Dedicated

to

my family,

friends, teachers, and

Lord Shiva.

CERTIFICATE

It is certified that the work contained in this thesis entitled "**Multifeature Analysis based Link Prediction and Influence Maximization in Dynamic Social Networks**" being submitted by "**Ashwini Kumar Singh**" to the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, is a record of bonafide research work carried out under my supervision and guidance. The thesis, in my opinion, is worthy of consideration for the award of the degree of Doctor of Pholosophy of the institute. To the best of my knowledge, the results embodied in this thesis have not been submitted to any other University or Institute for the award of any other Degree or Diploma. It is further certified that the student has fulfilled all the requirements of Comprehensive Examination, Candidacy and SOTA for the award of PhD. Degree.

Date: 15th July 2021

Place: **VARANASI**

Signature of Supervisor

Dr. Lakshmanan Kailasam

Assistant Professor

Department of Computer Science and Engineering

Indian Institute of Technology (BHU), Varanasi, India

DECLARATION

I, **Ashwini Kumar Singh**, certify that the work embodied in this thesis is my own bonafide work and carried out by me under the supervision of **Dr. Lakshmanan Kailasam** from July-2017 to July-2021, at the Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree or 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 others' work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertations, thesis, etc., or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 13-07-2021
Place: **VARANASI**

Signature of Student
(Ashwini Kumar Singh)

CERTIFFICATE BY THE SUPERVISOR

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

Date: 13 - 07 - 2021
Place: **VARANASI**

Signature of Supervisor
(Dr. Lakshmanan Kailasam)

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

Signature of Head of Department/Coordinator of School

आचार्य व विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg.
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)
वाराणसी–२२१००५/Varanasi-221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the thesis: **Multifeature Analysis based Link Prediction and Influence Maximization in Dynamic Social Networks**
Name of Student: **Ashwini Kumar Singh**

COPYRIGHT TRANSFER

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

Date: 13-07-2021
Place: **VARANASI** (**Ashwini Kumar Singh**)

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

Acknowledgements

I would like to take this opportunity to thank everybody who inspired me, helped me and contributed directly or indirectly in realizing this thesis. I express my deep gratitude and profound regard to my supervisor Dr. Lakshmanan Kailasam, Assistant Professor, Department of Computer Science and Engineering, IIT (BHU), Varanasi for his exemplary guidance, monitoring and constant encouragement throughout the course of this thesis. The meaningful discussions with him not only helped me in culminating the problems of this thesis but also helped me to understand the nuances of doing research. I would like to say thank for all his support from the core of my heart.

I would like to express my sincere thanks to Prof. Sanjay Kumar Singh, Head of, Computer Science and Engineering Department, for his kindness and valuable support in carrying out the research. I am enormously grateful to all the members of my Research Program Evaluation Committee, Dr. Bhaskar Biswas and Dr. Lavanya Selvaganesh, and DPGC Convener Dr. Prateek Chottopadhyay for their valuable suggestions, appreciation, and encouragement. I would like to convey my sincere gratitude to the other faculty members of the Department of Computer Science and Engineering, Prof.(Rtd.) R. B. Mishra, Prof. K.K Shukla, Prof. A.K Tripathi, Prof. Rajeev Srivastava, Dr. Sukomal Pal, and Dr. Ruchir Gupta for their extensive and inspiring guidance throughout this tenure.

A very special thanks go out to my colleagues and friends with special mention to Dr. Tribikram Pradhan, Mr. Sushant Kumar Pandey, Mr. Ashish Kumar Sharma, Mr. Nigmendra Pratap Yadav, and Mr. Ashish Ranjan for being great friends and the best advisors I could ever have. Their advice, encouragement, and critics were always a source of inspiration. This thesis would not have been possible without their invaluable remarks and persistent help. My heart lied thanks to my lab mates Mr. Vaibhav Padhye, Ms. Sheetal Arya, Mr. Sumit Kumar, Mr. Saurabh Arora, and Mr. Vipin Maurya. We really enjoyed our discussions, which spanned on every academics, social, and political topic. I wish to extend my gratitude towards my friends from outside this institute, particularly Dr. Dilip Verma, Dr. Vijay Singh, Dr. Avadh Kishor, Dr. Narayan Chaturvedi, Dr. Arun Chauhan, Mr. Ajay Singh, Dr. Avinash Pandey, Mr. Abhishek Agarwal, Mr. Devesh Rai, Mr. Devesh Sumeet, Mr. Abhishek Singh, Mr. Abhishek Yadav, Mr. Anil Yadav, and Mrs. Sneha Yadav.

