THE RANKING OF HATE SPEECH PROPAGATORS IN SOCIAL MEDIA USING A MULTIMODAL PROFILING FRAMEWORK

Weerakoon Achchige Suresha Nilmini Perera

218031V

Degree of Doctor of Philosophy

Department of Computer Science and Engineering

Faculty of Engineering

University of Moratuwa

Sri Lanka

August 2025

THE RANKING OF HATE SPEECH PROPAGATORS IN SOCIAL MEDIA USING A MULTIMODAL PROFILING FRAMEWORK

Weerakoon Achchige Suresha Nilmini Perera

218031V

Thesis submitted in partial fulfillment of the requirements for the degree

Degree of Doctor of Philosophy

Department of Computer Science and Engineering

Faculty of Engineering

University of Moratuwa

Sri Lanka

August 2025

DECLARATION

I declare that this is my own work and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 20th August 2025

The above candidate has carried out research for the PhD under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. G.I.U.S. Perera

Name of Co-Supervisor: Dr. (Ms) S. Ahangama

Signature of the Supervisor: Date: 21/08/2025

Signature of the Co-Supervisor: Date: 21/08/2025

DEDICATION

I dedicate this to my parents, husband, my daughter and son.

ACKNOWLEDGEMENT

I would like to extend my deepest gratitude to those who have supported and guided me throughout this journey.

My heartfelt thanks to my supervisor, Prof. GIUS Perera, for his exceptional guidance and constant encouragement. His patience and understanding have been invaluable, especially during occasional setbacks. I deeply appreciate his insightful contributions and encouragement, which provided direction in many situations. Furthermore, his exceptional knowledge and enthusiasm for research were inspiring and became a driving force for my motivation. I am also immensely grateful to my co-supervisor, Dr.(Ms) S. Ahangama, for her continuous support, invaluable knowledge, and unwavering encouragement. Her outstanding support and guidance throughout this journey helped me to reach this level.

I would like to thank the progress review evaluation team, including Dr. Charith Chitraranjan, Dr. Shehan Perera, and Dr. L. Ranathunga, for their insightful feedback, suggestions for improvement, and their appreciation and encouragement.

I would like to express my deep gratitude for the free education opportunities in Sri Lanka, which enabled me to pursue my studies from primary school to postgraduate level.

I am grateful to the staff of the Department of Computer Science and Engineering, Faculty of Engineering, University of Moratuwa, for sharing their knowledge and supporting my academic pursuits.

Special thanks to my cousin, Dr.(Ms) I. Udayamalee, for encouraging me during challenging times and providing me with the necessary articles throughout the study.

I am thankful to the Higher Education Expansion and Development (AHEAD) of Operation of the Ministry of Higher Education, which is funded by the World Bank, for providing the necessary funds to publish my papers.

I would also like to thank Mr. Duminda De Silva for motivating me to start this journey and for his continued encouragement. I also thank the entire staff of the Department of Computer Science, Faculty of Natural Sciences, The Open University of Sri Lanka.

Finally, I sincerely thank my family and friends for their continuous support and motivation.

ABSTRACT

Social media platforms have revolutionized communication by enabling the sharing of opinions, ideas, and beliefs at an unprecedented scale. However, the rapid growth of these platforms has also facilitated the widespread propagation of hate speech and harmful behaviour, posing both technological and societal challenges. Understanding hate speech propagators, their influence, reach, and impact is crucial for mitigating the spread of harmful content and fostering safer online communities. This research aims to address this challenge by developing a robust hate speaker ranking system to identify, analyze, and rank hate users based on their behaviour and influence. The proposed framework combines profile features, characteristics, and machine learning techniques to evaluate hate speech. The system provides a dynamic, multimodal ranking mechanism to rank hate users into three levels labelled as: HIGH, MEDIUM, and LOW, based on their level of influence. The framework is trained and validated on datasets from Twitter, ensuring scalability and adaptability across diverse environments. Results indicate that the system effectively identifies influential hate users with high accuracy, offering actionable insights for moderation teams. This study also addresses ethical considerations, ensuring fairness, reducing bias, and maintaining freedom of expression. The research highlights the unequal influence of hate speech propagators and emphasizes the importance of personalized strategies to combat their impact. By providing a scalable and dynamic platform for ranking hate users, this study contributes to the ongoing efforts to combat online hate, supporting moderation efforts and promoting a healthier digital environment.

Keywords: Hate Speech, Influence, Social Media Platforms, Social Network Analysis, User Ranking,

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
ACRA	Aggregation Consensus Rank Algorithm
AHP	Analytic Hierarchy Process
API	Application Programming Interface
AS	Audience Sentiment
BC	Betweenness Centrality
BFPTM	Big Five Personality Traits Model
CC	Closeness Centrality
CDFs	Cumulative Distribution Functions
CLC	Local Centrality with a Coefficient
DC	Degree of Centrality
DEV	Daily Engagement Velocity
EDA	Exploratory Data Analysis
EC	Eigenvector Centrality
EC-SLR	Exclusion Criteria Systematic Literature Review
EO	Engagements & Outreach
FB	FaceBook
FF	Forest Fire Model
flerCnt	Followers Count
flingCnt	Followings Count
GoSL	Government of Sri Lanka
hashCnt	Hashtags Count

HEV	Hourly Engagement Velocity
HITS	Hyperlink-Induced Topic Search
HMCDM	Hierarchical Multiple-Criteria Decision-Making
IB	Influencing Behavior
IC	Independent Cascade Model
IC-SLR	Inclusion Criteria Systematic Literature Review
ICCPR	International Covenant on Civil and Political Rights
IDC	In-Degree Centrality
imgUrl	Availability of the Profile Image URL
lang	Language
len	Length of the Tweet
LIWC	Linguistic Inquiry and Word Count
listCnt	Listed Count
loc	Location
LSTM	Long-Short Term Memory
menCnt	Mentions Count
MPR	Motif-based PageRank
MRC	Medical Research Council
mimgUrl	Availability of the User's Mini Profile Image URL
NB	Naive Bayes
NLTK	Natural Language Toolkit
NLP	Natural Language Processing
OF	Overall Footprint

OSN	Online Social Network
PR	PageRank
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
retweetCnt	Retweets Count
RNN	Recurrent Neural Network
SDS-PBRR	Statistic Dice Similarity-Based Probabilistic Binomial Regression and Ranking
sim	Similarity of the Tweets
SIR	Susceptible-Infected-Recovered Model
SMP	Social Media Platform
SMPs	Social Media Platforms
SMN	Social Media Network
SMUs	Social Media Users
SNA	Social Network Analysis
SPLICE	Structured Programming for Linguistic Cue Extraction
SLR	Systematic Literature Review
SVM	Support Vector Machine
tweetCnt	Tweets Count
tweetFavCnt	Tweet Favorites Count
UDHR	United Nations Universal Declaration of Human Rights
urlCnt	URLs Count
URL	Uniform Resource Locator
userFavCnt	The Number of Favorites the User Has Marked on Others' Tweets
VPNs	Virtual Private Networks

LIST OF APPENDICES