Trust Judgment of Web Information A Content and Provenance based Framework for Trustworthy Web Information

Ph.D Thesis

By

Saba Mahmood
91-FBAS/PHDCS/S13

Supervisor

Dr.Anwar Ghani
Lecturer, DCS&SE, FBAS, IIU

Co-Supervisor

Dr. Humaira Ashraf
Assistant Professor, DCS&SE, FBAS, IIU

Department of Computer Science & Software Engineering
Faculty of Basic & Applied Sciences
International Islamic University, Islamabad

2020

A dissertation submitted to the
Department of Computer Science & Software Engineering,
International Islamic University, Islamabad
as a partial fulfillment of the requirements
for the award of the degree of
Doctor of Philosophy in Computer Science.

Department of Computer Science & Software Engineering
International Islamic University Islamabad

Date: *October 20, 2020*

Final Approval

It is certified that we have examined the thesis report submitted by Ms. Saba Mahmood, Registration No. 91-FBAS/PhD(CS)/S13, and it is our judgment that this thesis is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for the Doctor of Philosophy in Computer Science.

Committee:

External Examiners

Dr. Muazzam A. Khan ____________________
Associate Professor
Qauid-e-Azam University Islamabad

Dr. Arif Ur Rahman ____________________
Associate Professor
Bahria University Islamabad Islamabad

Internal Examiner

Dr. Tehmina Amjad ____________________
Assistant Professor
DCS&SE, International Islamic University Islamabad

Supervisor

Dr.Anwar Ghani ____________________
Lecturer
DCS&SE, International Islamic University Islamabad

Co-Supervisor

Dr. Humaira Ashraf ____________________
Assistant Professor
DCS&SE, International Islamic University Islamabad

Declaration

I hereby declare that this thesis, neither as a whole nor as a part thereof has been copied out from any source. It is further declared that no portion of the work presented in this report has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning.

Saba Mahmood

Dedication

Dedicated to My family, especially to my parents, my husband and my kids *Sarah, Ibrahim*.

Saba Mahmood

Acknowledgments

I am very grateful to *ALLAH* the *ALMIGHTY* for without His grace and blessing this study would not have been possible.

Foremost, I would like to express my sincere gratitude to my supervisor *Dr. Anwar Ghani* for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Ph.D study.

I would like to extend immeasurable appreciation and deepest gratitude for the help and support of my Co-Supervisor *Dr. Humaira Ashraf*, who gave all her knowledge, guidance and support to boost my confidence and learning. Her mentoring and encouragement have been specially valuable, and her early insights launched the greater part of this dissertation.

I would also like to acknowledge and extend my gratitude to especially *Dr. Ali Daud* whose valuable feedback, reviews, and technical help refined the research.

I would also like to thank my husband who has supported me patiently and firmly during completion of this task.

I would like to admit that I owe all my achievements to my truly, sincere and most loving parents and friends who mean the most to me, and whose prayers have always been a source of determination for me.

Abstract

The web is a decentralized repository of information where anyone can contribute regardless of their knowledge and expertise. Web content exists in various forms for example in micro-blogs, QA forums, intra organizational blogs email, and content available on websites. Web information is judged for its trustworthiness through its origin and content. Trust is achieved through credibility. Information can be regarded as credible if it flows from a credible source, techniques like digital signatures, and public key infrastructure are employed to find source credibility. We are however, considering the scenarios when false information might flow from credible sources. We are also considering the scenario when users are interested to find the credibility of information shared by unknown or unpopular sources. The web information seekers use their intuition to judge the trustworthiness and credibility of information present on web.

There are two broad categories of finding credible information including Human Based and Computational Approaches. The authors have proposed a Reputation Based Approach with a strong basis on Bayesian calculation structures. The previous techniques are based on a simple summation calculation structure making them prone to many reputation attacks. Furthermore, the techniques only take into account active and positive interactions disregarding passive and negative interactions. The authors have put forth the hypothesis that information reviewed or edited by an expert is regarded as credible. This hypothesis is further verified by a survey study conducted as a part of this research. Therefore this thesis has also addressed expert ranking and identified limitation of previous techniques in terms of the inability to take into account the quality of links. The previous techniques that are inspired by page rank like algorithms are also unable to address the issue of negative referrals. Expert rank techniques are mostly domain specific and lack generalization. The authors have however proposed a generic solution with specific application to the tacit expertise of employees in large online enterprises. Since the author has proposed a reputation based approach for the problem so it is prone to different reputation based attacks such as Sybil, Slander, and White Wash attack. These attacks are targeted towards increasing the rank of a malicious node or maligning the rank position of any legitimate node.

The proposed approach takes into account different kinds of interaction. The interactions information is then utilized by the Bayesian based mathematical model based upon updating of the beta probability density function. The proposed approach also includes defense strategies by proposing the inclusion of the history of interactions, limiting the number of negative interactions, and weighing the reputation value by reputation rank of the rater against the reputation based attacks.

The proposed framework is evaluated by performing different experiments. The performance metrics of Mean Average Error, Precision, and Correlation are utilized. The comparison is twofold. The comparison is performed against the baselines of content credibility techniques revealing that the proposed technique is independent of the density and pattern of rating. The results show 27% and 18% improvement in terms of precision for the two experiments conducted on two different datasets. The correlation results are also significant in both experiments with significant values of 0.39 and 0.87, thereby showing a linear relationship between the predicted and original data. Experiments conducted to carry out the comparison against the baseline expert ranking technique also, revealing an improvement in precision by 7% on average for the experiments conducted on three different datasets. Finally, the defense mechanism is also evaluated in three different scenarios revealing the fact that the technique is capable enough in defending against the reputation attacks.

Contents

List of Figures

List of Tables

List of Publications

- **Published Articles in IEEE ACCESS JOURNAL *(IF = 4.12)***
  - S. Mahmood, A. Ghani, A. Daud, and S. Shamshirband, "Reputation-based approach toward web content credibility analysis," IEEE Access, vol. 7, pp.139 957–139 969, 2019.
- **Papers under review in journals**
  - S. Mahmood, A. Ghani, A. Daud, " EER: Enterprise Expert Ranking using employee reputation " ETRI Journal, Wiley, 2020.
  - S. Mahmood, A. Ghani, A. Daud and H. Ashraf," A novel defense mechanism against attacks on Reputation based Web Credibility Systems" International Journal of Web Services Research, ACM, 2020.
- **Other Publications**
  - T. A. Haq, K. Mansoor, and S. Mahmood, "Congestion avoidance adaptive routing protocol for manets using network coding," in 2019 International Conference on Communication Technologies (ComTech). IEEE, 2019, pp.47–52.
  - U. Ayaz, T. A. Haq, S. Taimour, K. Mansoor, and S. Mahmood, "An enhanced biometric based rfid authentication scheme defending against illegitimate access," in 2018 14th International Conference on Emerging Technologies (ICET). IEEE, 2018, pp. 1-6.
  - F. Akif, A. ul Asar, S. Mahmood, M. F. Wyne, and S. Akhtar, "Swarm intelligence based reputation model for establishing trust in wireless sensor networks," Journal of Next Generation Information Technology, vol. 4, no. 2, p. 113–124, 2013.
  - S. Mahmood, A. ul Asar, H. Suguri, and H. F. Ahmad, "Swarm intelligence based reputation model for open multi agent systems," in Multi-Agent Applications with Evolutionary Computation and Biologically Inspired Technologies: Intelligent Techniques for Ubiquity and Optimization. IGI Global, 2011, pp. 248–266.
  - S. Mahmood, S. Iftikhar, H. F. Ahmad, and F. Mahmood, "Formal modeling of trust in semantic grid registry," in 2009 International Conference on Information and Communication Technologies. IEEE, 2009, pp. 65–69.

Chapter 1

Introduction

The new era of web has posed some new challenges for the researchers. Finding trustworthy web information is one such challenge. We define Trust as the ability of an entity to behave as expected. Evaluating the content of information is a complex process. Kato [1] in his work has suggested that the trust placed on information content is Credibility. Trust and Credibility are most of the times used synonymously. However, Trust is achieved through Credibility. Trust judgment is done through credibility assessment of the information, the authenticity of the source of information, expert opinion about the source of information. People are using the information on the web from blogs, websites, e-magazine, e-books, e-journals, social networks on a variety of subjects. There is no limitation regarding the contribution of information over these channels as far as expertise and authority are concerned. Thus the credibility of information is dubious. Examples of people utilizing information from the web for their daily decision making includes scenarios such as information related to medicine, illness and other health related issues, news related to stocks, business and investments, daily political information, students browsing the web for their knowledge and subject information are some uses. In such cases and several others, credibility is a crucial virtue. One way of checking the credibility of information is by verifying the source of information. Techniques like Digital Signature, Public Key Infrastructure are utilized for the purpose. We are however taking into account the situations where false information might flow from credible sources. Researchers [2, 3] have proposed origin-based approaches that address the issue of evaluating content based upon its origin or source. The recent works towards finding credible information on the web are all knowledge base driven, and the analysis is dependent on efficient techniques of retrieving facts. This is quite hard since information is growing in size at a rapid speed and keeping track of the facts in that information and building a knowledge base is not an easy approach. Among many factors [4] that contribute towards the evaluation of content for trustworthiness, one that is the focus of this report is the identification of the expert user. Other factors are context, provenance, popularity, authority, bias, direct experience, incentive, recommendation, related resources. This thesis has shown a relationship of trust judgment to credibility. The authors have discussed various approaches towards content credibility analysis. The approach proposed is a hybrid one involving collaborative filtering and user ratings approach. This is achieved through reputation information. In this thesis, the author has proposed an evaluation of content through expert analysis. A lot of information is present that is sometimes false, contradictory, and misleading. This is complex in situations when information is shared by an unknown user. Let us take an example of reviews written by people on product websites. The trustworthiness of that information is questionable, an expert advice can regard it as trustworthy. The piece of information is regarded as trustworthy if it is reviewed by an expert. The term expert defines an agent or person with a high degree of knowledge in certain subject [5]. Previous researchers have given different models for expert identification. These models either use document analysis technique or link analysis based technique except for a few that utilizes both approaches thus a hybrid technique. Document-based approaches use mining techniques to associate a topic with the author. This technique is however useful in scenarios where users produce a large number of documents. With the advancement of the social network, researchers proposed to identify the central influential node as the expert node. Most of the techniques are extended versions of the standard PageRank algorithm. However, those techniques are unable to address the collusion issue and do not take into account the quality of content. This thesis has proposed the technique for expert identification that has a strong statistical basis i.e. Beta Probability Distribution. It utilizes the reputation information of the user. Beta reputation models are already proposed [6] that try to identify the reputation of the agents in the eCommerce scenario. We utilized the expected value of the distribution to identify the node with the highest rank given different types of interactions. A user/node can have both positive and negative interactions with other nodes. Identifying the types of interaction between nodes in the form of messages or posts is also discussed. The proposed technique is reputation based, since it utilizes past interactions of the user for expert ranking. We thus compared the reputation based content credibility framework against other recent techniques of content credibility. The thesis also discussed the various attacks on the reputation systems and the countermeasures taken up by the proposed algorithm. As a methodology, a survey study was carried out to get feedback from users about different criteria for web content credibility thereby developing a dataset that can be treated as ground truth. The data is utilized to calculate Precision, Recall, and Correlation. The proposed technique is tested through experiments that are performed on five different datasets.