I extend special thanks to the non-teaching staff in the department, particularly, Mr. Ravi Kumar Bharti, Mr. Ritesh Singh, Mr. Shubham Pandey, Mr. Prakhar Kumar, Mr. Manoj Kumar Rai, Mr. Viplav Biswas, and Mr. Akhilesh Kumar Pal.

My acknowledgments would not be complete unless I express my heartfelt gratitude and deepest appreciation to all my family members who continuously supported me throughout this journey. I would like to thank my mother Smt. Tara Devi, and my deceased father Late Daulat Singh, who continues to be a great inspiration for me. I would like to thank them for their numerous sacrifices and efforts for my studies. I would like to thank my brother Mr. Ajit, and sister Mrs. Archna, for their continuous support and encouragement. I would like to express my heartfelt appreciation to my wife, Renu, for her patience, tolerance, and sacrifices during this path. Thanks to my cute little daughter Yashika (Kuhu) and my son Yakshit (Vashu) for making me stress-free with their lovely smiles. Additionally, I would like to thank my mother and father-in-law for their continuous encouragement and support during this journey.

Finally, I am grateful to God for his blessings and for giving me the strength to persevere throughout the long and arduous journey towards a Ph.D.

Ashwini Kumar Singh
Varanasi, India.

Abstract

Network science has emerged as a fast-expanding research area in the last decade and has brought significant advances to our knowledge about the modern science of graphs. It includes the study of social networks that have gained attention from researchers due to the abundance of its data on the web. The rapid increase in the number of users to the social platforms (such as Facebook, Twitter, Instagram, and other blogs, dating sites, friends making sites) provided by the web has shown unseen human relationships and motivated the researchers to use this to extract meaningful information. This thesis deals with the two important challenges of social network analysis (also of dynamic/complex networks analysis): *Influence Maximization and Link Prediction*.

Influence Maximization is the problem of finding a small set of highly influential users in the social networks. The influence spreads according to an explicit influence propagation model. Influence Maximization is an essential component in many applications such as Network Monitoring and Viral Marketing. In this thesis, we study the Influence Maximization problem in a social network that evolves with time and proposes two new frameworks: *Link Prediction based Influential Node Tracking (LPINT)*, and *Multifeature based Influential Nodes Tracking (MINT)*.

Link prediction aims to predict the missing interactions in evolving networks that may appear in the future. It has practical importance in various real-world applications, ranging from friendship recommendation, knowledge graph completion, target advertising, and protein-protein interaction prediction. In this thesis, we present two models for link prediction in dynamic social networks. The first model uses conditional temporal Restricted Boltzmann Machine for predicting the links that may appear in the network by considering the evolutionary networks' temporal and structural patterns. The second model presents a modified Latent Dirichlet Allocation and Hidden Naive Bayesian (HNB)-based link prediction technique named *Popularity, interests, location used hidden Naive Bayesian-based (PILHNB) model* for link prediction in dynamic social networks by considering behavioural controlling elements like relationship network structure, nodes' attributes, location-based information of nodes, nodes' popularity, users' interests, and learning the evolution pattern of these factors in the networks. Extensive experiments are performed over various real social network datasets to demonstrate the effectiveness of the proposed methods over the existing ones.

Keywords: Online Social Networks, Influence Maximization, Link Prediction, Latent Dirichlet Allocation, Hidden Naive Bayesian, Restricted Boltzmann Machine.

