Machine Learning In Network Science

Final Project Report

Qihang PU, Wenxin WU, Yi-Ting KUO, Sri Vennela SEELAM

April 14, 2024

Abstract

This research analyzes sentiment dynamics in Reddit's network of subreddits through the lens of hyperlinks, with the objective of predicting inter-community relations and identifying potential conflicts. Previous works have explored textual similarity and troll-like behaviors within communities but lacked the predictive capacity for future interactions. Our methodology employs the JODIE and GraphSAGE models, integrating dynamic and static embeddings via recurrent neural networks and graph neural networks respectively. These models adeptly manage the temporal and structural complexities inherent in subreddit interactions. The performance of our models shows significant potential in capturing the nuanced dynamics of subreddit interactions, offering insights that could improve community management and moderation tools. Our results advocate for the integration of advanced embedding and multimedia content to deepen the understanding of online communities, paving the way for more robust and insightful analyses that can inform the decisions of moderators, social scientists, and policymakers.

1 Introduction and Motivation

The advent of digital platforms has catalyzed the formation of numerous online communities, among which Reddit stands out due to its unique structure of subreddits—distinct community-driven spaces centered around specific interests. Each subreddit is not only a collection of topics but also a network of users who actively shape the community's culture and discourse through their interactions. These interactions often occur through the creation of hyperlinks in posts, which serve as bridges connecting one subreddit to another, thus weaving a complex web of community relationships.

This project is devoted to analyzing the sentiment of these inter-community hyperlinks. Each hyperlink is annotated with three key properties: the timestamp, indicating the temporal context of the post; the sentiment of the source community post towards the target community post, which can be positive (+1) or negative (-1); and the text property vector of the source post, encapsulating features such as word count, readability, and thematic categories. This rich annotation enables a nuanced understanding of the dynamics at play. The network itself is directed, reflecting the one-way nature of each hyperlink, signed, capturing the positive or negative sentiment, temporal, highlighting the changing dynamics over time, and attributed, providing additional context through the properties associated with each link.

The motivation behind this project stems from the need to understand and preemptively identify negative interactions between subreddit communities that could lead to conflict. The ability to predict the sentiment of these hyperlinks is vital for maintaining the health of online discussions and community engagements. By understanding and predicting potential conflicts indicated by negative hyperlink sentiments, moderators and community managers can intervene proactively to foster a more constructive dialogue. This not only enhances the user experience but also contributes to the overall stability and growth of the online community ecosystem. This is particularly important

given the vast scale and dynamic nature of discussions on platforms like Reddit, where negative interactions can escalate quickly and disrupt community harmony. The importance of this problem is underscored by potential applications in community management, moderation tools, and social media analytics, where understanding and managing community dynamics are crucial.

- **Community Management**: Effective management of online communities relies heavily on understanding the sentiments and interactions within the community. By predicting the sentiment of hyperlinks between subreddits, community managers can gain insights into the underlying emotions and potential conflicts, allowing them to proactively address issues that might disrupt community harmony.

- **Moderation Tools**: The ability to predict negative sentiments in subreddit interactions can significantly enhance the functionality of moderation tools. By integrating sentiment analysis into these tools, moderators can be alerted to potentially harmful interactions automatically. This capability allows for quicker responses to negative content, thereby maintaining the quality and safety of discussions within the community. Additionally, these tools can help in fine-tuning the algorithms that detect and manage inappropriate or aggressive content, ensuring that the community standards are upheld effectively.

- **Social Media Analytics**: Analyzing the sentiment of links between communities provides valuable data that can be leveraged in social media analytics to understand trends, community health, and the impact of certain discussions or posts. This analysis can reveal how sentiments evolve over time and how they influence user behavior across different platforms. Insights gained from this analysis can guide strategic decisions in content creation, marketing strategies, and community engagement initiatives, aiming to enhance user experience and engagement on a larger scale.

By emphasizing these applications, this study not only addresses a significant gap in the current understanding of online community interactions but also proposes practical solutions that can be implemented to improve the overall health and productivity of these digital spaces.

