UNDERSTANDING NETWORK STRUCTURE, DIFFUSION PROCESS, OPINION LEADERS, AND SENTIMENTS IN SOCIAL NETWORKS: A CASE STUDY IN HEALTHCARE

a Thesis submitted by

ABHINAV CHOUDHURY

(D15049)

for the award of the degree of Doctor of Philosophy

SCHOOL OF COMPUTING & ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY MANDI

January 22nd, 2021

THESIS CERTIFICATE

This is to certify that the work contained in the thesis entitled "Understanding network structure, diffusion process, opinion leaders, and sentiments in social networks: A case study in healthcare" being submitted by Mr. Abhinav Choudhury (Enroll. No: D15049) has been carried out under my supervision. In my opinion, the thesis has reached the standard fulfilling the requirement of regulation of the Ph.D. degree. The results embodied in this thesis have not been submitted elsewhere for the award of any degree or diploma.

January 22nd, 2021

Dr. Varun Dutt (Supervisor)
Associate Professor
School of Computing and Electrical Engineering
School of Humanities and Social Sciences
Indian Institute of Technology Mandi
Kamand, Himachal Pradesh, India
Email: varun@iitmandi.ac.in

Declaration by the Research Scholar

I hereby declare that the entire work embodied in this Thesis is the result of investigations carried out by me in the ***School of Computing and Electrical Engineering***, Indian Institute of Technology Mandi, under the supervision of ***Dr. Varun Dutt***, and that it has not been submitted elsewhere for any degree or diploma. In keeping with the general practice, due acknowledgments have been made wherever the work described is based on findings of other investigators.

Place: IIT Mandi, Kamand

Date: January 22nd, 2021 Name: Abhinav Choudhury

Declaration by the Research Advisor

I hereby certify that the entire work in this Thesis has been carried out by ***Abhinav Choudhury***, under my supervision in the ***School of Computing and Electrical Engineering***, Indian Institute of Technology Mandi, and that no part of it has been submitted elsewhere for any Degree or Diploma.

Signature: 

Name of the Guide: Dr. Varun Dutt

Date: January 22nd, 2021

ABSTRACT

India has a significant shortage of trained physicians nationwide with a physician to population ratio of only 1:1674. Thus, identifying critical physicians is imperative as the proper diffusion of medical information to these physicians is of utmost importance. To identify critical physicians, one first needs to understand the network structure of the physician's social network as well as the underlying dynamics of adoption. The main contribution of this thesis is to understand how the network structure and the underlying network dynamics affect the diffusion of innovation that takes place inside physician social networks, and to develop novel strategies for identifying critical physicians to accelerate the diffusion process. In the first experiment, a binary approach and a weighted approach was proposed for creation of physician social networks. These approaches relied upon the similarity between physician attributes to assign weights to relationships. Results indicated that the weighted approach had a higher accuracy compared to the binary approach in predicting a physician’s medicine adoption.

While network structures play a pivotal role in modeling information diffusion, one also needs to understand the underlying network dynamics. In the second experiment, the effect of network dynamics on the diffusion of innovation was investigated. First, innovation diffusion was analyzed from a time-series perspective. Diffusion (Roger’s and Bass model), statistical (seasonal autoregressive integrated moving average), and machine learning (linear regression and gradient boosting regression) models were developed for predicting growth in the number of adopters of pain medications. The best performance was obtained by the time-series models, followed by machine learning and diffusion models. Second, innovation diffusion process inside a physician's social network was analyzed by predicting information cascades using multi-layer perceptron (MLP) and long short-term memory (LSTM) neural

networks. A systematic evaluation of different graph embedding techniques and the effect of embedding dimensions on the prediction of information cascades was also performed. Results indicated that the embedding techniques that preserved both the first-order and second-order proximity performed better in cascade prediction compared to those that only preserved one of the proximity measures. Furthermore, MLPs performed better compared to LSTMs in predicting information cascades.

Next, the third experiment investigates the effect of communication channels like electronic word of mouth (eWOMs) communications (e.g., tweets) on the innovation diffusion process. A sentiment analysis of tweets of followers and non-followers of two rare disease medication manufacturers and their respective medications was performed to understand their effect on the overall perception. Results indicated a counter intuitive finding: there was no significant difference in the average sentiment of followers and non-followers regarding rare disease medications, indicating that followers may not possess positive sentiments.

