

UNDERSTANDING INFORMATION PROPAGATION IN SOCIAL MEDIA

NIMISH JOSEPH



**DEPARTMENT OF MANAGEMENT STUDIES
INDIAN INSTITUTE OF TECHNOLOGY DELHI**

JULY 2023

UNDERSTANDING INFORMATION PROPAGATION IN SOCIAL MEDIA

by

NIMISH JOSEPH

Department of Management Studies

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the



**INDIAN INSTITUTE OF TECHNOLOGY DELHI
NEW DELHI – 110016, INDIA**

July 2023

CERTIFICATE

This is to certify that the thesis titled '*Understanding Information Propagation in Social Media*', submitted by **Nimish Joseph** (2015SMZ8338), to the Indian Institute of Technology Delhi, for the award of the degree of Doctor of Philosophy, is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Arpan Kumar Kar

Professor, Department of Management Studies

Indian Institute of Technology Delhi

Date:

Dr. P Vigneswara Ilavarasan

Professor, Department of Management Studies

Indian Institute of Technology Delhi

Date:

ACKNOWLEDGEMENTS

There are many people whom I would like to thank for contributing to this thesis directly or indirectly. Firstly, sincere gratitude to my thesis supervisors Prof. Arpan Kumar Kar and Prof. P Vigneswara Ilavarasan, who groomed me, stood by my side, and helped me during each stage of this programme. Their experience, knowledge, work ethics, and commitment to research scholars are certain qualities I would love to emulate in my life. I would have never achieved this milestone had these people not been part of my journey.

My profound gratitude to SRC Chair Prof. M P Gupta for his constant guidance and support. I would also like to thank other committee members, Prof. Sanjay Dhir and Prof. Samaresh Das, for their valuable suggestions that helped me complete this work and think further from a broader perspective. I was lucky to be part of various projects during this tenure. I would love to thank my professors who helped me connect with multiple organizations viz LIRNEasia, bitgrit, Digital India Awards, and National Digital Literacy Mission. I also received a chance to interact and work with the organizers of IFIP I3E 2017, 2018 and ICEGOV 2017. Thank you to everyone who gave me various opportunities to learn and contribute.

Thank you to all faculty members from DMS who helped me during the PhD programme. I cannot forget the support of non-teaching staff members, visiting faculty, guest lecturers, RAs and interns. Thank you to the entire Research Scholars community in DMS, especially my department seniors and the MIS team. I cannot miss naming Ashish Rathore, Syed Mustafa, Anushruti Vagrani, Pooja Sarin, Purva Grover, Reema Aswani, Rana Kidwai,

Ravinder Kumar Verma, Harjit Singh, Saroj Bijarnia, Swapnil Sharma, Umar Mir, and Akash Katiyar. DMS would have been a lonely place without all of you.

Thank you to all my dear friends and comrades from campus – Vickey Nandal, Hardeep Singh, Ajay Pratap Singh Lodhi, Sayan Dev, Kasimuthumanian S, Vinay Saini, Dharmendra Jangra, Sajan Wahi, Soham Das, Sumallya Mokopadhyay, Aljo Alexander and Deepak Singh. I know I can count on these friends whenever I am in need. A big thank you to IIT Delhi Malayalee Association, residents, and staff of Shivalik Hostel.

As a representative of PhD students, I got a chance to serve the IIT Delhi community as a member of the Student Affairs Council and Board of Academic Programmes. Thank you to all IIT Delhi Deans, office staff, and other student representatives who supported me during various occasions. I am also deeply indebted to my alma mater and all my teachers. It is their blessings that is driving me every single day.

Lastly, but most importantly, I am happy to value the unconditional love and encouragement provided by my family, friends, and relatives. This was a journey of ups and downs. I must not have sailed it through without my younger brother Neeraj Joseph's support, and my parents Sheeba Joseph's and Joseph Varky's love.

In an ambiguous world with full of uncertainties, it is difficult to walk alone, accepting the social norms and not being discriminated. This work is dedicated to you, A.

NIMISH JOSEPH

ABSTRACT

Social Media (SM) platforms play a significant role in the sharing of information. The rapid surge of SM users has brought in broad interest amongst governments, individuals, businesses, and other organizations to understand the behaviours exhibited by these platforms. The information shared by a SM user receives attention from other users in views, likes, comments, and shares based on various parameters. This thesis helps readers understand some of the significant reasons behind information propagation. Twitter, a microblogging platform that connects around 330 million monthly users (Twitter, 2021), acts as one of the primary sources of user-generated content. It is extensively used to share information in the form of text data or media. The studies in this thesis use Twitter as the source of information and explain various gaps in the literature related to SM and information propagation.