2 Problem Definition

The problem of our project is predicting the sentiment of a hyperlink between two subreddit communities. Let $G = (V, E)$ be a directed, signed, temporal, and attributed graph where V represents subreddits, and E denotes the edges corresponding to hyperlinks between these subreddits.

Notation: Each edge $e \in E$ is defined by a tuple (u, v, t, s, p), where u and v are the source and target nodes (subreddits), t is the timestamp of the interaction, s is the sentiment label of the hyperlink (positive +1 or negative 1), and p is the text property vector of the post from u linking to v.

Definitions and Constraints:

- **Subreddit Embeddings[1]:** Each node u in V is associated with an embedding vector that captures characteristics of the subreddit. These vectors represent aggregated community attributes derived from collective activities and thematic engagements.

- **User Embeddings[2]:** Each subreddit contains many users. For each user, we have user embeddings which represent user-level characteristics like personal interaction styles, posting frequency, and the diversity of subreddit engagement.

- **Text Property Vector:** p includes features such as the number of words, readability score, and category labels extracted from the post.

Objective: Our goal is to develop a predictive model that maximizes the accuracy of sentiment prediction for hyperlinks. Formally, we aim to maximize:

$$\text{Accuracy}(\mathcal{M}) = \left(\frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} 1_{\{\text{sign}(\text{pred}_i) = \text{actual}_i\}} \right)$$

Problem Hardness: The problem is complex due to the high dimensionality of the feature space and the temporal dynamics of community interactions. The temporal aspect requires that our model adapt over time to the evolving nature of subreddit interactions. Furthermore, the prediction of sentiment as a binary classification (positive or negative) introduces challenges due to potential class imbalances and the subtleties of interpreting textual content in varying contexts.

3   Related Work

To identify misaligned inter-group links and communities, Srayan Datta et al. use term frequency-inverse document frequency (TF-IDF) to present Textual similarity include word embedding and topic modeling[3]. Justin Cheng et al. show that troll-like behaviors are concentrated in small number of communities by using bag-of-words analyzing users who were banned from these communities[4]. S. Kumar et al. provide new insights into how less than 1% of communities initiate 74% of conflicts on Reddit, highlighting the role of specific community dynamics in online conflicts.

Discussion on limitations of previous models that either focus only on static properties or do not predict future interactions effectively. The need for a model that can handle large-scale data and predict future embeddings efficiently.

In the literature, dynamic embeddings are recognized for their capacity to evolve over time, thereby providing a nuanced representation of the changing behaviors and attributes of users and items. These embeddings form trajectories that capture the temporal dynamics of user interactions and item popularity. In contrast, static embeddings, which are typically initialized with one-hot vectors, are limited in their ability to reflect the temporal evolution of user interests and item states, leading to suboptimal modeling of user behavior.

4   Methodology

4.1   Dataset Description

The subreddit-to-subreddit hyperlink network is extracted from the posts that create hyperlinks from one subreddit to another. We say a hyperlink originates from a post in the source community and links to a post in the target community. Each hyperlink is annotated with three properties: the timestamp, the sentiment of the source community post towards the target community post, and the text property vector of the source post. The network is directed, signed, temporal, and attributed.

The dataset utilized encompasses a time span from January 2014 to April 2017, centered on the subreddits of the Reddit platform. This dataset contains a total of 55,863 nodes, with each node representing a distinct subreddit. The edges, which signify hyperlinks between subreddits, amount to 858,490, illustrating the interconnectedness within the network. Each edge is assigned a weight of either -1 or +1, indicative of the qualitative attribute of the hyperlink. The negative edge takes 7.27% of all edges. Furthermore, the edges are characterized by textual property vectors, laying the groundwork for more nuanced feature analysis and pattern recognition.

