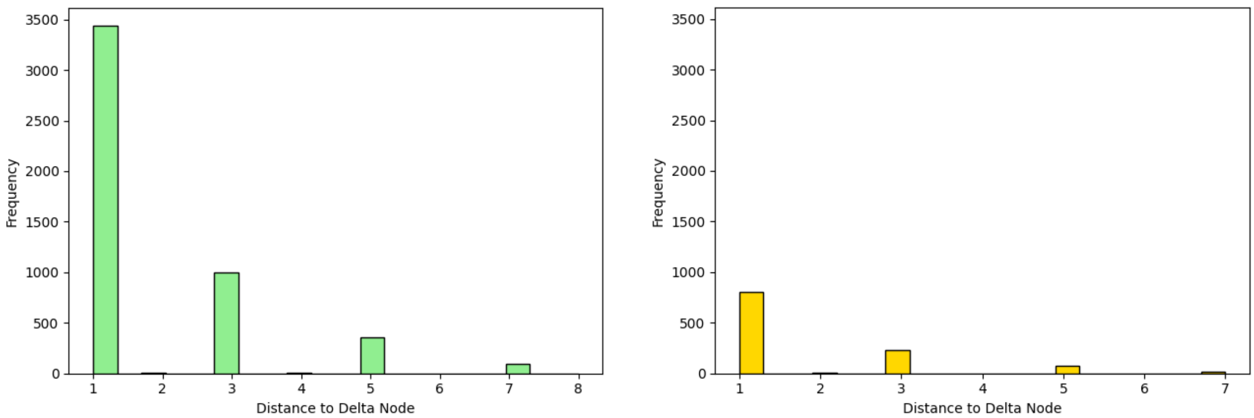


Figure 1: Example Conversation Graph with OP Nodes (as root), Delta node in black and All Edges.



(a) Train Set.

(b) Test Set.

Figure 2: Distribution of Distances to Delta Nodes from Root.

These preprocessing steps and insights provide the foundation for our subsequent modeling approach, which integrates Graph Neural Networks (GNNs) with distance-aware message passing techniques to predict persuasion success.

Methodology

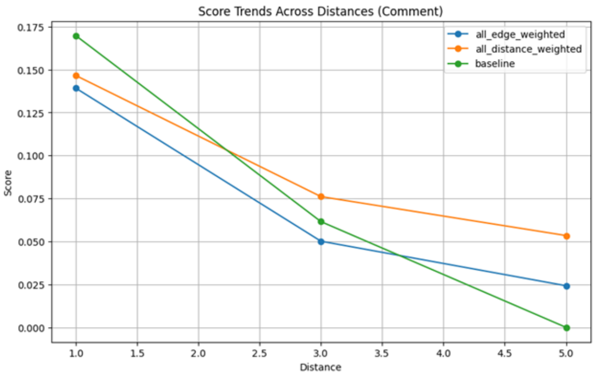
Our methodology involves three modeling approaches to predict the likelihood of a comment receiving a delta in the CMV subreddit. The baseline model utilizes BERT embeddings for both comments and OPs, treating each node independently without incorporating graph-based propagation. In contrast, the Distance-Weighted GNN model assigns predefined edge weights based on a comment's depth within the conversation tree, leveraging two Graph Convolutional Network (GCN) layers to aggregate features and enhance message passing. Lastly, the Edge-Weighted GNN model applies learnable Multi-Layer Perceptrons (MLPs) to transform both node features and edge attributes, allowing for more dynamic message propagation across the conversation structure.

To analyze the effectiveness of these models, we explore different feature groups: (1) embeddings derived only from comment nodes using either BERT or GNN-based methods, (2) embeddings that combine comment representations with OP embeddings obtained from standard BERT, and (3) embeddings that integrate comment features with OP representations fine-tuned specifically for persuasion prediction. Performance evaluation is conducted using accuracy and F1-score metrics, with additional analysis focusing on how model performance varies across different conversation depths.