To achieve reach, to organize masses, to educate people, to inform about an activity and to stop spreading rumours in SM, it is necessary to understand the reasons behind the sharing of information. Group behaviour related theories from sociology explain information sharing between individuals. This research explores the possibility of understanding these theories from a SM perspective. A set of individuals or SM users subscribing to particular information were treated as a group to explore the possibilities of associating with these traditional theories.

Hashtags are one of the significant features available in a SM platform. It plays a major role in the grouping of information. The convergence behaviour exhibited by the usage of

hashtags helps in identifying the information being shared on a particular topic. Study on hashtags demonstrated how SM analytics might assist in discovering debate themes from a pool of information being shared, as well as determining who the influencers are and how much they influence information dissemination.

The thesis also explains the role of close friends and acquaintances in the sharing of information. Identifying cliques and then learning the pattern of liking and retweeting helped in understanding the significance of close friends in information propagation. This thesis proves that it is not the close friends that play a significant role in information propagation; rather, it could be the acquaintances who play an important role in disseminating information. This work also explains the contagion behaviour of SM users and proves that the sentiments of content did not have a significant impact on the information virality.

This research also explores the role of tweets with images and information propagation. Using learning algorithms, tweets were classified as viral or non-viral. These tweets were further processed to determine the role of various parameters in making a tweet with an image go viral, as well as to predict the likelihood of a tweet going viral. Since the contribution of image parameters in the popularity of tweets was minimal, this work was extended to learn the features of popular images. Different image parameters were studied to identify the dominant colours, the relevance of the number of objects, and to perform emotion analysis.

This thesis contributes to the studies on information propagation in Social Media by explaining several perspectives related to hashtags, cliques, and images, as well as their contribution to contagion and convergence behaviour.

सार

सोशल मीडिया प्लेटफॉर्म सूचना साझा करने में महत्वपूर्ण भूमिका निभाते हैं। सोशल मीडिया उपयोगकर्ताओं की तीव्र वृद्धि ने इन प्लेटफॉर्मों द्वारा प्रदर्शित व्यवहारों को समझने के लिए सरकारों, व्यक्तियों, व्यवसायों और अन्य संगठनों के बीच व्यापक रुचि पैदा की है। एक सोशल मीडिया उपयोगकर्ता द्वारा साझा की गई जानकारी विभिन्न मापदंडों के आधार पर अन्य उपयोगकर्ताओं के विचारों, पसंद, टिप्पणियों और साझा करने पर ध्यान आकर्षित करती है। यह थीसिस पाठकों को सूचना प्रसार के पीछे के कुछ महत्वपूर्ण कारणों को समझने में मदद करती है। ट्विटर, एक माइक्रोब्लॉगिंग प्लेटफॉर्म जो लगभग 330 मिलियन मासिक उपयोगकर्ताओं को जोड़ता है (ट्विटर, 2021), उपयोगकर्ता-जनित सामग्री के प्राथमिक स्रोतों में से एक के रूप में कार्य करता है। इसका उपयोग टेक्स्ट डेटा या मीडिया के रूप में जानकारी साझा करने के लिए बड़े पैमाने पर किया जाता है। इस थीसिस में अध्ययन ट्विटर को सूचना के स्रोत के रूप में उपयोग करता है और सोशल मीडिया और सूचना प्रसार से संबंधित साहित्य में विभिन्न अंतरालों की व्याख्या करता है।

पहुंच हासिल करने, जनता को संगठित करने, लोगों को शिक्षित करने, किसी गतिविधि के बारे में सूचित करने और सोशल मीडिया में अफवाहें फैलाने से रोकने के लिए, जानकारी साझा करने के पीछे के कारणों को समझना आवश्यक है। समाजशास्त्र से समूह व्यवहार संबंधी सिद्धांत व्यक्तियों के बीच सूचना साझा करने की व्याख्या करते हैं। यह शोध इन सिद्धांतों को सोशल मीडिया परिप्रेक्ष्य से समझने की संभावना तलाशता है। विशेष जानकारी की सदस्यता लेने वाले व्यक्तियों या सोशल मीडिया उपयोगकर्ताओं के एक समूह को इन पारंपरिक सिद्धांतों के साथ जुड़ने की संभावनाओं का पता लगाने के लिए एक समूह के रूप में माना गया था।