Given the extensive volume of the dataset, which comprises a substantial number of edges and nodes, there are significant computational demands. To facilitate feasible and efficient model training within the constraints

of computational resources, a strategic subset of the most active subreddits was selected for analysis. This selection criterion, based on predefined activity metrics, ensures that the model training is viable even with limited computational capacity, while also maintaining the representativeness and relevance to practical applications. The approach adopted strikes a balance between the breadth and depth of the dataset, offering an effective method for edge prediction analysis in large-scale social networks.

4.2   Joint Dynamic User-Item Embedding Model

The JODIE model utilizes a dual embedding system for users and items, comprising static and dynamic components. Static embeddings, initialized with one-hot vectors, remain constant and represent the enduring characteristics of the entities. In contrast, dynamic embeddings are designed to capture the changing behaviors and attributes over time, with each interaction creating a trajectory of evolving embeddings.

At the core of JODIE are two recurrent neural networks (RNNs) responsible for the dynamic embedding updates. During an interaction, both user and item embeddings are concurrently updated. The user RNN, denoted as RNNU, refines the user's dynamic embedding by incorporating the previous state of the item and other interaction-specific features. In a similar fashion, the item RNN, RNNI, revises the item's dynamic embedding based on the user's preceding state. This reciprocal update process effectively models the intricate interplay between users and items, enhancing the representational quality of the embeddings.
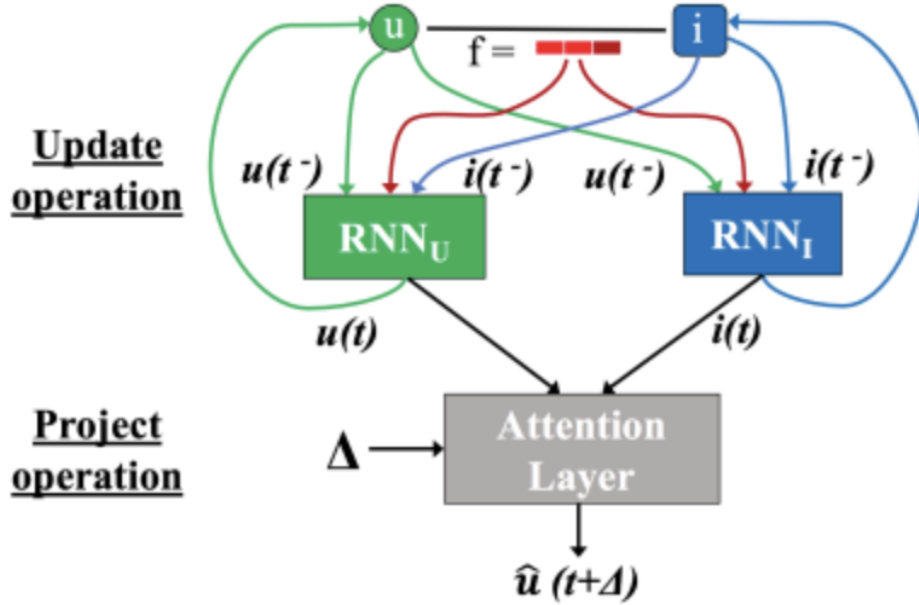

Figure 1: The JODIE model: After an interaction (u, i, t, f) between user u and item i, the dynamic embeddings of u and i are updated in the update operation with RNN_u and RNN_i, respectively. The projection operation predicts the user embedding at a future time $t + \Delta$.

JODIE is its capability to forecast future embeddings. The model employs a projection operator that extrapolates a user's embedding to any future time instance. This is facilitated through a temporal attention mechanism, which adjusts the user's current embedding in accordance with the time elapsed since their last interaction.

Both user embeddings and subreddit embeddings are generated toward one numerical vector in low dimensional space (a.k.a. embeddings) for each element. The embeddings are 300 dimensions each.

4.3 Modelling Methods

SAGE

As a variant of Graph Neural Networks (GNNs), SAGE employs a neighborhood aggregation strategy. This allows for the learning of node representations by sampling and aggregating features from a node's local neighborhood. In the context of social networks where interaction dynamics are complex, this method effectively captures the relational dependencies and community structures inherent within the network. This feature is particularly beneficial for our dataset, which, despite our focus on the most active subreddits, the graph remains substantial in size and complexity.