The experiments are conducted not only to verify the effectiveness of the expert rank algorithm but also its overall contribution in assessing the credibility of the content. The results show that the proposed technique outperforms the standard technique and is an effective tool for judgment of the credibility of the content. The results in terms of Mean Average Error, Correlation, Recall, and FMeasure support this argument. It is also worth mentioning that the proposed technique showed independence from the pattern of rating and density of data compared to previous techniques. The related subproblem of the thesis i.e expert identification is specifically explored in the context of large global organizations for employee ranking. The experimental results in comparison to other ranking techniques of the same domain show better performance. The defense mechanism adopted by the proposed technique shows promising results in a simulation scenario.

1.1 Motivation

There is plenty of information present on the web and social media. However, all of the information is not credible. Sometimes there is contradicting information. There is an increasing trend in users to turn towards the world wide web and social media for information seeking purposes. Information can be judged for credibility through its source, but sometimes even a credible source provides misleading information. Thus it is also important to evaluate the content of the information. The content of information can be judged by an expert for its trustworthiness and credibility. Expert identification is another problem of the World Wide Web. This all led us to formulate the research problems.

1.2 Problem Statement

Finding trustworthy information on the web is a challenge. Researchers [7–9] have proposed theoretical models for finding trustworthy information. However, there is no practical solution that could guide the users about the trustworthiness of the piece of information. Given below are the problems addressed in this thesis.

- The existing technique [7] for content credibility based on reputation information does not take into account passive and negative behaviors. Furthermore, the calculation structure adopted by the technique is a simple summation that is prone to different reputation attacks such as Sybil, Slandering, and Whitewash.

- Another problem addressed in this thesis is the identification of experts. The earlier techniques however lack ability to differentiate between the negative and positive interaction for ranking purposes [10–13]. Also, the techniques are based on the link analysis techniques, that cannot capture the quality of the links. With the emergence of large online enterprises there is a need to find employees with undocumented and tacit expertise [14–17].

- The thesis has also addressed the attacks on reputation systems in relation to expert ranking and web credibility systems [18, 19].

1.3 Research Objectives

Every day, lots of content is shared on the web. The web content serves as a primary source of information. This content is shared, edited, and reviewed by users from diverse knowledge backgrounds. Sometimes even incorrect information flows from credible sources. Thus the credibility of the content present on the web is questionable. However, experts can verify their credibility. In this background, we have thus established the research objectives that are listed below.

- To develop a framework for trustworthy web information that can guide users about the trustworthiness/credibility of the web content.
- A content credibility technique based on reputation information that has the ability to take into account the positive, negative, active, passive behavior of the creator, editor or reviewer of the content.
- A expert rank technique based on quality of interactions.

  A domain specific expert rank technique for large enterprise where no documentary proof of employee exists.
- A defense mechanism to counter attacks on reputation based content credibility systems such as Sybil, Slander, and the Whitewash attack.

To achieve the above objectives, the authors have developed a theoretical relationship of credibility to trustworthiness. We have proposed a content credibility technique. The performance of the proposed framework is evaluated on standard metrics. The implementation of the proposed technique is done in MATLAB. The results are compared against previous baselines.

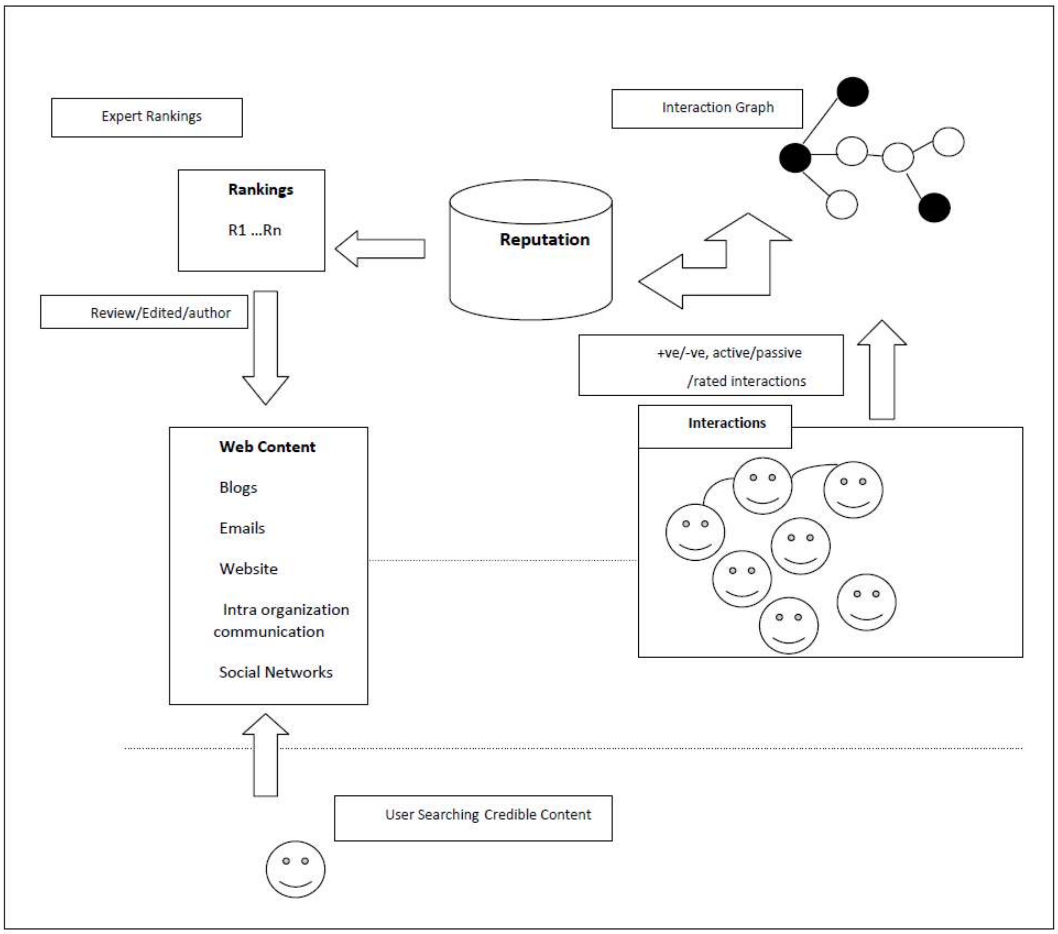

Figure 1.1: The Problem Domain

1.4 Application Scenario

There are various application scenarios of web content trustworthiness and credibility. Most of them are domain specific. Some of them are not directly addressing the problem; however, the approach is targeted towards finding credible information. We have given details of both types of applications targeted towards the identification of experts and trustworthy information.

1.4.1 Moleskiing

Moleskiing is an interesting application of a trust aware recommendation system. Skiing is always an exciting and thrilling experience. But unknown weather conditions and avalanches can pose a great risk to skiers. But if the snow conditions and routes situation is found in advance this can be greatly helpful in reducing the risk of emergencies. Thus moleskiing approach utilizes the information from trusted users about the conditions. In this system, the skiers are asked to rate the routes and other sky mountaineers about their perception of the information.

1.4.2 Twitterholic.com

This service basically is a third party service that scans twitterers few times in a day to find the top tweets, number of followers, and followees. The objective is to find the most influential, user.

1.4.3 Answer Garden

Answer Garden [20] is a feedback tool that can be utilized by organizations, classrooms and conferences for brainstorming sessions. A question is posted and shared with other users. Any answer generated is then shared with all the users. The users of this tool include teachers, for finding the knowledge level of the students. A teacher posts a question and in response gets answers from the class. It is used for digital brainstorming in conferences and workshops, to actively involve all the participants. The objective of the tool is to come up with an answer that is believed to be correct by all the participants.

1.4.4 Contact Finder Agent

Contact Finder [21] is an intelligent agent that assists users by referring them to users who can help them. It categorizes the messages and finds the topic of interest. Contact Finder reads questions from bulletin boards proactively rather than acting in response to the questions.

1.4.5 Twitter Who To Follow

Twitter has provided a who to follow service [22]. It utilizes graph based algorithms by generating connections between users based upon common interests. The service utilizes the self disclosed information of twitterers to compute the user recommendation. The service can also be used for searching other services, products.

1.4.6 TweetCred

TweetCred [23] finds credible tweets. It tries to find credible tweets and separates rumors and false news based on the credibility score. Such systems are again restricted by the context and trustworthiness of the scores. It utilized semi supervised learning by utilizing already labeled tweets.

1.4.7 Health Information

More recently health and medical information available on the web is gaining quite an attention for its credibility and truthfulness. Recent researches [24, 25] are trying to propose protocols to judge the credibility of health information.

