A Cross Platform Framework for Social Media Information Diffusion Analysis

H.M.M.Caldera

198132D

Doctor of Philosophy

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

November 2023

A Cross Platform Framework for Social Media Information Diffusion Analysis

H.M.M.CALDERA

198132D

Thesis submitted in partial fulfillment of the requirements for the degree
Doctor of Philosophy

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

November 2023

DECLARATION

I declare that this is my own work and this thesis does not incorporate without acknowledgment 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 acknowledgment is made in the text. Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature of the candidate: Date: 21.11.2023

Mr.H. M. M. Caldera

The above candidate has carried out research for the PhD thesis under my supervision.

Name of the supervisor: Prof. G. I. U. S. Perera

Signature of the supervisor: .... Date: 21.11.2023

Acknowledgements

First and foremost, I want to thank my supervisor, Prof. Indika Perera, for his unbeatable support throughout the study. He has taught me, both consciously and unconsciously, how well experimental research is done. I appreciate all his time, ideas, and recommendation for the AHEAD grant to help fund my Ph.D. study.

Next, I would like to thank AHEAD Grants for financial support through research and give special thanks to Dr. Lochandaka Ranathunga as research coordinator for all financial approvals. Dr. Shalinda Adhikari for winning the research grant. I won't complete this degree if you have not won the grant.

As a fellow researcher, Mrs. Nadeera Meedin provided unbelievable support for success during the last few years. She has fully supported my success in many aspects. She shares personal research experiences and motivates me a lot for success.

Further, I thank all the research fellows who worked in the "social media analytics research group" for helping me. Thanks to reviews (including internal ones) for providing valuable feedback.

All authors develop valuable content and publish it on the web. I have gained much knowledge from your videos, documents, research articles, etc. Thanks to all the online content authors. Thanks to Alexander T., Research gate, and other online portals for making many scientific articles publicly available. Further, all scientists who kept their research data as a public asset and let other fellow researchers explore their research.

Finally, I would like to thank my father, mother, wife, son, and all other family members for motivating me and staying with me during this challenging period.

Abstract

In the current digital era, social media platforms have emerged as one of the most effective channels for the diffusion of information. People may readily access and exchange information, news, and opinions from anywhere worldwide because of increasing social media usage.

Information diffusion across multiplex social media platforms is one of the most prominent research problems ever. Social media content generators diffuse information on multiplex social media platforms by targeting many objectives such as popularity, online presence, hate targets, and customer engagement. Regardless of the ”content” posted on social media platforms, evaluating the dissemination velocity of each piece of content published on those platforms is essential. It will help to get an overall picture of ”how it flows” throughout the social media platforms. Most social media platforms have a platform-specific algorithm for calculating the degree of information diffusion on those platforms. The main objective of this research was to develop a method to calculate the velocity of information diffusion across multiplex social media platforms.

Existing literature on information diffusion strategies, effects, and measurements was used to develop the proposed algorithm. The information diffusion velocity of social media influencers varies according to the content. The platform-specific algorithms for diffusion strength detection vary based on the platform. Somehow, these platform-specific algorithms influence the community to engage with the trending content. i.e., platforms support increasing the strength of information diffusion.

Conventional information diffusion algorithms were designed to measure content diffusion speed on a simplex social media platform, which might be content-specific. The missing dimension is ubiquitous nature. Hence, regardless of the platform, it is mandatory to calculate a ubiquitous information diffusion velocity over multiplex social media platforms.

Both structured information diffusion in a graph for diffusion in a closed network and unstructured patterns in an open-ended coarse-grained information diffusion model check the importance of information diffusion on multiplex social media platforms. Time is another critical factor in defining velocity. i.e., a time series of information diffusion provides a rich picture of information diffusion.

Event-driven architecture is a well-known software architectural approach that facilitates the implementation of microservice-based solutions. The suggested algorithm utilizes an event-driven architecture to manage the information flow by processing social media events. Eventually, this research uses the event-triggering process to understand how information is propagated through an event-driven microservice architecture.

Data science and artificial intelligence are being employed in information diffusion

studies. Understanding how information spreads and the variables and features that influence it is another crucial study area of this research. There are several techniques for studying information dissemination using artificial intelligence. Applying artificial intelligence to information diffusion studies might improve our knowledge of ”How information travels” and ”how to disseminate information” in various circumstances efficiently. The research used natural language processing to evaluate the textual content of the social media post. That is to find a general textual meaning given by the end-user reactions.

Event-driven architecture is one of the best possible for information diffusion analytics. Using event-driven architecture, data may be delivered in real-time to various analytics services, allowing for the speedy and effective processing of enormous amounts of data. This is especially true in today's data-driven world, when businesses and organizations must make quick, well-informed decisions based on real-time data. Because of its event-driven nature, it is also simple to interface with other systems and services, making it a highly adaptable and versatile option for information distribution analytics. Since the diffusion of information starts with an event's occurrence, it follows numerous steps to flow among the community. An event-driven micro-services architecture that uses artificial intelligence methods (like natural language processing to evaluate textual information) has been experimented with to propose a simple solution for this complex problem.

As per the research work, I can summarize the key findings. I have proposed a tree-structured diffusion tree that can explain how information flows through multiplex social networks. Under this multiplex context, I have experimented with multiple trees and a more robust graph that focused on the diffusion of information. The diffusion strength was based on the SIR model, and the time series analysis focused on how quickly information spread throughout the network. The proposed solution was tested in several real-world cases. Technique-specific tests like seasonality and autocorrelation were conducted to evaluate how the time-series model works in a graph context. Further tests like cohesiveness and robustness were tested, and the proposed algorithm achieved good robustness (an average of 75%) and cohesiveness (an average of 70%) in each case. The best experimental results show an average of more than 80% accuracy in any given instance, and it constructs the tree in less than a second. Most of the predicted values generated an average accuracy of around 70%.

In summary, social media platforms have emerged as prominent channels for information propagation within the contemporary digital landscape. Quantifying the velocity at which information propagates across diverse social networks presents a notable challenge in research. While algorithms tailored to specific platforms influence community engagement, a ”universal metric for information dissemination strength” is necessary across multiple social media platforms. The envisioned algorithm considers time series data, integrating structured and unstructured patterns during construction.

Keywords: Information diffusion analysis, Social Media Data Analytics, Graph Learning, Time series analysis, Event-driven micro-services, Artificial Intelligence, Natural Language Processing.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial intelligence

API Application Programming Interfaces

ARIMA Autoregressive Integrated Moving Average

BC Betweenness centrality

DLR Dynamic Linear Regression

EC Eigenvector centrality

ETS) Exponential Smoothing

FMTS Fixed model time series

LDA Latent Dirichlet Allocation

OMTS Open model time series

SNA Social network analysis

SNS Social Network Sites

STL Seasonal Decomposition of Time Series

SVM Support vector machine

TDC Total degree centrality

TFP Total-From-Partial