SAGE Module:
conv1: SAGEConv (in_feat; out_feat; aggregator_type)
conv2: SAGEConv (in_feat; out_feat; aggregator_type)
DotProductPredictor (pred)

Figure 2: Diagram of the GraphSAGE model architecture

Unlike transductive methods, which learn representations only for the seen nodes, SAGE can generalize to unseen nodes after training. This inductive property is particularly beneficial for evolving social networks where new nodes (subreddits) continually emerge, ensuring that the model remains applicable over time without the need for retraining from scratch. The GraphSAGE model stands out for its capacity to incorporate multifaceted node feature information, which is paramount when analyzing social networks replete with rich textual and attribute data. The architecture of the model facilitates the integration of intricate subreddit textual properties into the inductive learning process. This integration is essential for the enhancement of the quality of the representations learned, allowing for a nuanced analysis that is sensitive to the context and content of the subreddits.

Another pivotal advantage of the GraphSAGE model is its flexibility in the utilization of diverse neighborhood aggregation functions. These functions, which include mean aggregators, Long Short-Term Memory (LSTM) networks, and pooling mechanisms, provide the model with the adaptability necessary to be tailored to the particular characteristics of the dataset and the analytical tasks it undertakes. Specifically, the mean aggregator can distill the average influence of neighboring nodes. This feature is particularly beneficial for discerning the overarching sentiment or thematic orientation of a subreddit, derived from the nature of its hyperlinked neighbors, thereby contributing to a deeper understanding of the social network structure. The structural sparsity commonly observed in social networks, where certain subreddits may have scant connections, poses a significant challenge in network analysis. The SAGE model is robust to such sparsity, thanks to its aggregation mechanism that can effectively operate across neighborhoods of varying sizes. This mechanism does not rely on dense connectivity to function, making it particularly adept at handling the sparse interaction patterns inherent in large-scale social networks.

Figure 3: Visual illustration of the GraphSAGE sample and aggregate approach

5   Evaluation

In our evaluation, two main approaches are used to analyze the Subreddit social network and assess the performance of the model:

1. **Graph Analysis**: This approach involves the examination of the structural properties of the Subreddit social network graph, the key elements include -

   - **Description:**The network, comprising 35,776 nodes and 124,330 edges, reflects a complex system with numerous interactions. It is not connected and contains 497 distinct components, indicative of numerous, possibly niche, sub-networks isolated by specific reddit interests. The graph's density, a mere 0.000194, suggests a sparse network where connections are not fully utilized, common in large-scale networks where interactions are strategically significant.

   - **Degree Analysis:** The network exhibits a maximum degree of 2336, the average degree stands at 6.95. Suggesting that each subreddit, on average, directly interacts with approximately seven other subreddits. This relatively low average degree indicates that most subreddits are not densely connected within this network. The observed disparity between the maximum and average degrees suggests that the network likely follows a long-tail distribution.

   - **Edge Sentiment Distribution** : The percentage of negative edges in the graph is 7.273%. indicates that the vast majority of relationships (92.73%) are positive, suggesting that the network predominantly exhibits positive interaction characteristics, indicative of a community with strong cooperation or good relations. Although negative edges are less frequent, their impact on the network can be significant. Negative relationships may highlight potential conflicts, disputes, or points of division within the community. When aiming to improve social network management and optimization, analyzing negative edges is crucial for a deeper understanding and enhancement of the network's structure and functionality.
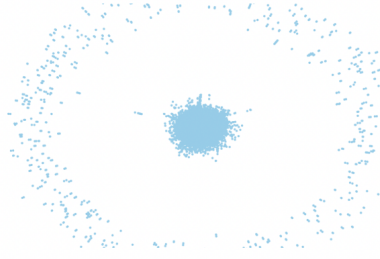

