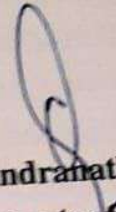


Dedicated to my beloved parents.

# CERTIFICATE

It is certified that the work contained in the thesis titled **Models to Find Influential Users for Viral Marketing in Multilayer Networks** by **Katakamsetty Venkatakrishna Rao** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree. It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of Ph.D. Degree.

  
Supervisor  
Dr. Ravindranath Chowdary C.  
Dept. of Computer Science & Engg.

पर्यवेक्षक/Supervisor  
संगणक विज्ञान एवं अभियंत्रण विभाग  
Department of Computer Science & Engg.  
Varanasi - 221005  
भारतीय प्रौद्योगिकी संस्थान  
Indian Institute of Technology  
(काशी हिन्दू विश्वविद्यालय)  
(Banaras Hindu University)  
वाराणसी/Varanasi-221005

## DECLARATION

I, **Katakamsetty Venkatakrishna Rao**, certify that the work embodied in this thesis is my own bona fide work and carried out by me under the supervision of **Dr Ravindranath Chowdary C.** from July-2018 to December-2022, at the Department of Computer Science Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, theses, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 25-05-2023

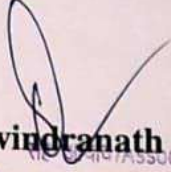
Place: Varanasi

K. Venkata Krishna Rao

(**Katakamsetty Venkatakrishna Rao**)

## CERTIFICATE BY THE SUPERVISOR

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

  
**Dr. Ravindranath Chowdary C.**

Associate Professor  
IIT(BHU), Varanasi  
भारतीय प्रौद्योगिकी संस्थान / Indian Institute of Technology  
(बनारस हिन्दू यूनिवर्सिटी) / (Banaras Hindu University)

  
**Signature of Head of Department/Coordinator of School**

अधीनस्थ विभागाध्यक्ष  
Professor & Head  
समणक विज्ञान एवं अभियांत्रिकी विभाग  
Department of Computer Sc. & Engg  
भारतीय प्रौद्योगिकी संस्थान  
Indian Institute of Technology  
(बनारस हिन्दू यूनिवर्सिटी)  
(Banaras Hindu University)  
पता-221005 / Varanasi-221005

# COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: **Models to Find Influential Users for Viral Marketing in Multilayer Networks**

Name of Student: **Katakamsetty Venkatakrishna Rao**

## Copyright Transfer

The undersigned hereby assigns to the Indian Institute of Technology (Banaras Hindu University), Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the Doctor of Philosophy.

Date: 25-05-2023

Place: Varanasi

*K. Venkata Krishna Rao*

**(Katakamsetty Venkatakrishna Rao)**

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

# Acknowledgments

This thesis is the result of the research carried out since August 2018 as part of my Ph.D. degree. During this tenure, I have associated with many people and express my sincere gratitude for their support and help. I express my sincere thanks and gratitude to my supervisor Dr. Ravindranath Chowdary C, who consistently supported and helped me. I am deeply grateful for his invaluable advice, for giving me the liberty to discuss my ideas, and for creating an exciting work environment. He was instant to listen my troubles and tried his best to help me out. Without his support, it would not be possible to finish my degree.

I would also like to thank my Research Progress Evaluation Committee (RPEC) members, Dr. Ravi Shankar Singh and Dr. Kishor P. Sarawadekar, for their guidance, evaluation of my research work, and suggestions at all levels. The knowledge and wisdom they have imparted to me during this tenure greatly help my career. I also acknowledge the Ministry of Education, Government of India, for providing fellowship to carry out the Ph.D. program at IIT Varanasi. It helped me to continue my research work with peace of mind without financial hurdles. I also thank my fellow lab mates, friends, and family members for their support and help.

I convey special thanks to my mentor and guardian Shri. Ram Madhav Varanasi, who enquired about my research progress and encouraged me all the time. He always guided me to work hard with dedication and commitment. He also gave me lots of motivation and encouragement to achieve the goal. It's always been an amazing and enlighting experience to work under him. He has always been very instrumental in my career. It is always my pleasure and wishes to work with him.