Contents

List of Figures

List of Tables

Symbols

List of Symbols for Chapter 3

$G(\mathbb{V}, E)$	a simple graph
$\mathbb{V}$	the vertex set
E	the edge set
$\rho = \{G_t\}_1^T$	set of snapshots of an online social network
$G^t = (\mathbb{V}, E_t)$	a snapshot of ρ at time t
E^t	the edge multiset of G_t
S	the seed set
S_t	the seed set at timestamp t
k	the size of seed set
$\sigma(S)$	expected number of nodes influenced by S
$p_{u,v}^G$	the propagation probability between node u and v in snapshot G
N	window size
W_A	weight matrix of ctRBM at time $t-1$
θ_m	parameter of ctRBM model R_m
L_m	ctRBM for node m
$\mathbb{L}$	collection of ctRBMs for all nodes
η_m^t	neighbours impact on node m at time t
V	visible layer in RBM
$\hat{p}$	total number of nodes in the visible layer
H	hidden layer in RBM
$\widetilde{V}$	reconstructed data (model's estimation)

$N_{\bar{V}}$	number of visible variables in ctRBM
$N_{\bar{H}}$	number of hidden variables in ctRBM
β	hyperparameter balancing temporal variations
γ	hyperparameter controlling node neighbours impact
x	bias for visible layer $\bar{V}$ in RBM
$\bar{x}^t$	adaptive bias of V at t for ctRBM model
y	bias for hidden layer $\bar{H}$ in RBM
$\Delta_{e,e_s}(S_t)$	the replacing gain of changing from e_s to e

List of Symbols for Chapter 4

G^t	a snapshot of graph at time t.
V^t	the vertex set at time t
E^t	the edge set at time t
A^t	the attribute vector at time t
n_t	number of nodes/users at time t
a^i	attribute information of i^{th} node
L^t	location vector at time t
d_{v_i}	degree of node v_i
$\mathscr{P}^t$	popularity vector
$\mathscr{P}_{v_i}$	popularity of node v_i
Ω	user behaviour pattern distribution
I^t	interest vector
$\mathscr{S}_{v_i}$	attribute similarity vector of node v_i
c_{ij}	common neighbour vector
α	individual dependency
β	combined dependency
k	represent the existance of links
$\oplus$	superposition operator

List of Symbols for Chapter 5

G^t	a snapshot of graph at time t.
V^t	the vertex set at time t
E^t	the edge set at time t
D^t	text documents associated with nodes at time t
n_t	number of nodes/users at time t
L^t	location vector at time t
l_i	check-in information of i^{th} user
$N(v_i)$	set of neighbours of user v_i
$N_{out}(u)$	set of outgoing neighbours of user u
$N_{in}(u)$	set of incoming neighbours of user u
$d_{v_i}(t)$	degree of node v_i at time t
$\mathscr{P}^t$	popularity vector at time t
$\mathscr{P}_{v_i}$	popularity of node v_i
z	topic distribution
$\mathscr{T}$	number of topics
S_m	seed set for message m
k	size of seed set
$\sigma(S)$	influence spread by seed set S

Chapter 1

Introduction

This chapter establishes the key concepts and vocabulary used in the rest of the thesis. We begin with a general introduction to the network science in Section 1.1, followed by a brief description of social networks, link prediction, and influence maximization in sections 1.2, 1.4, and 1.5, respectively. In Section 1.6, the general limitations of existing methods of link prediction and influence maximization in online social networks are illustrated. In Section 1.7, we provide motivation for our work. Section 1.8 summarise the main contributions of the thesis. Finally, Section 1.9 presents the layout of the rest of the thesis.

1.1 Network Science

Network Science has been getting a lot of attention recently. Broadly, it deals with complex networks such as telecommunication networks, computer networks, biological networks, cognitive and semantic networks, and social networks, considering distinct elements or actors represented by nodes and the connections between the elements or actors as links. This field of research is interdisciplinary based on the concepts and algorithms of graph theory [1] from mathematics, statistical mechanics [2] from physics, data mining [3] and information visualization [4] from computer science, inferential modelling [5] from statistics, and social structure [6] from sociology.