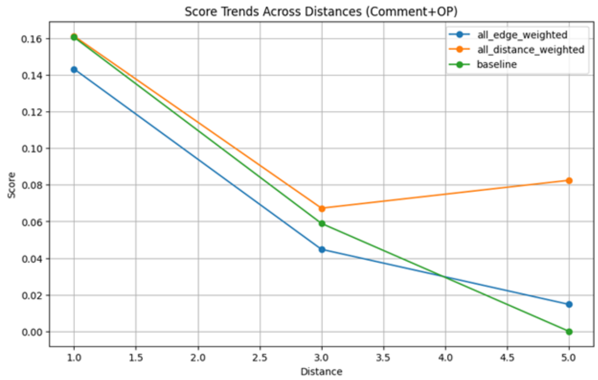
Feature groups include:

- Comment-only embeddings (BERT or GNN-based).
- Comment + OP embeddings from standard BERT.
- Comment + OP embeddings from fine-tuned BERT.

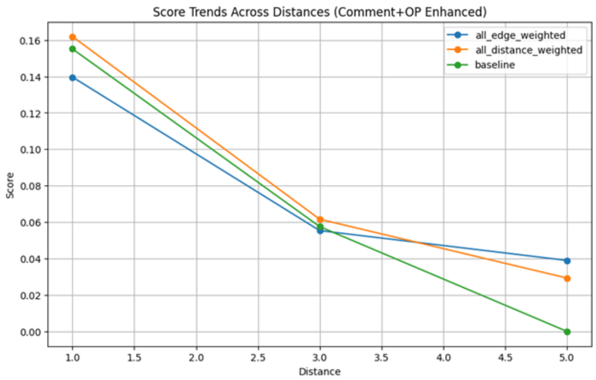
Results and Discussion



(a) Only Comment



(b) Comment+OP



(c) Comment+Enhanced OP

Figure 3: Results by Model and Input

Distance Impact on Performance

Our findings indicate that GNN-based models exhibit greater resilience when predicting persuasion at deeper conversation levels compared to traditional BERT-based models. In particular, distance-aware modeling strategies, such as the Distance-Weighted GNN, successfully preserve performance across increasing comment depths, while the baseline models tend to degrade significantly. This pattern suggests that persuasion dynamics are highly dependent on hierarchical relationships, and capturing these relationships through graph-based message passing is essential for maintaining predictive accuracy.

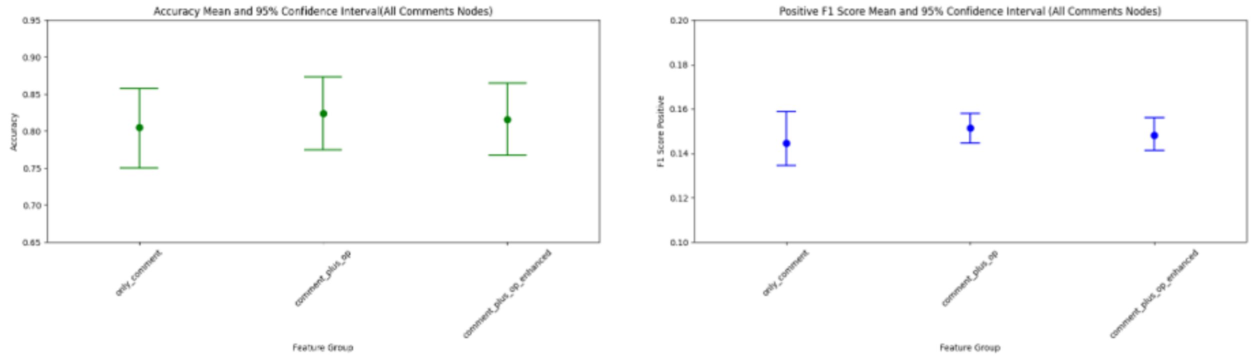
Figure 3 illustrate how model performance varies across different distances from the Original Poster (OP) for three different feature configurations:

Figure 3a presents the score trends for comment-only embeddings, comparing the Edge-Weighted GNN, Distance-Weighted GNN, and the baseline BERT model. Figure 3b extends the analysis to comment and OP embeddings, incorporating OP-specific features to examine whether they contribute to preserving performance across depths. Figure 3c evaluates an enhanced OP embedding strategy, which aims to provide a more fine-grained representation of OP malleability. From Figure 3a, we observe that the Distance-Weighted GNN outperforms the other models beyond distance 3, where the baseline model's performance deteriorates sharply.

Introducing OP embeddings in Figure 3b results in slightly improved performance at shorter distances for GNNs, but the downward trend remains similar beyond depth 3, indicating that OP-specific features alone do not sufficiently counteract the loss of predictive power.