I thank my parents, my brother, and his family for their support and love for finishing my Ph.D. They always heard my pain when everyone ignored me and gave their support and inspiration. I will never be able to repay them for the love they showed me growing up. Finally, I am forever indebted to my wife, Uma Venkata Maheswari, and my son Yuvin Sriram Krishna for their unconditional selflessness and endless support. Completing this Ph.D. would not have been possible without their support, and I will be thankful to them forever.

Date: 18-05-2023

*K. Venkata Krishna Rao*  
Katakamsetty Venkatakrishna Rao

# List of Tables

2.1 Classification of various IM algorithms. . . . .	16
3.1 Notations used in CIM. . . . .	20
3.2 Properties of Datasets . . . . .	26
4.1 Notations used in this chapter . . . . .	36
5.1 Notations used in this chapter. . . . .	52
6.1 Notations used in this chapter . . . . .	74
6.2 Influence spread (IS) and Running time (RT in sec) on various datasets. . . . .	80

# List of Figures

1.1	Overview of IM model	3
1.2	Thesis organization	9
2.1	Benchmarking framework for IM [3]	13
2.2	Seed selection process in various IM frameworks [67]	15
3.1	Illustration of identifying five seed nodes from each layer after finding maximal cliques.	23
3.2	Illustration of ignoring noted nodes for influence maximization.	24
3.3	(a) Influence spread of 50 seed nodes for different algorithms under IC model on Synthetic Network before ignoring noted nodes (b) Influence spread of 50 seed nodes for different Algorithms under IC model on Synthetic Network after ignoring the noted nodes.	28
3.4	(a) Running time of different algorithms for finding 50 seed nodes in Synthetic Network. (b) Running time of IC Model for influence maximization on Synthetic Network. (c) Running time of IC Model on Synthetic Network after ignoring noted nodes.	29
3.5	(a) Influence spread of 50 seed nodes for different algorithms under IC model on Celegans multiplex GPI network. (b) Influence spread of 50 seed nodes for different algorithms under IC model on Celegans multiplex GPI network after ignoring noted nodes.	30
3.6	(a) Influence spread of 50 seed nodes for different algorithms under IC model on Drosophila multiplex GPI network for influence maximization. (b) Influence spread of 50 seed nodes for different algorithms under IC model on Drosophila multiplex GPI network after ignoring <i>noted nodes</i> for influence maximization.	30
3.7	(a) Influence spread of 50 seed nodes for different algorithms under IC model on Homo multiplex GPI network for influence maximization. (b) Influence spread of 50 seed nodes for different algorithms under IC model on Homo multiplex GPI network after ignoring noted nodes for influence maximization.	31

3.8	(a) Running time of different algorithms for finding 50 seed nodes in Celegans multiplex GPI network (b) Running time of IC Model on Celegans multiplex GPI network for influence maximization. (c) Running time of IC Model on Celegans multiplex GPI network after ignoring <i>noted nodes</i> for influence maximization.	32
3.9	(a) Running time of different algorithms for finding 50 seed nodes in Drosophila multiplex GPI network (b) Running time of IC Model on Drosophila multiplex GPI network for influence maximization. (c) Running time of IC Model on Drosophila multiplex GPI network after ignoring <i>noted nodes</i> for influence maximization.	33
3.10	(a) Running time of different algorithms for finding 50 seed nodes in Homo multiplex GPI network (b) Running time of IC Model on Homo multiplex GPI network for influence maximization. (c) Running time of IC Model on Homo multiplex GPI network after ignoring <i>noted nodes</i> for influence maximization.	34
4.1	(A) Shows the two graphs in layer 1 and layer 2. (B) Shows the different identified communities in layer and layer 2, red nodes are seed nodes. (C) Red nodes are influenced nodes i.e. active nodes and remaining are inactive nodes.	38
4.2	(a) Number of influenced nodes for various algorithms using IC model on Celegans Multiplex GPI network dataset.	43
4.3	(a) Number of influenced nodes for various algorithms using IC model on Arxiv netscience multiplex dataset. (b) Number of influenced nodes for various algorithms using IC model on Homo multiplex GPI network dataset.	44
4.4	(a) Number of influenced nodes for various algorithms using IC model on Moscow Athletics 2013 dataset. (b) Number of influenced nodes for various algorithms using IC model on Newyork climate change 2014 dataset.	44
4.5	(a) Running time of various algorithms for identifying 50 seed nodes on Celegans Multiplex GPI network. (b) Running time of various algorithms for information spread on Celegans Multiplex GPI network using IC model.	45
4.6	(a) Running time of various algorithms for identifying 50 seed nodes on Homo multiplex GPI network dataset. (b) Running time of various algorithms for information spread on Homo multiplex GPI network dataset using IC model.	45
4.7	(a) Running time of various algorithms for identifying 50 seed nodes on Arxiv netscience multiplex dataset. (b) Running time of various algorithms for information spread on Arxiv netscience multiplex dataset using IC model.	46
4.8	(a) Running time of various algorithms for identifying 50 seed nodes on Moscow Athletics 2013 dataset. (b) Running time of various algorithms for information spread on Moscow Athletics 2013 dataset using IC model.	46



