EARLY IDENTIFICATION OF EXPERTS IN ONLINE QUESTION ANSWERING COMMUNITIES USING NEURAL NETWORKS

J.H.M.M.D Herath

229330E

Degree of Master of Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

March 2024

DECLARATION

“I declare that this is my own work and this thesis does not incorporate without acknowledgement 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 acknowledgement is made in the text.

Also, I hereby grant to 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: Date: 03/06/2024

The supervisor/s should certify the thesis with the following declaration.

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

Name of the supervisor: Dr. Charith Chitraranjan

Signature of the supervisor: Date 03/06/2024

ABSTRACT

Community Question Answering (CQA) platforms designed to enable users to ask questions and seek answers from the community have gained an increased popularity over the years. Prior research has shown that these communities thrive largely due to a small subset of expert users who provide comprehensive and often accurate answers. Identification of such experts in community question answering (CQA) platforms is an important, yet challenging task.

Many prior studies employ variants of classical machine learning and link analysis techniques to tackle this challenging task. However, studies leveraging novel advancements in neural networks to tackle this challenging problem are quite sparse. This thesis work models community question answering forums as heterogeneous graphs and leverages novel Graph Neural Networks to learn rich representations of users. Next, this thesis work adapts a neural network model to leverage learned user representations and embeddings of questions extracted from a pre-trained large language model (LLM) to identify expert users who are best suited to answer a given question.Through experiments conducted on a real-world CQA dataset, the author demonstrates the effectiveness of the proposed approach.

Furthermore, existing studies have not sufficiently explored the effects of considering different degrees of relevance among neighbors in CQA interaction graph, and largely assume all neighbors to be of equal relevance. This thesis work explores this under-explored area, and experiments with Graph Attention Networks (GAT) which makes use of attention mechanisms to weigh the messages passed from neighboring nodes. Furthermore, this thesis work proposes a novel edge-weighting scheme based on temporal decay to explicitly assign relevance scores to interactions among nodes. Through experiments, the author demonstrates how capturing different degrees of relevance among neighbors helps improve the identification of community experts.

Keywords : Expert Identification, Expert Recommendation, Neural Networks, Graph Neural Networks, Community Question Answering

ACKNOWLEDGMENTS

I would like to extend my utmost gratitude to my supervisor, Dr. Charith Chithraranjan for all the invaluable support given to make this research project a success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
CQA (Platforms)	Community Question Answering Platforms
GNN	Graph Neural Network
LLM	Large Language Model
GAT	Graph Attention Networks
GCN	Graph Convolutional Networks
NLP	Natural Language Processing
LDA	Latent Dirichlet Allocation
MRR	Mean Reciprocal Rank

LIST OF APPENDICES