Finding critical physicians may be summarized as an influence maximization (IM) problem, which aims to select a subset of physicians from an influence social graph, such that the diffusion of information is maximized. In the fourth experiment, a reinforcement learning (RL)-based framework for solving the IM problem was proposed. The RL framework consisted of an edge-based graph neural network (GNN) that generated the node embeddings, which were then fed to a double deep Q-network (DDQN) to learn a Q-function to predict the solution set. The edge based GNN was an ensemble architecture comprising of structure2Vec, multi-headed self-attention, and edge enhanced graph neural network (EGNN). The framework was trained on social graph with 20% of its edges randomly removed and tested on the whole graph. Results revealed that the framework was able to

generalize to an unknown graph and gave a spread difference of 8% compared to a heuristic IM algorithm implemented on the whole graph.

In the fifth experiment, the problem of online influence maximization (OIM) was investigated, where OIM is a variant of the IM problem. The objective of OIM is to identify critical physicians in the absence of influence probabilities in a social graph. A new explore-exploit ensemble approach based on the Exponential-weight, Exploration, and Exploitation using Expert advice (EXP4) algorithm was proposed for solving the OIM problem. Results indicated the ensemble approach performed better compared to other current OIM algorithms.

Lastly, in the sixth experiment, the problem of volume maximization (VM) and different algorithmic solutions to the VM problem were investigated. The objective of the VM problem was to select a set of physicians who are both influential as well as have frequent interactions with patients. For solving the VM problem, these two frameworks were proposed: a reinforcement learning framework that developed Q-learning and SARSA models; and an instance-based learning (IBL) framework that developed IBL models. A greedy based algorithm called weighted-CELF was also proposed for the VM problem. The proposed models were compared with multiple IM algorithms like PMIA. Results revealed that the weighted-greedy and weighted-CELF algorithm gave the best weighted influenced spreads while PMIA gave the best influence spread. In contrast, the RL-frameworks: Q-learning and SARSA gave good weighted influence spread, and influence spread different initial seed set sizes (k). This thesis highlights the utility of the proposed approaches in identifying critical physicians and shows the effect of network structure and the underlying dynamics on the diffusion of innovation in a physician's social networks.

Keywords: *Social network analysis, Network structure, Influence maximization, Online influence maximization, Healthcare, Reinforcement learning, EXP4, Q-learning, SARSA, Graph neural network, Sentiment analysis, Volume maximization, Instance based learning*

Acknowledgments

It is with immense gratitude and profound thanks that I acknowledge the support of my Ph.D. supervisor, Dr. Varun Dutt, in completing this thesis. I am incredibly grateful for his immense patience while dealing with me. His invaluable guidance, constructive feedback, and constant encouragement are the bases for the success of this research. I want to express my sincere gratitude towards him for welcoming me into his research team, providing me an excellent research atmosphere. His mentorship has significantly influenced, the way I understand and approach research work.

I have great pleasure in acknowledging my gratitude towards the co-chairs of my dissertation committee: Dr. Aditya Nigam, Dr. Sriram Kailasam, and Dr. Devika Sethi for generously offering their time, guidance, and goodwill throughout my Ph.D. duration. I would also like to acknowledge Dr. Samar Agnihotri and Dr. Bharat Singh Rajpurohit for their feedback on my work. Without their constant push, I would not have shaped my thesis in its current form.

I warmly thank my friend and colleague Shruti Kaushik for creating a cordial working environment and supporting me in difficult times. I am also thankful to my fellow lab members (Palvi Aggarwal, Zahid Maqbool, and Akash K Rao) for their invaluable support in the lab.

I am extremely grateful to my parents Dr. Suraja Choudhury and Mr. Ajit Choudhury, for their unconditional support and my brother (Arpan) for being the best brother in the world. Without their love, support, and constant encouragement throughout my life, I would not have completed this thesis or anything else for that matter.

Lastly, I would like to express my gratitude towards Mr. Larry A. Pickett, Jr., and Mr. Sayee Natarajan from RxDataScience Inc. for providing me with the funding for my research work.

ABHINAV CHOUDHURY

TABLE OF CONTENTS