1.4.8 Other Examples

Another interesting application of expert finding technique could be in case of reviews written against the products in e-Commerce sites. The technique can be employed in showing the expert level of those reviews regarding that particular product. The new arising mobile applications of service providers such as food and taxi systems that are community based, can utilize the framework in providing information from the experts about the trustworthiness of the services provided.

1.5 Thesis Organization

The thesis is organized in seven chapters with three sections. First Section is about literature review. While Section 2 is the Proposed Technique containing chapters 4,5,6. The final section is related to the conclusion and future directions. Route map of the thesis is shown in the figure 1.2.

Chapter 3 discusses the literature related to the problems of the thesis, that includes literature related to content credibility and the problem of expert ranking.

Chapter 4 then describes the proposed framework along with its mathematical model and the algorithm. The chapter also discusses the experiments performed for the evaluation of the proposed framework and the related results.

Chapter 5 describes the proposed reputation based expert rank algorithm and its comparison to other ranking algorithms. The experiments and the related results are also discussed.

Chapter 6 is small but identifies the major attacks on reputation systems in the literature, the proposed defense mechanism to counter the reputation attacks and the related simulations.

Chapter 7 finally concludes the thesis and also highlights the future research directions.

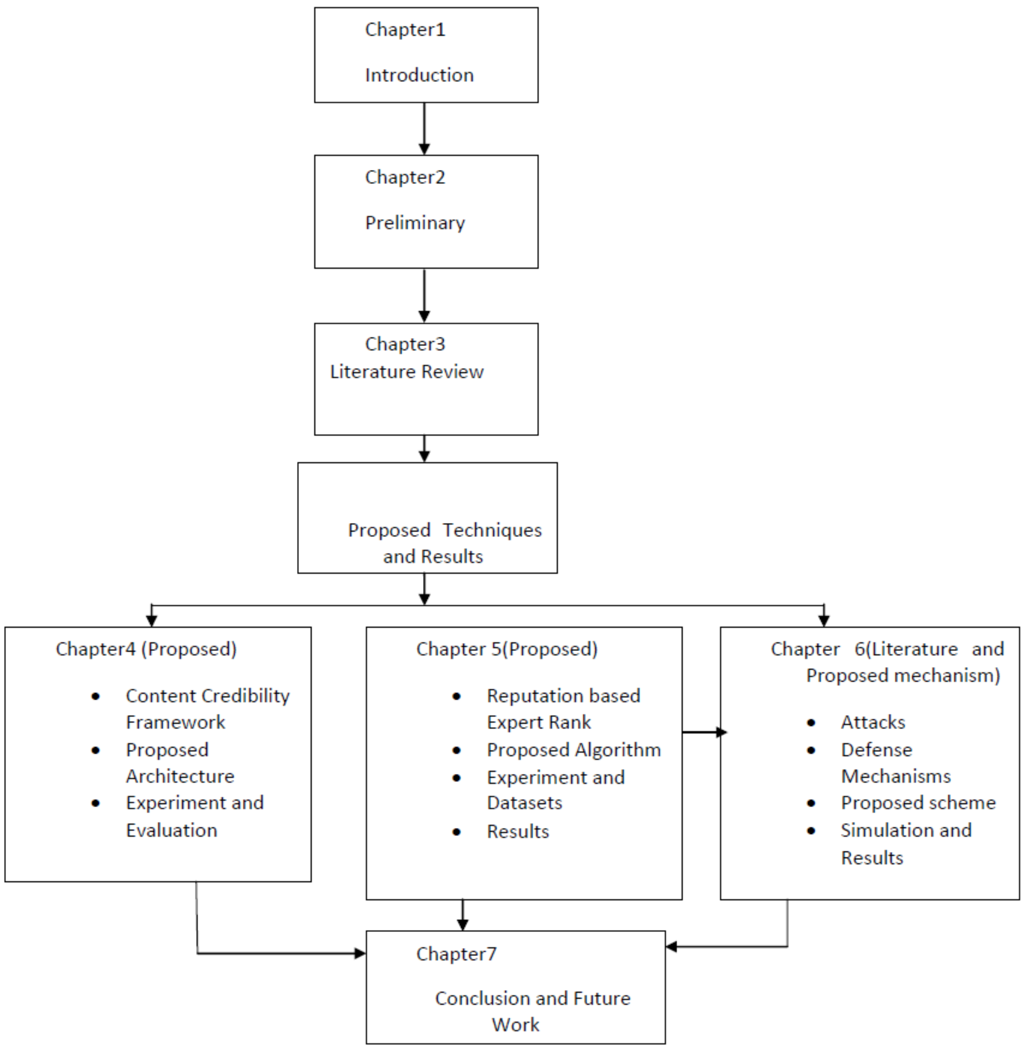

Figure 1.2: Thesis Road-map

Chapter 2

Preliminaries

In this chapter, the author has given details of background concepts that would be helpful to the readers in understanding the following chapters.

2.1 Trust And Security

The web has become an open decentralized repository, where lots of unknown entities and piece of information is available. In such an open environment existing security techniques are not appropriate. In such an open and uncertain territory people, organizations, and software agents cannot trust strangers. All this has evolved under the umbrella of the web of trust. However, the notion of trust has not received any formal definition so far. Currently, two definitions of trust exist. One that looks into trust from hard security approach while the other one defines trust based upon the quality of service. Researchers have come across a detailed definition of these approaches in their models and applications. Given below is the detail of various trust categories.

2.2 Trust Categories

Research in the domain of trust falls into two broad categories of policy based and recommendation based, however, there are some domain, specific general models, as well.

2.2.1 Policy Based Trust

These models of trust focus on establishing the credentials and enforcing the policies. Entities that have sufficient credentials are allowed access. We must say that this category of trust looks into trust from hard security perspective KeyNote PolicyMaker [26] are such models.

2.2.2 Recommendation Based Trust

This category of trust judges the behavior of an entity based upon its past behavior. The past behavior record can act as a guiding tool for future recommendations. If for example, no such information exists then referrals are used that combine information about the behavior of interaction from others [27–29].

2.2.3 General Models Of Trust

Lot of generic models have been developed that try to associate trust decisions related to agents, systems, etc. These models have tried to develop the important criteria for trust evaluation including, trust metrics, methodologies of finding trusted entities. Most of the models are inspired by sociology and psychology.

2.2.4 Trust In Information Sources

Trust has become an important theme for the information present on the web. How much the information on a particular website is reliable? What are the factors(design plus content) that affect the trustworthiness of the information present on the web?

2.3 Trust Elements

The trust model establishes trust in the entity, system, and content as described in the studies. Given below is the brief discussion.

2.3.1 Entity

Entity trust refers to the models that try to assess the trustworthiness of the entities involved in the interaction, for example, customers and business, users trying to find trustworthy websites, software agents locating trustworthy agents in a distributed environment, etc. Most of the previous research work in this domain is focused on problems of authentication and reputation of the entity.

2.3.2 System

System trust is the trust placed on the stable or predictable functions or behaviors of a system. System trust may appear as professional membership based trust, characteristics-based trust, institution-based trust, or regulation-based trust. System trust is especially useful in cases where a user has to interact with strangers and there are no trust paths to them in social networks.

2.3.3 Content

Content Trust focuses on the nature and use of the information being exchanged. Trust judgment on a particular piece of information in a given context is referred to as Content Trust. Many factors can determine the trustworthiness of the content. Content trust is subjective and there are various challenges related to the content trust. Gil and Artz [4] has tried to address these challenges and has identified the key factors that influence content trust.

2.4 Reputation Systems

Reputation is another termed used in the domain of trust management; though trust and reputation are used sometimes interchangeably they are two different concepts. Reputation is defined as " collected and processed information about one entity's former behavior as experienced by others". A wide variety of trust and reputation models have been developed in the past few years. There are two broad categories of these models i.e., centralized and decentralized.

2.4.1 Centralized Reputation Mechanism

These models propose management of the reputation of all the users in a center, they are specifically applied to online marketplaces such as eBay and SPORAS.

2.4.1.1 Ebay Model

It is implemented as a centralized system where users can rate the interactions of other agents in the past and also leave some textual comments about their behavior. For example on eBay, for interaction, a user can rate its partner as positive, neutral, and negative on the scale –1, 0, and +1 respectively. The reputation value is calculated by aggregating ratings from six months stored centrally [30].

2.4.1.2 SPORAS

SPORAS[28] introduced a new reputation system that updates the reputation of the involved party each time rating is gathered. SPORAS provides a more sophisticated model as compared to the previous model however it is designed for a centralized system that is unable to address the issues related to open MAS.

2.4.2 Decentralized Systems

In decentralized systems, each agent can carry out trust evaluation themselves without a central authority. The following section gives details of some of the decentralized models.

2.4.2.1 Jurca And Faltings

The popular reputation system [29] is built on the basic concept of awarding the agents that provide truthful interaction information. The objective of the system is to gather truthful information from the agents. The system has proposed a set of broker agents whose main purpose is to purchase the reputation information from other agents and producing an aggregate report. The reports are generated by using a simple average method as the calculation structure. The broker agents can then sell back information to other agents when they need it. The architecture of this system is distributed but the reports are generated by the agents in the centralized capacity, thus making it unadjustable to a highly dynamic environment; as computation is still carried out centrally.

2.4.2.2 Regret

Regret [31] model adopts a decentralized trust evaluation process. Due to decentralization, each agent can calculate the reputation value of other agents with whom it has interacted. This is achieved by rating every interaction that an agent had with other partner agents. The agent then stores the result in its locally maintained database. Regret uses direct trust and “witness” reputation to compute the final reputation value of the agent. REGRET agents share the information without any incentive system. The social network is utilized to choose and weight the witness to be used, however, one shortcoming in REGRET is the inability to specify the mechanism of creation of the social network. In summary, we can say that REGRET equips the agent to evaluate the trust value itself, thus working in a decentralized manner. Therefore this model is compatible with the requirements of an open multiagent system(MAS), however, its limitation to address how each agent can build the social network makes it less applicable.