हैशटैग सोशल मीडिया प्लेटफॉर्म में उपलब्ध महत्वपूर्ण सुविधाओं में से एक है। यह सूचनाओं के समूहीकरण में प्रमुख भूमिका निभाता है। हैशटैग के उपयोग द्वारा प्रदर्शित अभिसरण व्यवहार किसी विशेष विषय पर साझा की जा रही जानकारी

की पहचान करने में मदद करता है। हैशटैग पर अध्ययन से पता चला है कि कैसे सोशल मीडिया एनालिटिक्स साझा की जा रही जानकारी के पूल से बहस के विषयों की खोज करने में सहायता कर सकता है, साथ ही यह निर्धारित कर सकता है कि प्रभावित करने वाले कौन हैं और वे सूचना प्रसार को कितना प्रभावित करते हैं।

थीसिस जानकारी साझा करने में करीबी दोस्तों और परिचितों की भूमिका भी बताती है। समूहों की पहचान करने और फिर लाइक और रीट्वीट करने के पैटर्न को सीखने से सूचना प्रसार में करीबी दोस्तों के महत्व को समझने में मदद मिलती। यह थीसिस साबित करती है कि यह करीबी दोस्त नहीं हैं जो सूचना प्रसार में महत्वपूर्ण भूमिका निभाते हैं; बल्कि, यह परिचित ही हो सकते हैं जो सूचना प्रसारित करने में महत्वपूर्ण भूमिका निभाते हैं। यह कार्य सोशल मीडिया उपयोगकर्ताओं के संक्रामक व्यवहार को भी समझाता है और साबित करता है कि सामग्री की भावनाओं का सूचना वायरलिटी पर कोई महत्वपूर्ण प्रभाव नहीं पड़ा है।

यह शोध छवियों और सूचना प्रसार वाले ट्वीट्स की भूमिका का भी पता लगाता है। लर्निंग एल्गोरिदम का उपयोग करके, ट्वीट्स को वायरल या गैर-वायरल के रूप में वर्गीकृत किया गया था। किसी छवि वाले ट्वीट को वायरल करने में विभिन्न मापदंडों की भूमिका निर्धारित करने के साथ-साथ किसी ट्वीट के वायरल होने की संभावना का अनुमान लगाने के लिए इन ट्वीट्स को आगे संसाधित किया गया। चूंकि ट्वीट्स की लोकप्रियता में छवि मापदंडों का योगदान न्यूनतम था, इसलिए लोकप्रिय छवियों की विशेषताओं को जानने के लिए इस काम को बढ़ाया गया था। प्रमुख रंगों की पहचान करने, वस्तुओं की संख्या की प्रासंगिकता और भावना विश्लेषण करने के लिए विभिन्न छवि मापदंडों का अध्ययन किया गया।

यह थीसिस हैशटैग, क्लिक्स और छवियों से संबंधित कई दृष्टिकोणों के साथ-साथ संक्रामक और अभिसरण व्यवहार में उनके योगदान की व्याख्या करके सोशल मीडिया में सूचना प्रसार पर अध्ययन में योगदान देती है।

TABLE OF CONTENTS

Chapter 1	1
Introduction.....	1
1.1 Social Media	1
1.2 Impact of Social Media – Individuals, Businesses and Governments	4
1.3 Social Media and Information Propagation	9
1.4 Social Media and Information Systems Research	11
1.5 Motivation for the Research.....	12
1.6 Research Objectives.....	13
1.7 Twitter as a Social Media Platform	17
1.8 Structure of thesis	17
1.9 Chapter Summary	18
Chapter 2	19
Literature Review	19
2.1 Benefits of Social Media	19
2.2 Information Propagation.....	20
2.3 Virality of Social Media Posts.....	21
2.4 Role of Networks.....	21
2.5 Role of images in Social Media.....	22
2.6 Images and Information Propagation	23
2.7 Social Media Analytics.....	24
2.8 Classification and Clustering.....	26