Figure 4: Visualization of subreddit-to-subreddit hyperlink network

   - **Sub-Graph Sampling** : Creating a new graph by filtering nodes that have less than 3 degrees. In subreddit social networks, some users with only a few connections may not be very active or their data may not be sufficiently representative. Subsequently, we partitioned the graph into training, testing, and validation subgraphs according to proportions of 0.5, 0.2, and 0.3, respectively.
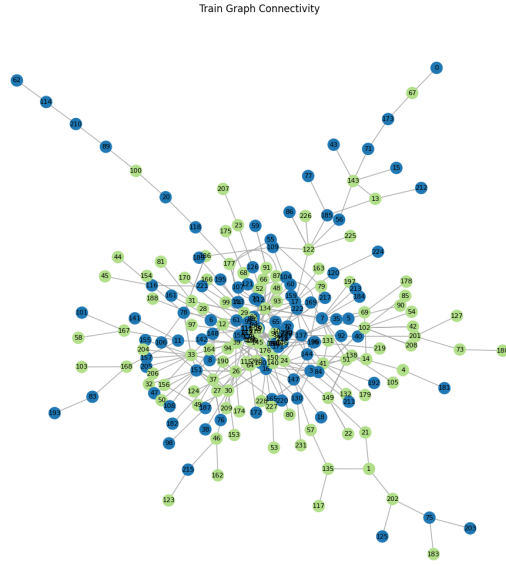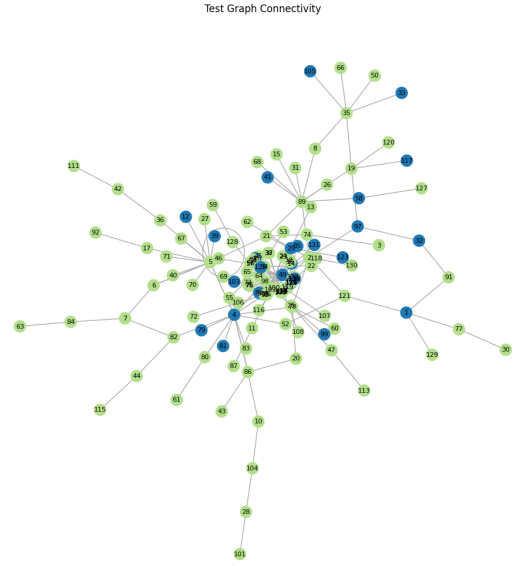
Figure 5: Train Sub Graph

Figure 6: Test Sub Graph

2. **Machine Learning Model Evaluation**: This approach focuses on assessing the performance of the Graph-SAGE model built to predict the sentiment of the links between the subreddits, it involves -

- **Model Training and Validation** : Using a split of the graph's edges for training and validation, the GraphSAGE model's ability to predict link sentiment is evaluated. Accuracy metrics are calculated for both sets.

- **Loss Calculation** : Use MSE metrics. Monitoring the model's loss during training provides insights into how well the model is learning to predict the sentiment of subreddit interactions.

- **Accuracy Tracking** : By tracking the accuracy of sentiment predictions on both the training and validation sets across epochs, the effectiveness of the model and its generalization capability are assessed.

Through these approaches together offer a comprehensive evaluation of both the structural properties of the Reddit social network and the predictive performance of the GNN model.
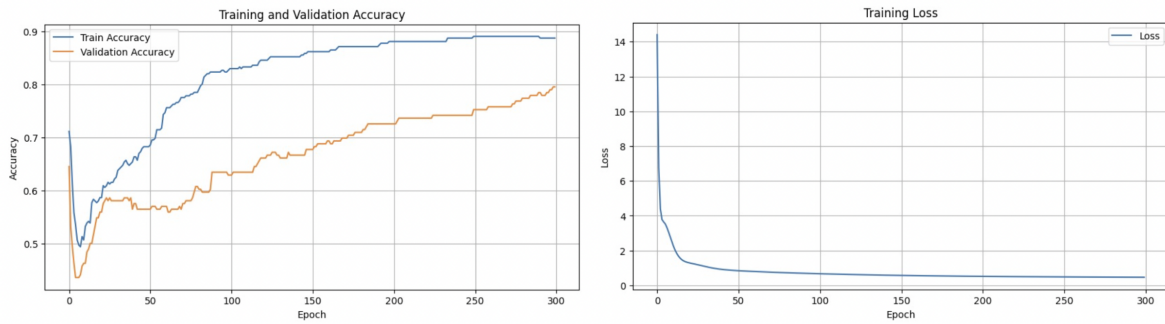