4.9	(a) Running time of various algorithms for identifying 50 seed nodes on Newyork climate change 2014 dataset. (b) Running time of various algorithms for information spread on Newyork climate change 2014 dataset using IC model. . . . .	47
4.10	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Sanremo music festival 2016 dataset. (b) Execution time for various algorithms to find 50 seed users on Sanremo music festival 2016 dataset.	47
4.11	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Moscow Athletics 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Moscow athletics 2013 dataset. . . . .	48
4.12	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Newyork climate change 2014 dataset. (b) Execution time for various algorithms to find 50 seed nodes on Newyork climate change 2014 dataset.	48
4.13	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Martin Luther King’s 50th anniversary 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Martin Luther King’s 50th anniversary 2013 dataset. . . . .	49
4.14	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Cannes film festival 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Cannes film festival 2013 dataset. . . . .	49
5.1	Overview of CBIM model. . . . .	53
5.2	(A) Weight distribution to the edges for node 1 and EWS for node 1 is 0.12622 (B) Weight distribution to the edges for node 2 and EWS for node 2 is 0.4622 .	59
5.3	Influence spread of 50 seed nodes for different similarities with our proposed CBIM algorithm under various datasets. . . . .	63
5.4	(a) Influence spread of 50 seed nodes for different algorithms using IC model on London Multiplex Transport Network. (b) Influence spread of 50 seed nodes for different algorithms using IC model on Arxiv netscience multiplex for influence maximization. . . . .	63
5.5	(a) Influence spread of 50 seed nodes for different algorithms using IC model on Celegans multiplex GPI network. (b) Influence spread of 50 seed nodes for different algorithms using IC model on Homo multiplex GPI network for influence maximization. . . . .	64
5.6	(a) Influence spread of 50 seed nodes for different algorithms using IC model on Moscow Athletics 2013. (b) Influence spread of 50 seed nodes for different algorithms using IC model on Newyork climate change 2014 for influence maximization. . . . .	64

5.7	(a) Influence spread of 50 seed nodes for different algorithms using LT model on Celegans multiplex GPI network. (b) Influence spread of 50 seed nodes for different algorithms using LT model on Homo multiplex GPI network for influence maximization. . . . .	65
5.8	(a) Influence spread of 50 seed nodes for different algorithms using LT model on Moscow Athletics 2013. (b) Influence spread of 50 seed nodes for different algorithms using LT model on Newyork climate change 2014 for influence maximization. . . . .	66
5.9	(a) Execution time of different algorithms for finding seed nodes under Celegans multiplex GPI network. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in Celegans multiplex GPI network. (c) Execution time of different algorithms for spreading of 50 seed nodes using LT model in Celegans multiplex GPI network. . . . .	66
5.10	(a) Execution time of different algorithms for finding seed nodes under Homo multiplex GPI network. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in Homo multiplex GPI network. (c) Execution time of different algorithms for spreading of 50 seed nodes using LT model in Homo multiplex GPI network. . . . .	67
5.11	(a) Execution time of different algorithms for finding seed nodes under Moscow Athletics 2013. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in Moscow Athletics 2013. (c) Execution time of different algorithms for spreading of 50 seed nodes using LT model in Moscow Athletics 2013. . . . .	67
5.12	(a) Execution time of different algorithms for finding seed nodes under Newyork climate change 2014. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in Newyork climate change 2014. (c) Execution time of different algorithms for spreading of 50 seed nodes using LT model in Newyork climate change 2014. . . . .	67
5.13	(a) Execution time of different algorithms for finding seed nodes under Arxiv netscience multiplex. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in Arxiv netscience multiplex. . . . .	69
5.14	(a) Execution time of different algorithms for finding seed nodes under London Multiplex Transport Network. (b) Execution time of different algorithms for spreading of 50 seed nodes using IC model in London Multiplex Transport Network. . . . .	69
5.15	(a) Number of influenced people for 50 seed users for various algorithms using IC model on Sanremo music festival 2016 dataset. (b) Execution time for various algorithms to find 50 seed users on Sanremo music festival 2016 dataset. . . . .	70

