Preventing Brigading in Reddit

A PROJECT PROPOSAL

Submitted in partial fulfillment for the course

19 CSE 356 Social Network Analytics

School of Computing

Submitted by

Register No	Names of Students
<AM.EN.U4EAC22028>	<Heman Sakkthivel MS>
<AM.EN.U4EAC22010>	<Aswinth Narayan A>
<AM.EN.U4EAC22048>	<Nandakumar S M>

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

Amritapuri Campus (INDIA)

March 2025

Motivation

0.1 Background

Reddit is one of the largest social media platforms, with millions of users engaging in discussions, sharing content, and voting on posts and comments. It is organized into communities (subreddits) that cater to specific interests, topics, or ideologies.

Brigading is a form of coordinated behavior where a group of users manipulates Reddit's voting system or floods a subreddit with content to influence discussions, promote agendas, or harass others. This undermines the platform's integrity and disrupts genuine user interactions. Brigading can lead to:

- **Manipulation of Public Opinion**: Artificially boosting or suppressing content can skew perceptions.
- **Harassment and Toxicity**: Targeted brigading can create hostile environments for users.
- **Erosion of Trust**: Frequent brigading can reduce user trust in the platform's fairness and authenticity.

While Reddit has moderation tools and anti-brigading measures, they are often reactive, rule-based, and insufficient to handle sophisticated or large-scale brigading.

0.2 Problem Statement

Brigading is a growing problem on Reddit, and existing solutions are limited in their ability to:

- **Detect Coordinated Behavior**: Identifying groups of users working together is complex, especially when they use multiple accounts or subreddits.
- **Analyze Context**: Understanding the motivation behind brigading requires analyzing both user behavior and content.
- **Scale Effectively**: Reddit's massive user base and dynamic nature make real-time detection and prevention difficult.

0.3 Objective

The primary goal of this project is to build a system that detects and prevents brigading on Reddit using **graph-based modeling** and **deep learning techniques**. The specific objectives are:

- Model Reddit data as a graph, representing users, posts, comments, and subreddits as nodes and their interactions as edges.
- Detect coordinated groups using community detection algorithms.
- Identify anomalous behavior in voting, posting, or commenting patterns.
- Analyze the content and context of targeted posts using NLP techniques.
- Develop strategies to prevent brigading, such as flagging suspicious activity or notifying moderators.
- Evaluate the system's performance using metrics like precision, recall, and F1-score.

0.4 Scope of the Project

The project will focus on the following areas:

- **Data Collection**: Collect Reddit data (posts, comments, votes, user activity) using the Reddit API or publicly available datasets.
- **Graph Modeling**: Build a graph representation of Reddit data, incorporating users, posts, comments, and subreddits as nodes and interactions as edges.
- **Community Detection**: Use graph algorithms (e.g., Louvain, Girvan-Newman) and deep learning techniques to identify coordinated groups.
- **Anomaly Detection**: Detect unusual patterns in user behavior using statistical methods, machine learning, or deep learning.
- **Content Analysis**: Use NLP techniques (e.g., sentiment analysis, topic modeling) to analyze the content of posts and comments.
- **Prevention Strategies**: Develop tools to flag suspicious activity, notify moderators, or restrict brigading behavior.
- **Evaluation**: Test the system on historical Reddit data and measure its performance using appropriate metrics.

0.5 Organization of the Project Report

The project report will be organized as follows:

- **Introduction** - Provides an overview of the project, including the background, problem statement, objectives, and scope.
- **Literature Review** - Reviews existing research and solutions related to brigading detection and prevention.
- **Methodology** - Describes the data collection, graph modeling, community detection, anomaly detection, and content analysis techniques used in the project.
- **Implementation** - Details the implementation of the system, including the algorithms, tools, and frameworks used.
- **Results and Discussion** - Presents the results of the system's performance and discusses their implications.
- **Conclusion and Future Work** - Summarizes the findings, discusses limitations, and suggests directions for future research.

The report will follow a logical and coherent flow, starting with the problem definition and ending with actionable insights and future directions.