2.4.3 TRAVOS

TRAVOS [32] is a model of trust and reputation that can support decision making for good interactions in a virtual organization that is based upon the GRID environment. If a group of agents is to form a VO(Virtual Organization), then they need to choose the most appropriate partner. This model is built upon the probability theory. TRAVOS utilizes direct experience based information and recommendation based information in calculating the overall trustworthiness of an agent. The computed value is considered as the probability with which an agent interacts either successfully or unsuccessfully termed as an untrustworthy behavior. Thus the model considers the outcome of an interaction of an agent as a binary value.

2.4.4 The FIRE Model

This model [33] has taken up the scenario of producers and consumers. The consumers select the producer based upon their reputation. Reputation is measured in terms of utility gain (UG). The model uses four different sources of input that includes interaction reputation (IR), witness reputation (WR), certificate reputation (CR), and role based reputation (RR). IR utilizes direct interaction of agents, an agent 's' might give ratings for agent'b' for the products it received from agent 'b'. The witness's reputation is the information gathered from other agents about a particular producer. Role based trust is the reputation ranking of an agent according to certain rules set by the organization etc to which these agents belong. While certificate based trust is the reputation information that the rated agents provide about itself. The certificate of ratings that it has received from different sources. The model then incorporates these sources of information in a composite value by weighted mean method, whereby the weight is decided by the user to reflect the importance of these sources. Based on the reputation value the providers are categorized as good, bad, intermittent, or ordinary providers. The model compared against SPORAS, REGRET models produced better results showing that agents can select a reliable provider by utilizing the model. The model is decentralized in nature thus making it suitable for open MAS.

2.5 Credibility

The definition of credibility is discipline specific. From the qualitative research domain, credibility is the first thing that needs to be established. It expects researchers to link the research findings to already available facts. From a psychological perspective, credibility is the believability and reliability of the information received from outside the resource. The Merriam Webster [34] dictionary defines credibility as "the quality or power of inspiring belief". Fogg et al. [35] has defined that credibility is based on trustworthiness and expertise.

2.6 Expert

According to Merriam webster [34] an expert is defined as "having, involving, or displaying special skill or knowledge derived from training or experience, one with the special skill or knowledge representing mastery of a particular subject".

2.7 Bayesian Algorithms

Bayesian [36] based reputation algorithms are easy to understand and can be easily applied to a wide variety of application domains. Bayesian based systems are binomial or multinomial. Binomial bayesian systems are applicable to scenarios with binary outcomes for example good, bad. The binary outcomes represent the service quality of an entity. The Beta distributions belong to continuous distribution functions in a binary mode that uses two parameters, commonly named as alpha and beta.

Multinomial Bayesian reputation systems allow ratings to be provided over k different levels which can be considered as a set Bayesian based reputation systems have shown promising results and are easy to implement. Bayesian reputation systems have a sound basis in classical statistics that makes them effective and adaptable to various contexts.

2.8 Techniques For Expert Identification

Expert identification is considered a newer area, that emerged in recent times. The techniques utilized in expert identification are discussed below.

2.8.1 PageRank

The techniques discussed above are an extended version of the popular PageRank [37] algorithm developed by Google to list the top web pages of a query. The basic theory behind PageRank is that it measures the number of links to a page. The page with maximum links is then brought as the top result. PageRank is a link analysis algorithm that assigns weightage to the elements being hyperlinked thereby finding the relative importance within a set. Its history dates back to the concept of Eigenvector. PageRank of a page 'a' can be mathematically represented as,

$$PR(a) = \sum k \in S_u \frac{PR(b)}{L(b)} \tag{2.1}$$

PageRank based algorithms are proposed to find the central influential element of the network, referred to as an expert element. The Expert Rank, Twitter Rank algorithms are inspired by the PageRank algorithm.

2.8.2 HITS

There are two types of pages, one that represents an authoritative source such as CNN, BBC, etc called authority. The pages linking to authority pages are called hubs. A good hub connects to many good authority pages similarly a good authority is one that is connected by many good hubs. Thus each page can have two types of scores i.e. authority score and hub score. Expert Ranking systems as proposed by Zhang [38] utilized both PageRank and HITs to measure the rank of a node.

2.8.3 Document Analysis

Document based techniques evaluate the documents produced by the authors to relate a topic with the author thereby finding a topic related expert. Twitter based expert ranking technique Twitter Rank uses LDA to associate a topic to an expert user. The tweets, posts are used as documents to generate this association. Systems that utilize only document based techniques includes EEL, AnswerGarden System [20], ContentFinder [21]. Document analysis based systems have demonstrated acceptable performance results in combination with the link based techniques. These techniques are well suited for the scenarios where a large amount of documents and content are produced by the user.

2.9 Chapter Summary

This chapter helps familiarize the reader with the concepts utilized in this thesis. The chapter has given details of trust and reputation related concepts along with a discussion of some popular systems like REGRET, FIRE. The discussion shows that trust can be derived from the reputation information. Credibility is also defined from different perspectives. The reason for deriving definitions is to create a linkage between trust and credibility. The chapter has also highlighted popular trust models and the expert rank techniques along with the popular techniques utilized in these models. Thus the chapter also gives details of the basis of expert rank algorithms like PageRank and HITS. Furthermore, a section is devoted to the discussion of Bayesian systems and their mathematical model.

Chapter 3

Literature Review

This section gives details of the literature related to trust judgment, content credibility and expert ranking. Trust judgment of web information is a research issue that has been approached by different methodologies. For example in the earlier research in the field of question answer by Clark [39]. In this work, the author has addressed the issue in a scenario when a user gets multiple answers in response to a query. They proposed that most frequently occurring answers must be treated as true. Work in information filtering has addressed the concept of quality. Quality on the Web is discussed in literature [40] that has pointed out that lots of content on the web is becoming old and irrelevant. Lately question answer based microblogs and forums have risen due to an increase of trend in users to consult the web for answers to their queries. However, there is a problem with getting multiple answers for a single query. Clark [39] proposed that in such a problem most frequently occurring answer may be considered as a correct and valid answer.

A model [40] also demonstrated that frequently occurring answers can be utilized as a metric to label an answer as correct.

In much later work with the emergence of the semantic web, researchers proposed a browser Triql.p that could display information to users as a result of the policies defined by them in terms of the content, context, and source. Ding et al.[41]provides a method to find trustworthy information sources based on provenance and trust. The solution is based on semantic web technology that proposes an ontology to find an association between trust and provenance. The work proposes a mechanism to filter information gained from different sources. Agents [42] utilize the ability of the semantic web to find the context, together with the reputation information and context agents seek trustworthy information. The results enable the agent to ask other agents "which agent can I trust to get the weather?". In related work recommender systems on the Web, may filter information based on recommendations and/or trust ratings. Trust values, or recommendations, are computed within a group of "similar" users, and the resulting information is filtered accordingly. With the proliferation of social networks, the issue of trustworthiness is addressed as to how users can trust each other. In earlier work in this regard is by Goleback [43]. This work proposed a trust inference algorithm to find a trustworthy user. They applied their proposed technique to applications like FilmTrust [44]that recommends reviews from trustworthy users. STrust [45] proposed a framework that brought the concept of social trust and social capital. The framework builds and sustains an online social community. The model computes the engagement and popularity of an individual agent/user in the community. These earlier researchers are seeking reliable information from a source that can be trustworthy. A proposal in this regard is to get an opinion from a subject expert, but the identification of a subject expert is a complex problem. In our daily activity, people intend to find reliable information from an expert. In the World Wide Web finding a subject expert is a problem. Generally, an expert is identified through the credentials and opinions of others. Thus utilizing the same process researchers proposed techniques of finding experts in the www, however, judgment through credentials is only helpful for finding popular experts. Scenarios when unknown experts are to be searched, opinion, and content produced by the prospective expert need to be analyzed.

3.1 Trust

"Trust is the psychological state in which the trustor believes that the trustee behaves as expected in a specific context, based on evidence of the trustee's competence and goodwill [46]". Trust judgment is comprised of Trust assessment and Trust decision.

"Trust Assessment is the process whereby the trustor assesses the degree of belief that the trustee behaves according to the belief of the trustor". Trust Decision in literature has utilized decision trees and utility theory. Trust decision is done through the Trust degree that means the value that indicates whether to trust or not to trust.

Trustworthiness is defined as an estimate of the level of trust that the Trusting Agent has in the Trusted Agent [47]. Thus the title of the thesis " Trust Judgment of Web Information" employs a technique that is able to do trust assessment so that a trust decision could be taken. Trustworthy information implies information that is at a higher level of trust compared to others.

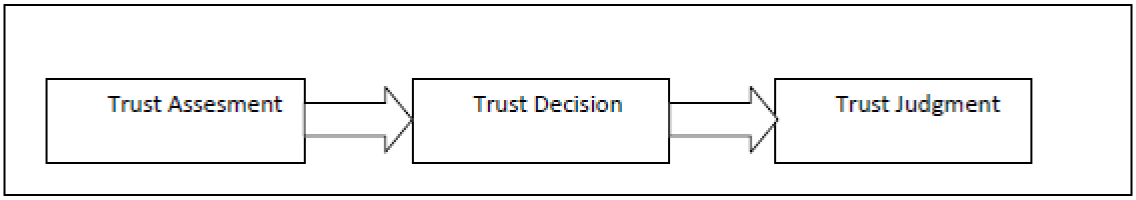

Figure 3.1: Trust Judgment Process

3.2 Provenance Trust

The details such as author, publisher, citations, etc give information about the source of information termed as provenance, which has been used in literature as a way to evaluate the trust. Provenance representation and tracking have been studied in the context of information sources. Provenance Trust is the trust placed on the source of information. Traditional security techniques of PKI(Public Key Infrastructure), Digital Certificates are used to identify authentic sources. Traditional security techniques are useful for well-known sources however with the emergence of web 2.0 whereby microblogs and social networks have become a major source of information, that shares most of the information from unknown sources. In such cases, the content of information can be evaluated for its trustworthiness.