Feature Group Insights

Despite expectations, fine-tuned OP embeddings did not significantly enhance classification performance. This may indicate that OP malleability, while relevant, does not provide sufficient standalone



(a) Feature Group Mean Accuracy Comparison (b) Feature Group F1 Score Comparison

Figure 4: Feature Group Comparison

predictive power. Instead, conversational context and structure appear to be the dominant factors in determining persuasion success.

From Table 1, we observe that while models incorporating OP embeddings achieve slightly higher accuracy at distance 1, their overall accuracy slopes decline more steeply than the comment-only model. This suggests that OP embeddings provide initial advantages but do not contribute meaningfully as depth increases. Similarly, the F1-score positive slope decreases significantly with the addition of OP embeddings, indicating that they do not enhance predictive performance for persuasive comments at greater depths.

To further analyze the impact of different feature groups, we evaluate model accuracy and F1-score at various depths. The table below summarizes the results across models:

Model	Distance 1 Acc.	Acc. Slope	Dist. 1 F1	F1 Slope
Only Comment Node	0.77	-0.003	0.251	0.005
Comment Node + OP Node	0.79	-0.006	0.265	0.002
Comment Node + OP Fine-Tuned Node	0.79	-0.004	0.262	0.001

Table 1: Comparison of accuracy and F1-score changes across different depths and feature groups.

Future Work

- **Feature Engineering**: Explore additional features such as sentiment progression and user engagement metrics.
- **Advanced Graph Architectures**: To further improve long-distance relationship modeling, future research could explore attention-based GNNs, such as Graph Attention Networks (GATs), which dynamically weight influential comments. This may help capture persuasive influence that is not solely determined by tree depth but rather by conversational style and argument strength.
- **Real-World Applications**: Potential applications include automated moderation, marketing analysis, and research in social psychology.

Since the accuracy and F1-score slopes indicate diminishing returns for OP embeddings, future research should focus on refining OP representations or exploring alternative conversational features, such as sentiment trends or discourse structure.

Conclusion

Our results emphasize the critical role of leveraging conversation structure rather than relying solely on user-specific embeddings. While previous studies have highlighted the significance of original poster

(OP) characteristics in persuasion, our findings suggest that the success of persuasive comments is more closely tied to the broader conversational context and hierarchical relationships between responses. This underscores the need to model persuasion as an interactional phenomenon rather than an individual trait.

By incorporating graph-based approaches, we were able to capture the relational dynamics within discussions, revealing that structural positioning and contextual dependencies play a greater role than previously assumed. These insights suggest that future work should focus on refining graph-based propagation techniques, integrating more sophisticated hierarchical attention mechanisms, and exploring multimodal cues to further enhance the identification of persuasive comments. Additionally, expanding the dataset to include diverse discussion platforms and examining domain-specific persuasion strategies could provide a more comprehensive understanding of online persuasion dynamics.

Reflection

Working on this project provided valuable insights into both the technical and conceptual challenges of modeling persuasion in online discussions. One of the key lessons we learned was the complexity of integrating Graph Neural Networks (GNNs) with language models like BERT. While we initially expected that enriched OP embeddings would significantly enhance performance, our experiments revealed minimal gains, emphasizing the importance of focusing on conversational structure instead. Additionally, our statistical analysis highlighted the critical role of discussion depth in persuasion modeling, challenging some of our early assumptions. We also encountered challenges with data preprocessing and model evaluation, particularly in ensuring consistency across experiments. Despite these hurdles, the iterative process of refining our models and interpreting the results deepened our understanding of both the technical and theoretical aspects of computational persuasion.

Work Division

The project was a collaborative effort with each team member contributing to distinct yet interconnected components:

- **Dennis** was responsible for data acquisition, cleaning, and exploratory data analysis (EDA), as well as implementing GNN architectures and contributing to the report writing.
- **Gad** focused on the design of GNN architectures and the planning and execution of statistical experiments, alongside report writing.
- **Yuval** handled data preprocessing, implemented BERT fine-tuning, and contributed to writing the report.
- **Noam** was in charge of model evaluation, developing visualization tools for interpreting the results, and preparing the project presentations.

This division of labor ensured that each aspect of the project, from data processing to model development and analysis, was handled efficiently while fostering collaborative input across all stages.

References

[1] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. *Proceedings of the 25th International Conference on World Wide Web*, 2016.

[2] Zhongyu Wei, Yang Liu, and Yi Li. Is this post persuasive? ranking argumentative comments in online forum. In *Annual Meeting of the Association for Computational Linguistics*, 2016.