Figure 7: Model performance

6 Results

The results underscore the complexity and the richness of the Reddit social network, characterized by a mixture of supportive and contentious interactions among its subreddits. The effectiveness of the GraphSAGE model highlights the utility of the graph neural networks for social media analysis, especially in understanding and predicting the dynamics of user interactions based on network structure and node features and resulting in a graph with 622 nodes and 88,022 edges, with nodes representing subreddits and edges denoting the sentiment of hyperlinks between them. The edges were equipped with 86-dimensional feature vectors and binary sentiment labels.

The model was trained over 300 epochs using the Adam optimizer with a default learning rate of 0.001, employing mean squared error as the loss function. Training yielded an average accuracy of approximately 88.7%, while the validation accuracy averaged around 74.6% with a peak at approximately 79.6%. The test accuracy reached 72.6%. The training and validation accuracy curves suggested effective learning but hinted at possible overfitting, given the variability in validation accuracy.

The limitations of our research are -

- **Model Generalization**: The GraphSAGE model performed well on the current dataset, its ability to generalize to other types of social networks or datasets with different characteristics (such as those with less structured or sparser connections) remains untested. The model's performance could vary significantly in different contexts, affecting its utility in broader applications.

- **Imbalance in Sentiment Labels**: If the dataset exhibits a significant imbalance in sentiment labels, it could lead to biases in the model's predictions towards the majority class. This could affect the model's accuracy in realistically predicting negative sentiments or detecting subtle nuances in user interactions.

- **Scalability Issues**: Handling large-scale data effectively is a critical concern, especially as the size of the graph grows with the addition of more subreddits and links. The computational complexity and memory requirements for training graph neural networks might pose scalability challenges.

- **Overfitting Potential**: While the model showed good generalization in the initial tests, there's always a risk of overfitting, especially as the complexity of the model increases or if the training data is not sufficiently diverse.

7 Future Work

Looking ahead, there are several avenues we aim to enhance the analyzing social networks like Reddit using graph neural networks and focus on enhancing model robustness and feature integration. Testing the model's generalization across different social media platforms and incorporating richer datasets will be crucial. Introducing advanced features such as natural language processing for text, temporal data analysis, and multimedia content could offer a more comprehensive understanding of community dynamics. Addressing data imbalance through sophisticated sampling techniques and improving scalability with distributed computing or more efficient algorithms will ensure the model's applicability to larger datasets. Moreover, implementing regularization techniques to prevent overfitting will further validate the model's utility in diverse scenarios, paving the way for more robust and insightful social media analysis tools.

8 Conclusion

In conclusion,the tailored edge prediction model we developed for the Reddit social network effectively utilized the graph neural networks (GNNs) to analyze the complex web of interactions within the Reddit social network, revealing a predominantly positive sentiment landscape interspersed with areas of contention. Through detailed graph analysis and the application of a GraphSAGE model, our model successfully mapped the intricate community dynamics and predicted interaction sentiments with high accuracy. The results underscore the effectiveness of the GNNs in capturing and interpreting the nuanced patterns of digital social interactions, offering valuable insights for community moderators, social scientists, and policymakers interested in fostering positive online environments.

References

[1] Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 933–943. International World Wide Web Conferences Steering Committee, 2018.

[2] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1269–1278. ACM, 2019.

[3] Srayan Datta, Chanda Phelan, and Eytan Adar. Identifying misaligned inter-group links and communities. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW), dec 2017.

[4] Justin Cheng, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Antisocial behavior in online discussion communities. *Proceedings of the International AAAI Conference on Web and Social Media*, 9, 04 2015.