3.3 Content Trust

Content Trust is trusting a particular piece of information in a particular context. Issues relevant to a content trust includes authority, topic, context, popularity, criticality, direct experience, related resources, provenance, bias, incentive, age, appearance, and user expertise [4]. Content trust is subjective and is context-dependent.

3.4 Web Information Trust Models

Given below is the detail of the trust models specifically designed for web information.

3.4.1 MoleTrust

MoleTrust [48] allows only relevant and reliable information according to the user's point of view of other author's trustworthiness. They have developed a MoleSkiing website that recommends various routes to the user and also allows users to show how much they trust other users.

3.4.2 Advogato

Advogato [49] has presented an analytical framework of utilizing the trust metric for public key certification. Appleased [50] has taken concepts from spreading activation models in psychology and has related to trust concepts within the semantic web. They have proposed a local group trust metric. Their trust model is somewhat similar to FOAF where all trust information is publicly accessible through personal home pages. Both Advogato and Appleased have utilized the local group trust metric.

3.4.3 TRELLIS

TRELLIS [51] system is an information analysis and decision making tool. TRELLIS annotates the decisions and tradeoffs and facilitates users to extend the vocabulary. It provides an interactive environment that allows users to add their opinions and conclusions as they analyze the information.

3.4.4 SNOPES

Snopes.com [52] separates a rumor from true information. They have a group of researchers who take up the popular news and do manual research to find the source and credibility of the information. They categorize the information as either true, false, rumor, or unknown.

3.4.5 EMERGENT Info

Emergent Info [53] is a real time rumor tracker; it is part of the project by the Tow center of journalism at Columbia University. The basic objective of the group is to find rumors and debunking of false information. Part of their project also highlights the importance of social media in spreading news.

Table 3.1: Web Trust Models Comparison

Web Trust Models	User can add information	Chain of inference	Social information	Personalized	Expert opinion
TRELLIS	Yes	Yes	No	No	No
SNOPES	No	No	No	No	No
EMERGENT INFO	No	No	Yes	No	No

To our knowledge, three models have tried to check the trustworthiness of web information. We compared them on five parameters, i.e. information; that means the users could add their reviews about the inference results. The second parameter is the chain of reference, only TRELLIS shows the steps of evaluating a piece of information, and how they evaluate the trustworthiness of the information. Social information means, models, utilizing the information from the social network, only emergent info explores the social network to assess the trustworthiness of the information. The last parameter is personalization. None of the models give personalization to the user, like for a particular user 'A', for example, what piece of information was evaluated and how much he was satisfied after getting the results from the models. All of these models do not take into account the expert level of the reviews.

3.5 Trust And Credibility

Trust and Credibility are most of the times used synonymously; however, the authors believe that trust is achieved through credibility. According to Merriam-Webster [34] Credibility is the "quality of being believed or accepted as true, real, or honest" while the term Trustworthy is defined as being able to be relied on to do or provide what is needed or right. Kato et. al [1] has related reliability to trust. Lincoln and Guba put forth an argument that insurance of credibility is one of the prime factors in the establishment of trustworthiness [54]. The trust judgment technique includes trust assessment and decision processes. The figure 3.2 depicts the relationship between credibility and trust judgment as derived through definitions, thus the basis of the proposed framework in this thesis. In this thesis, the proposed trust judgment framework is based on the credibility technique.

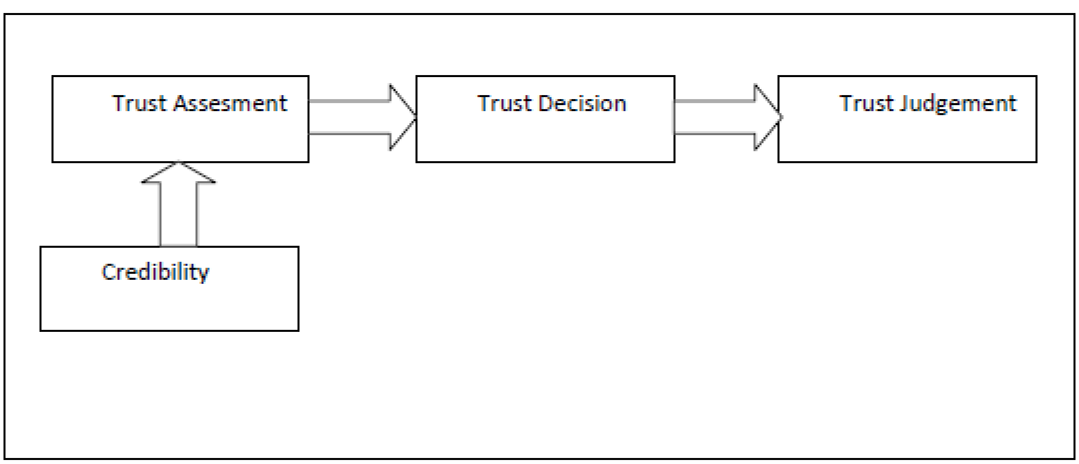

Figure 3.2: Trust Judgment Through Credibility

3.6 Content Credibility

The major problem discussed in this thesis is the assessment of the credibility of the content present on the web. Information on the web is present in various forms such as micro-blogs, email, websites, QA forums, etc. Thus *Content in this thesis is defined as the information present in the form of microblog(Facebook like) comments, email conversation, inter-organizational interactions, subject based interactions in student blogs*. Web content credibility can broadly be categorized based on the evaluation mechanisms they utilize, including only human based evaluations or computer based evaluations. Human based approaches are the ones that utilize the judgment through users, whereby they explicitly judge and rate the content. Computer based techniques include digital signature, machine learning techniques, collaborative filtering, semantic web, and credibility rating systems. Recently a comprehensive review work has highlighted these categories in detail [55], this problem is solved from different approaches, the one proposed in this thesis is Reputation Based. The following sections discuss the related literature. Web content exists in different forms for example content on websites, in social blogging sites, emails, QA forums, intra organization communication as shown in Figure 3.3. Researchers have proposed various methods of content credibility. Given below is the detailed description of different categories of research in this domain.

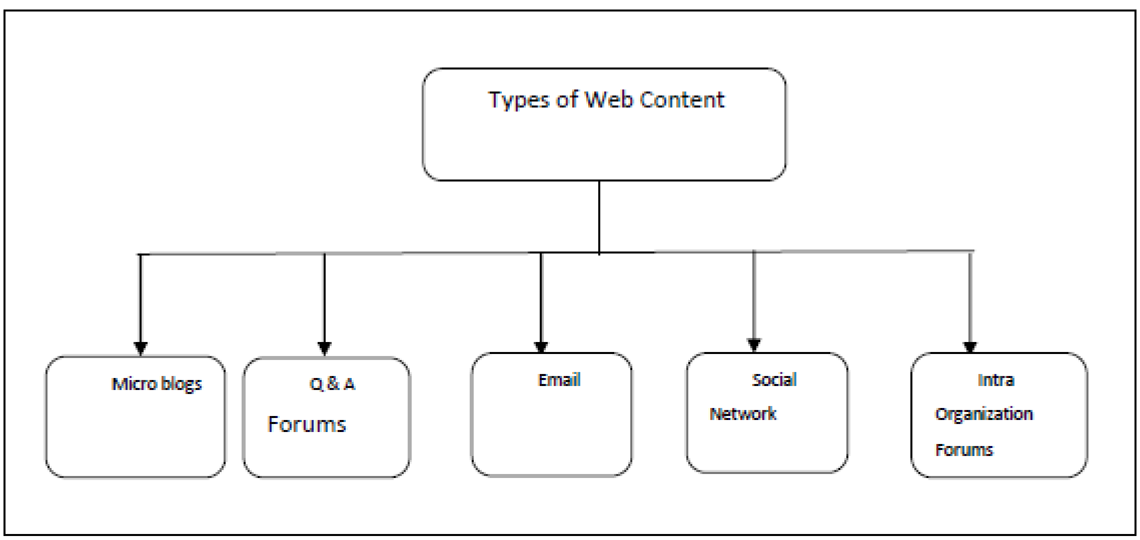

Figure 3.3: Types Of Web Content

3.6.1 Human Based Approaches

Human based approaches include techniques such as visual appearance, web layout, URL, date, personal belief, site familiarity, etc. Through various studies, [56–58] was found that users perceive information differently and thus assess the credibility of it differently and sometimes even the credibility of the source of information is neglected.

3.6.2 Computational Approaches

Computational approaches include various techniques such as:

3.6.2.1 Digital Signatures

Digital Signature [59] is an electronic method to provide proof that the document is authentic. The signature verifies that the document is indeed created by the mentioned author. However, situations, when inaccurate content flows from authentic and credible authors cannot be addressed by this technique.

3.6.2.2 Collaborative Filtering

In this technique, the content is evaluated by peers and experts. Peer review systems [60] for journals and publications are an example.

3.6.2.3 Machine Learning Approaches(ML)

Olteanue et al. [61] proposed a system that looks into social network features for predicting the credibility of the information. ML techniques require structured data and ground truth values. Although their results are 70% accurate but the solution is computationally expensive for everyday user of the web content [62, 63] as the user might be able to utilize such a service by purchasing it due to its cost. Authors of the latest work utilized [64] supervised learning techniques to find the credibility of information during high impact events. They used the random forest for classification; results showed improved accuracy with the approach. Techniques like [63–66] have utilized supervised techniques to find credible tweets. Fairbank et al [67] has evaluated text based techniques with respect to a structural approach in order to identify fake news compared to textual that is insufficient in finding credibility. More recently [68] authors proposed a trust based solution to find about the trustworthiness of the information resources, their technique employed a weighted average method whereby weights are dynamically assigned by moving weighted average and ordered weighted average.

3.6.2.4 Semantic Web

Semantic web [69] can judge the content using reasoners [70] is based on different criteria that cannot be addressed in other techniques. However, it cannot be effectively implemented since all the content on the web is not present in an organized manner.

3.6.2.5 Content Ratings

The credibility rating systems are based on the ratings given by users and experts on the content [71]. These systems, however, are unable to ascertain the expert level of the users giving the ratings.