5.16 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Moscow Athletics 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Moscow athletics 2013 dataset. . . . .	70
5.17 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Newyork climate change 2014 dataset. (b) Execution time for various algorithms to find 50 seed nodes on newyork climate change 2014 dataset.	71
5.18 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Martin Luther King’s 50th anniversary 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Martin Luther King’s 50th anniversary 2013 dataset. . . . .	71
5.19 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Cannes film festival 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Cannes film festival 2013 dataset. . . . .	72
6.1 Network is divided into two communities C1 and C2. . . . .	75
6.2 (a) Execution time for various algorithms to find 50 seed users on Sanremo music festival 2016 dataset. (b) Number of influenced people for 50 seed users for various algorithms using IC model on Sanremo music festival 2016 dataset.	81
6.3 (a) Execution time for various algorithms to find 50 seed users on Moscow athletics 2013 dataset. (b) Number of influenced people for 50 seed users for various algorithms using IC model on Moscow Athletics 2013 dataset. . . . .	81
6.4 (a) Execution time for various algorithms to find 50 seed nodes on Newyork climate change 2014 dataset. (b) Number of influenced people for 50 seed users for various algorithms using IC model on Newyork climate change 2014 dataset.	82
6.5 (a) Execution time for various algorithms to find 50 seed users on Martin Luther King’s 50th anniversary 2013 dataset. (b) Number of influenced people for 50 seed users for various algorithms using IC model on Martin Luther King’s 50th anniversary 2013 dataset. . . . .	83
6.6 (a) Execution time for various algorithms to find 50 seed users on Cannes film festival 2013 dataset. (b) Number of influenced people for 50 seed users for various algorithms using IC model on Cannes film festival 2013 dataset. . . . .	83
6.7 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Sanremo music festival 2016 dataset. (b) Execution time for various algorithms to find 50 seed users on Sanremo music festival 2016 dataset.	84
6.8 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Moscow Athletics 2013 dataset. (b) Execution time for various algorithms to find 50 seed users on Moscow athletics 2013 dataset. . . . .	84
6.9 (a) Number of influenced people for 50 seed users for various algorithms using IC model on Newyork climate change 2014 dataset. (b) Execution time for various algorithms to find 50 seed nodes on Newyork climate change 2014 dataset.	85

6.10 (a) Number of influenced people for 50 seed users for various algorithms using	
IC model on Martin Luther King's 50th anniversary 2013 dataset. (b) Execution	
time for various algorithms to find 50 seed users on Martin Luther King's 50th	
anniversary 2013 dataset. . . . .	85
6.11 (a) Number of influenced people for 50 seed users for various algorithms using	
IC model on Cannes film festival 2013 dataset. (b) Execution time for various	
algorithms to find 50 seed users on Cannes film festival 2013 dataset. . . . .	86

# List of Abbreviations

<b>IM</b>	Influence Maximization
<b>IC</b>	Independent Cascade
<b>LT</b>	Linear Threshold
<b>NOMC</b>	Number of Maximal Cliques
<b>SOC</b>	Size of Clique
<b>CIM</b>	Clique based Influence Maximization
<b>EWS</b>	Edge Weight Sum
<b>FIC</b>	Finding Initial Communities
<b>FC</b>	Final Communities
<b>DSC</b>	Dice Similarity Coefficient
<b>SN</b>	Seed node
<b>DIC</b>	Detecting initial communities
<b>MSC</b>	Merging small communities
<b>SIM</b>	Similarity based Influence Maximization
<b>CSC</b>	cut similarity community
<b>PCI</b>	Power community Index