2.9 Group Theories & Social Media.....	28
2.10 Chapter Summary	31
Chapter 3	32
Hashtags, information convergence & insights.....	32
3.1 Hashtags and information convergence.....	32
3.2 Social Media Analytics.....	35
3.3 #Internet Of Things	37
3.4 #Demonetization & #GST	55
3.5 Information Propagation and Hashtag Analysis	68
3.6 Chapter Summary	69
Chapter 4	70
Network attributes & Information virality	70
4.1 Information propagation and networks.....	70
4.2 Review of Literature	71
4.3 Method.....	76
4.4 Findings	82
4.5 Path Analysis	89
4.6 Discussion.....	91
4.7 Chapter Summary	95
Chapter 5	96
Images & Information propagation	96
5.1 Introduction.....	96
5.2 Review of Literature	98
5.3 Research Methodology	102

5.4 Results.....	109
5.5 Discussion.....	120
5.6 Chapter Summary	124
Chapter 6	126
Summary, limitations, and future work	126
6.1 Introduction	126
6.2 Theoretical Contributions	129
6.2.1 Hashtag analysis, Information convergence and Insights	130
6.2.2 Role of Networks.....	132
6.2.3 Role of images	133
6.3 Practical Contributions	135
6.4 Limitations and future work	136
References.....	139
Publications from Thesis	156

TABLE OF FIGURES

Figure 1.1: Functional Building Blocks of SM	2
Figure 1.2: SM in the presence of content creators	7
Figure 1.3: Communications Model	8
Figure 1.4: Framework to build fan power	9
Figure 3.1. Broad overview of analysis possible in social media analytics.....	36
Figure 3.2. Activity vs visibility statistics of the top active and visible users.....	42
Figure 3.3. Relationship between hashtags and commonly used words.....	45
Figure 3.4. Word clouds based on polarity and emotions.....	48
Figure 3.5. Sentiment analysis on the various themes identified.....	50
Figure 3.6. Network generated for Topological Analysis based on the tweets related to IoT.....	51
Figure 3.7. Results of community analysis of profiles related to IoT tweets	53
Figure 3.8 Most frequently used hashtags for Demonetization and GST.....	59
Figure 3.9. Wordcloud for demonetization.....	61
Figure 3.10. Wordcloud for GST	61
Figure 3.11. Polarity of words.	64
Figure 3.12. Polarity of Tweets.....	65
Figure 3.13. Sentiment Analysis by emotion.....	66
Figure 3.14. Sentiment Analysis	67
Figure 4.1. Theoretical model explored.....	75
Figure 4.2. Methodology followed in the study.....	78
Figure 4.3: Average sentiment score for each user.....	85

Figure 4.4. Cliques identified for a user	86
Figure 4.5: Path Diagram for cliques, reputation, and popularity	90
Figure 5.1. Overview to research methodology and stages in analysis	104
Figure 5.2: Feature importance for Decision Trees	112
Figure 5.3: Feature importance for Random Forest.....	112
Figure 5.4: Likelihood of joy	120

TABLE OF TABLES

Table 1.2: Image Virality Detection Parameters.....	16
Table 3.1. Term adjacency matrix for topic identification	45
Table 3.2. Hashtag adjacency matrix for topic identification	46
Table 3.3. Centrality Analysis	52
Table 3.4. User Statistics	57
Table 4.1: Pseudocode for identification of cliques.....	81
Table 4.2: Relationship between the number of cliques and average retweets	87
Table 4.3: Relationship between cliques and average favorite count.	87
Table 4.4: Significance of the presence of cliques	88
Table 4.5: Significance of user reputation on average favorite count	89
Table 4.6: Significance of user reputation on average retweet count	89
Table 5.1: Data for the unique users	111
Table 5.2: Contribution of each parameter in making a tweet popular	113
Table 5.3: Confusion Matrix - Decision Trees and Random Forest	114
Table 5.4: Confusion Matrix for Neural Networks.....	114
Table 5.5. Precision and Recall.....	114
Table 5.6: Initial Cluster Centroids.....	116
Table 5.7: Final Cluster Centroids	116
Table 5.8: Dominant colors.....	116
Table 5.9 Anova test to identify the significance of clusters identified.....	117

Table 5.10: Impact of human presence on image popularity	118
Table 5.11: Impact of number of objects in an image	119
Table 5.12: Impact of emotions on the number of retweets	120

ABBREVIATIONS

SM	Social Media
GST	Goods and Services Tax
IS	Information Systems