3.7 Reputation Based Credibility Evaluation

Reputation is defined as the Overall quality as seen and judged by users according to the Merriam-Webster's [34] online dictionary. The past behavior of the entities and opinions of others is utilized to find the reputation. The opinion is based upon the past history of interactions. Existing reputation systems are categorized either into a centralized and distributed system. The centralized systems have a central entity that records the experience of all the entities in the network, while a distributed one does not has a central store instead all the entities keep a record of their interactions with other entities. Systems like [72, 73] are based upon distributed reputation mechanisms. The distributed systems have to address the issue of propagation of reputation information in addition to the method of calculating the reputation scores. The author's area of interest is the calculation of reputation scores since it is the basis of comparison with respect to other models. The most common ways of calculations are; Summation, Average, Weighted Average, and Bayesian.

Summation, the simpler one where positive and negative scores are summed up separately for each individual. The method is employed by eBay [30] and is the easiest method to understand and implement.

Reputation systems employed by Amazon and Epinions [74] use averaging whereby the scores are averaged to present the reputation score of all the entities. In the Weighted Average method, the average reputation score is multiplied by a rating factor. The rating factor depends on the age of score, trustworthiness. Bayesian systems [75] are based upon computing reputation scores by calculating and updating the beta probability density function. The major advantage of the approach is that it provides the theoretically sound basis for computation of the scores. There is no known disadvantage of this approach.

More recently researchers have utilized PageRank based reputation models [76] to rank volunteered geographic information (VGI) system, thus considered as a new calculation structure of the reputation systems. Calculations can be performed in a centralized manner or decentralized manner for example in the case of peer-to-peer systems. Both structures have their pros and cons.

Another reputation structure utilized by researchers is based on the Normal Distribution Reputation(NDR) [42]. Different reputation models have been widely researched in various domains such as online market places, p2p systems for example eigentrust [77], regret [31]. It is worth mentioning that Bayesian reputation systems, addressed the limitations of the most popular eigen reputation systems [78].

Reputation approach towards Content Credibility technique [7], utilized reputation for credibility assessment of twitter data. The reputation model utilized engagement, popularity, and sentimental information for calculating the reputation rank of the user. This work has utilized simple summation as the calculation method that is centralized in structure. The authors of this work compared the result with other machine learning techniques of naive Bayes and logistic regression. The results in the form of precision and recall showed improvement with reputation based technique also overcoming computational cost associated with training of data in the case of machine learning methods. The technique is however unable to take into account the negative relationships and are prone to reputation attacks. Authors have treated this recent technique as the baseline technique along with other reputation baselines including Page Rank and NDR based approaches for comparison against the results produced through the proposed technique.

3.8 Credibility Literature Overview

Given below is the summary of the literature review of the key recent papers with their limitations and concepts. The limitations highlight the research gap addressed in this thesis. These limitations are categorized in terms of input and process. Limitations in $input$ means the parameters utilized by these researches and $process$ implies the limitations in the working of the algorithm and the methodology.

Table 3.2: Content Credibility Literature Overview

SNO	Title	Limitations	
		Input	Process
1	M. Alrubaian, M. Al-Qurishi, A.Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino,"Credibility in online social networks: A survey," IEEE Access, vol.7, pp. 2828–2855, 2018.	Review Paper	Review Paper
2	R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban, "Cat: Credibility analysis of arabic content on twitter," in Proceedings of the Third Arabic Natural Language Processing Workshop, 2017, pp. 62–71.		Computational cost associated with supervised learning, depends on human judgment and ground truth
3	M. Kakol, R. Nielek, and A. Wierzbicki, "Understanding and predicting web content credibility using the content credibility corpus," Information Processing and Management, vol. 53, no. 5, pp. 1043–1061, 2017.		Results obtained from survey data after statistical analysis, no computational algorithm utilized.
4	M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, M. M. Hassan, and A. Alamri, "Reputationbased credibility analysis of twitter social network users," Concurrency and Computation: Practice and Experience, vol. 29, no. 7, p. e3873, 2017.		Simple summation as reputation model, unable to take into account number of down votes, no. of unfollowings, no. of dislikes etc.
5	A. Abdel-Hafez, Y. Xu, and A. Jøsang, "A normal-distribution based reputation model," in International Conference on Trust, Privacy and Security in Digital Business. Springer, 2014, pp. 144–155.	Applicable to normally distributed variables	

6	C. Lodigiani and M. Melchiori, "A pagerank-based reputation model for vgi data," Procedia Computer Science, vol. 98, pp. 566–571, 2016.		Quality of interactions
7	A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, "Web credibility: Features exploration and credibility prediction," in European conference on information retrieval. Springer, 2013, pp. 557–568.	Structured data required, ground truth values needed, computationally expensive	
8	M. J. Metzger and A. J. Flanagin, "Credibility and trust of information in online environments:The use of cognitive heuristics," Journal of Pragmatics, vol. 59, pp. 210–220, 2013.		Only Human Judgment based
9	A. V. Pantola, S. Pancho-Festin, and F. Salvador, "Rating the raters: a reputation system for wiki-like domains," in Proceedings of the 3rd international conference on Security of information and networks. ACM, 2010, pp. 71–80.		Expertise of raters
10	X. Liu, R. Nielek, A. Wierzbicki, and K. Aberer, "Defending imitating attacks in web credibility evaluation systems," in Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013, pp. 1115–1122.		Expertise of raters
11	M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research," Journal of the American Society for Information Science and Technology, vol. 58, no. 13, pp. 2078–2091, 2007		Only Human Judgment based
12	R. Savolainen, "Judging the quality and credibility of information in internet discussion forums," Journal of the American Society for Information Science and Technology, vol. 62, no. 7, pp. 1243–1256, 2011.		Only Human Judgment based

3.9 Expert User Identification/Rank Systems

Recently lots of research has taken place related to expert identification [79]. The following sections give details of research carried out in the domain of expert identification/ranking. The expert ranking has been studied in different scenarios like finding an expert in micro-blogs. Newer areas of expert findings includes author ranking [80–82] and employee/contractor ranking in a large organization or online job portals [16].

3.9.1 Graph Based Expert Rank Systems

COGNOS [10] uses crowd wisdom to find a topic based expert on twitter. In twitter users can create lists of popular experts. The proposed technique crawls through these twitter lists and finds the topic based experts. The authors of this technique carried out a comparison against the twitter who to follow service by performing numerous experiments. The results showed that this technique produces better results. However this technique is vulnerable to attacks through the creation of fake lists by malicious users. Twitter has provided a who to follow service [22] enables users to find an expert in twitter. They use the disclosed information of twitter users to compute the expert twitterer. The detailed technique has not been disclosed by the researchers as yet. This work [83] brought important findings that the history of reviews of an advisor in an online marketplace plays important role in grading advisors as reliable. The study also revealed a strong relationship of trust to expertise. It was also found that extremely positive or negative advisors cannot be regarded as trustworthy. The advisors who write both kinds of reviews with no pattern are more trustworthy. Twitter Rank [11] is an advancement of PageRank algorithm [84] for Twitter. The algorithm first performs topic based distillation, then it finds topic based influence derived through the following relationship. The algorithm has justified the concept of homophily. The algorithm uses LDA [85] model to perform topic distillation from the tweets. Expert twitter accounts are located by finding the influence of topic based tweets. The results from the technique are compared against the indegree, PageRank algorithm, showing improved performance of the results. The authors of this work [86] comment that Twitter Rank generally identifies popular experts. The authors have taken up the problem when information sources are not well known. Thus they proposed a certain set of features that can be helpful that may include the number of tweets, number of followers, etc. This data is then clustered by the Gaussian Mix algorithm. A comparative analysis was performed against the result of the survey. The top ten authors identified by the technique are compared against the survey results. Expert Rank [87] proposed utilization of document based and link based techniques in order to find a topical expert. Using the LDA technique, they associate a topic with an author from the documents produced by the author. The algorithm then utilizes a PageRank inspired technique to compute the expert rank of an author in the microblogging sites. Through their results, they proved that both document and link based techniques when used together produce better results in terms of precision and recall. The technique has also proposed a weighted version of link based algorithm that addresses the issue of collusion. This work [13] utilizes chains of a social network. When a user receives a query request, the request is forwarded to the neighbors based on the profile match. If the request can not be fulfilled by immediate neighbors the request is forwarded to the next levels until the search is accomplished. The system then pays everyone in the social referral chain. Link analysis techniques [88] are employed to explore email communication in organizations. Similarly, graph based technique [89] is utilized to find an expert from the online discussion board of an online course. The table 3.3 summarizes various expert finding techniques. The third column tells the scope of these frameworks. Some of the frameworks are targeted towards finding the Macro level or worldwide known experts. The third column for differentiation among the models is Scope. The scope is of three types Macro, Micro, and Both. Micro means an individual's social circle. While Macro means worldwide. COGNOS, Profile History, Twitter Rank ranks the expert on a topic universally. That means they can find world known experts. While Expert Rank, Social Referral, find / rank experts on a given topic in a user's local social graph. Every ranking framework has utilized a different technique. As discussed earlier the common techniques employed for ranking experts are link based and document based. The only Expert Rank algorithm uses both techniques.

Table 3.3: Expert Finding Techniques

Framework	**Parameters**	**Scope**	**Technique**
COGNOS	Twitter List	Macro	Mining Twitter List
Profile History	Review History	Macro	Mining Review History
Twitter Rank	Followers	Macro	PageRank and LDA
Topical Authorities	Follow,Reply	Both	Gaussian Mixture Model
Expert Rank	Document Analysis, Social Rank	Micro	PageRank
Social Refferal	Social Connection	Micro	Profile Matching

Micro levels are the ones that try to find the experts at local /community levels. COGNOS explores the twitter list of twitter users and employs mining techniques to find the topic related to experts. Twitter rank is inspired by PageRank and utilized the LDA technique to relate a topic to the document produced by the author. The last three models try to find influential users in the network. Their techniques are Eigen Centrality and Gaussian Mixture Model to identify the influential node.

According to their work, an influential node is most likely the expert node. The document analysis technique utilized by Expert Rank is well suited for scenarios when users have produced a large number of documents such as in the case of organizations or blogs.

3.9.2 Document Based Systems

This technique is utilized by researchers to find experts related to a topic in scenarios when documents are present in an organized manner and the quality of documents is high. Mostly LDA [85] based technique is used to find topic based experts. Twitter rank utilized this to find topic based experts from tweets, post produced by twitterers. Probablistic topic models [90] utilized text analysis techniques to find a topic expert thus a document based technique. Recently researchers [91] utilized LDA technique on social annotations for topic modeling.

3.9.3 Hybrid Systems

These expert finding systems use document based and link based techniques to find an expert. Campbell et al [92] proposed an algorithm designed to find experts from the content of email communication called a HITS algorithm. It is a graph based algorithm with the concepts of hubs and authority. The proposed technique analyzes email communication patterns in an organization to locate an expert. The results of this work were compared against a simple content based algorithm that counts the number of emails sent on a particular topic. The results in terms of precision and recall showed better performance, however with a shortcoming that the technique is only well suited for small sized datasets.

Fu et al.[93] proposed to combine both contents of a document and social network information to identify an expert. A social network was built from email communication with the co-occurrence of people on webpages. The work demonstrated performance improvement however, they utilized organizational datasets where the quality of documents produced is much higher. Thus it is unknown if it can be applied to online knowledge communities where the quality and quantity of content are very low. Zhang et al.[94] have proposed an expert finding algorithm for online knowledge communities such as Java Forums, where the information is not present in an organized or structured manner. The work used document and link based techniques. The performance of the algorithm however could not be reported due to a lack of empirical evaluations. JWang et al. [95] predicted the user with the best answer as a subject expert through convoluted neural networks.

3.9.4 Reputation Based Expert Ranking

Faisal et al.[96] proposed a technique whereby a user is ranked as an expert based upon reputation information of co-existing users. The reputation of the user is calculated by finding the number of times the user appeared in question answer thread. Another reputation structure utilized by researchers is based on Normal Distribution Reputation(NDR) [42] where the weights are generated from a normal distribution is used by workerrank [17] expert ranking system. Recently a reputation model based on Hidden Markov Model(HMM) [16] has been proposed to carry out the evaluations in online markets. The research addresses issue of reputation staticity and inflated reputation scores in a dynamic environment. They propose that reputation rank may be derived from the recency of interactions. A similar issue has been addressed in time bound topic modeling for expert ranking [97]. Answer reputation [98, 99] was measured by the number of accepted answers or upvoted answers. The reputation model is simpler and did not take into account user consistency and tags [100].

3.10 Application Scenarios of Expert Finding Systems

Previously areas mostly targeted by researchers regarding expert identification and ranking are microblogs, and email communication in organizations.

3.10.1 Email and Organizations

Email communication [101, 102] is utilized by certain researchers in order to identify an expert in an organization since it represents the formal mode of communication. A network graph can be created from the communication and various link based techniques are applied to find the expert. However, none of the techniques have taken into account the quality of the content of the emails. The same methodology is applied for organizations. There are certain cases where employees rate each other on a certain scale regarding some required expertise. That information is also utilized to find an expert employee. Worker Rank [17] that ranks employees based on expertise utilizes a weighted average method to compute reputation value. The weights are calculated from the normal distribution. This domain has utilized graph based or document based techniques. Expert ranking in organizations is based upon the availability of the type of information. Sometimes information is not properly documented, the expertise or knowledge of an employee is not self disclosed and is therefore tacit. In scenarios where employers are interested to find an appropriate candidate for a task for which no documentary proof exists, reputation based techniques can be employed.

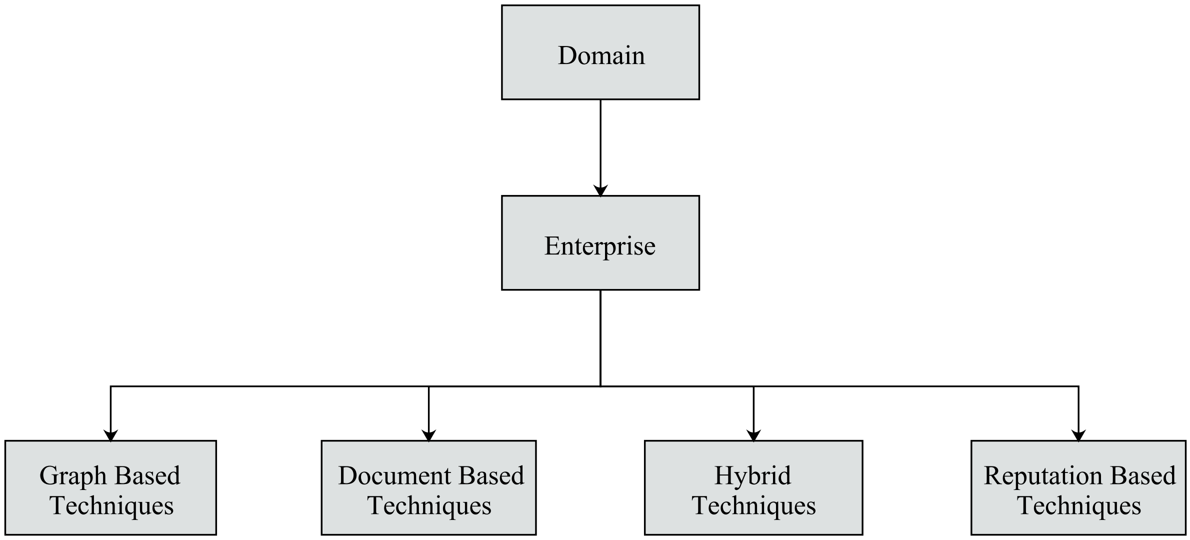

Figure 3.4: Techniques Of Expert Rank In Organizations

Given in figure 3.4 is the diagrammatic representation of the techniques utilized by researchers for employee ranking.

3.10.1.1 Microblogs

A large amount of work is done in the area of social networks and microblogs. Researchers have targeted microblogs [11, 12, 94, 103] by utilizing the features of the blog in order to identify an expert. The metric utilized by them includes the user answering most questions; that user is then regarded as an expert. These techniques however, fail to take into account the quality of those answers while ranking.

3.11 Expert Rank Literature Overview

Given below is the summary of the literature review of the key recent papers in the domain of expert ranking with their limitations and concepts. The limitations highlight the research gap addressed in this thesis. The limitations are categorized in terms of the inputs and processes. By $input$ we mean the parameters utilized by those researches while $process$ means the limitation in the actual working of the algorithm and the techniques.

Table 3.4: Expert Ranking Literature Overview

SNO.	Title	Limitations	
		Input	Process
1	M. S. Faisal, A. Daud, A. U. Akram, R. A. Abbasi, N. R. Aljohani, and I. Mehmood, "Expertranking techniques for online rated forums," Computers in Human Behavior, vol. 100, pp.168–176, 2019.		Pagerank based with limitations of negative referral and collusion
2	M. Kokkodis, "Reputation deflation through dynamic expertise assessment in online labor markets," in The World Wide Web Conference, 2019, pp. 896–905.		HMM Based with limitation of application to real scenarios.
3	M. Faisal, A. Daud, and A. Akram, "Expert ranking using reputation and answer quality of co-existing users." International Arab Journal of Information Technology (IAJIT),vol. 14, no. 1, 2017	Users who lie in answers(negative)	
4	M. Daltayanni, L. de Alfaro, and P. Papadimitriou, "Workerrank: Using employer implicit judgements to infer worker reputation," in Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. ACM, 2015, pp. 263–272.	Applicable to normally distributed variables	Based on Normal Distribution Calculation structure.
5	L. Zhang, X.-Y. Li, J. Lei, J. Sun, and Y. Liu, "Mechanism design for finding experts using locally constructed social referral web," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 8, pp. 2316–2326, 2014.		Negative Refferals
6	K. Wu, Z. Noorian, J. Vassileva, and I. Adaji, "How buyers perceive the credibility of advisors in online marketplace: review balance, review count and misattribution," Journal of Trust Management, vol. 2, no. 1, p. 2, 2015.	A Survey Study	A Survey Study

7	G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, "Expertrank: A topic-aware expert finding algorithm for online knowledge communities," Decision Support Systems, vol. 54, no. 3, pp. 1442–1451, 2013.		Negative referrals, Quality of answers
8	S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: crowdsourcing search for topic experts in microblogs," in Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012, pp. 575–590.	Structured data required, ground truth values needed, computationally expensive	
9	A. Pal and S. Counts, "Identifying topical authorities in microblogs," in Proceedings of the fourth ACM international conference on Web search and data mining. ACM, 2011, pp. 45–54.		Negative connections
10	J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010, pp. 261–270.		Negative connections, limitation of pagerank, attacks
11	A. Daud, J. Li, L. Zhou, and F. Muhammad, "Temporal expert finding through generalized time topic modeling," Knowledge-Based Systems, vol. 23, no. 6, pp. 615–625, 2010.		Change in expertise with time

3.12 Research Gap

The existing techniques of content credibility system, that are based on reputation models take into account the number of followers, number of likes, and number of positive links only. They do not take into account the number of unfollowing, down likes, and other negative interactions. Also, the reputation calculation structure is based on simple summation, that just sums up the number of positive interactions. The machine learning based approaches are already criticized for their utility in the content credibility systems where the size of information is increasing at a fast pace. Similarly, the expert ranking techniques are mostly dominated by the PageRank like algorithms that only takes into account the number of links without considering the quality of links. Some researchers have utilized reputation based models for expert ranking but again the calculation structure is based upon simple summation. The expert ranking has been explored in many domains but an emerging domain is the expertise ranking for large enterprises that are spread across continents. In such domains, expertise ranking can be an effective tool to identify the tacit or undocumented expertise of an employee. The reputation based approaches utilized in content credibility systems and expert ranks are prone to reputation attacks such as Sybil, Slander, and Whitewash. Thus to overcome these research gaps the authors have proposed techniques in the following chapters.

3.13 Chapter Summary

This chapter has three sections according to the literature review related to the Trust Judgment, Reputation based Credibility, and the Expert Ranking Techniques. The chapter begins with the definition of trusted computing, trust, trustworthiness. The security aspect of trust is also discussed. The authors have discussed the trust judgment process and its relationship with credibility analysis. The chapter also highlights popular trust models. The second section of the chapter gives a literature review of research related to content credibility. The approaches are categorized into human based and computational. The term reputation is defined. Reputation based approaches are discussed and compared with other approaches of content credibility in literature. The last section of the chapter gives detail of literature related to the expert ranking problem. The underlying techniques of the ranking algorithms are discussed in details along with the application of expert ranking algorithms to certain domains.

Chapter 4

Content Credibility Framework

This chapter gives details regarding the content credibility framework. The framework was developed after a detailed survey study, that helped in designing the architecture. The aim is to design a content credibility framework based on reputation information that can take into account positive, negative, or active and passive behaviors. To establish this, a hypothesis indicating the relationship of expertise with credibility is considered and then verified through a survey study. This chapter gives details of the proposed architecture and its different layers along with the mathematical model. The chapter then discusses the experiments conducted on different datasets. The results are discussed, compared, and analyzed with the baselines and presented according to standard performance metrics.

4.1 Architecture

In this section, the author has given details of the architecture. In our daily life, we always seek an expert for advice and recommendation. Thus we hypothesized that content evaluated by an expert can be regarded as credible. Let $U = u_1, u_2, \cdots u_n$ presents set of users.

$R = r_1, r_2, \cdots r_n$ presents reputation information of the users in the set U.

Then Maximum $(R) = E$ represents the Expert User. Thus Opinion of E is directly proportional to credibility (C) of Web information, presented mathematically as follows;

$$E \propto C \tag{4.1}$$

The initial hypothesis is that web content evaluated by experts is credible. Thus in this relation, the authors developed an architecture that identifies an expert. The reviews by the expert will render the content as credible. Michael Kohler et al.[62] also discussed a similar hypothesis in their work with results verifying it. To test the hypothesis a study was carried out. The survey study was double blinded since the participants did not know the authors and the research. Similarly, the authors were unaware of the identities of the participants. The survey study was conducted by a third party, thus resulting in an unbiased sampling. Survey questions were asked from the students of the University of Engineering and Technology Peshawar regarding their opinion about the tools that are effective in evaluating the credibility of the web information. The sample was chosen randomly and the participants were informed about the research and confidentiality of their responses. The tools are;

- The source of information
- Popularity of information
- Expert analysis of the information
- Recommendation by others

The question was asked in the context of blogs and social media. There were 80 participants who understood the questions and answered according to the given choices. Given below is the distribution of their opinion along with the ranking against each tool. In the case of Expert analysis, almost 35% of participants ranked it as the most important tool, while 44% considered it as the second most important tool. In the case of the source of information, 46% considered it to be the most important tool. The tool popularity of information, however, received 53% votes of least importance. The figures given below depict this distribution. In further data analysis, data was entered in the table where the rows represented the tools and the column entries represented the rank given by each participant. The average was calculated for each tool, the result showed that after the source of information, expert based analysis is considered to be the most important tool for content credibility assessment.

Thus from the results author developed an expert based framework for credibility assessment. Zieglar et al. [50] proposed a hybrid model that can effectively perform web information credibility judgment that involves assessment through computer and human judgment. Following this, the author has thus proposed a technique that terms information as credible if it is shared by a user who is considered most reputed and regarded as an expert of the context and area of interest. The reputation information can be gathered by the feedback provided by other users. If the rating

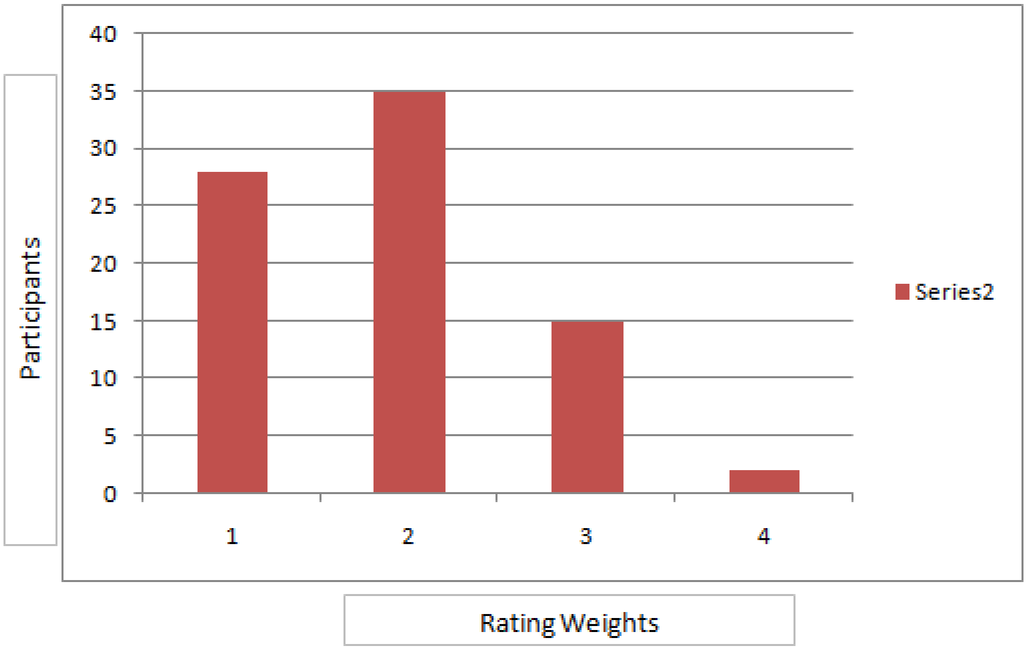

Figure 4.1: Expert Analysis

information is not available; the text shared by the user can be assessed for its sentiment value. Calefato et al. [104] found that sentiment plays an important role in the acceptance of the answers in question answer blogs.

A large amount of information is present on the websites, in online discussion forums, online question answer forums, Social networks, and other Microblogging sites. The author has particularly focused on the content available in social networks, microblogging sites, question answer forums, and the information available in large organizations in the form of emails or intra discussion forums. The proposed technique is a generic solution and applies to wide areas, however; experiments were conducted on a few including the social networking sites and Intra Organization Communication. The figure 4.4 shows the two dimensions to web content credibility assessment including human based judgment and computer based. The proposed framework has addressed these two dimensions by utilizing user feedback in the form of human judgment. While the reputation algorithm techniques are the computerized dimensions. The proposed architecture is layered and closed. This means each layer is dependent upon the next immediate layer. The architecture consists of four layers i.e interaction layer, reputation layer, expert layer, and the credibility layer. Figure 4.5 shows this architecture with interaction being the lowest layer. The user interacts through this layer. Finally, the last or top layer credibility is responsible to indicate the credibility level.

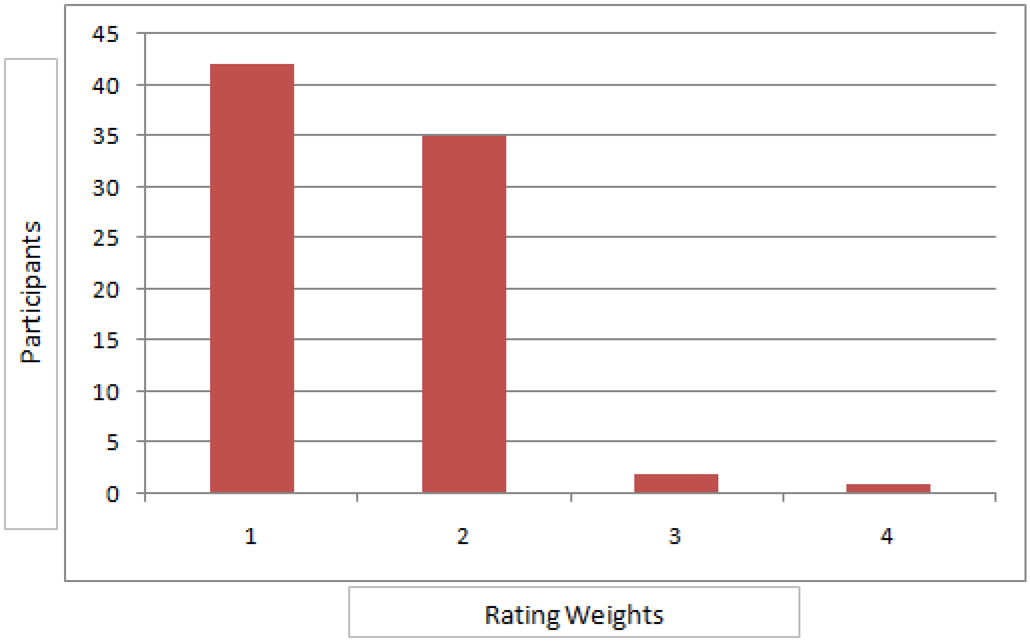

Figure 4.2: Source Of Information

4.1.1 Interaction Layer

In the case of the social network, an interaction can be a post, private message, like, tweet, retweet, following, upvotes, downvotes, rankings, etc. It is a generic term that can be adapted according to the target platform. Interactions as discussed in the work by Nepal S et al. [45] can be positive or negative/active-passive.

4.1.1.1 Parameters Of Interactions

An interaction between two users A, B can be through a post. User A posts a message, user B interacts by positively accepting it, shows a positive relationship however this could be negative as well. Another user C chooses not to respond thus remains passive. In another scenario such as in a large organization email, employee rankings, can all be treated as an interaction. The interaction data can be in the form of text, ratings, votes. The interaction layer is responsible for categorizing the interaction. If the interactions are present in the form of text, its sentimental value is used to categorize as either positive or negative. While if the interaction is a value, the cutoff decides its category. The layer is responsible to calculate the total number of positive and negative interactions.

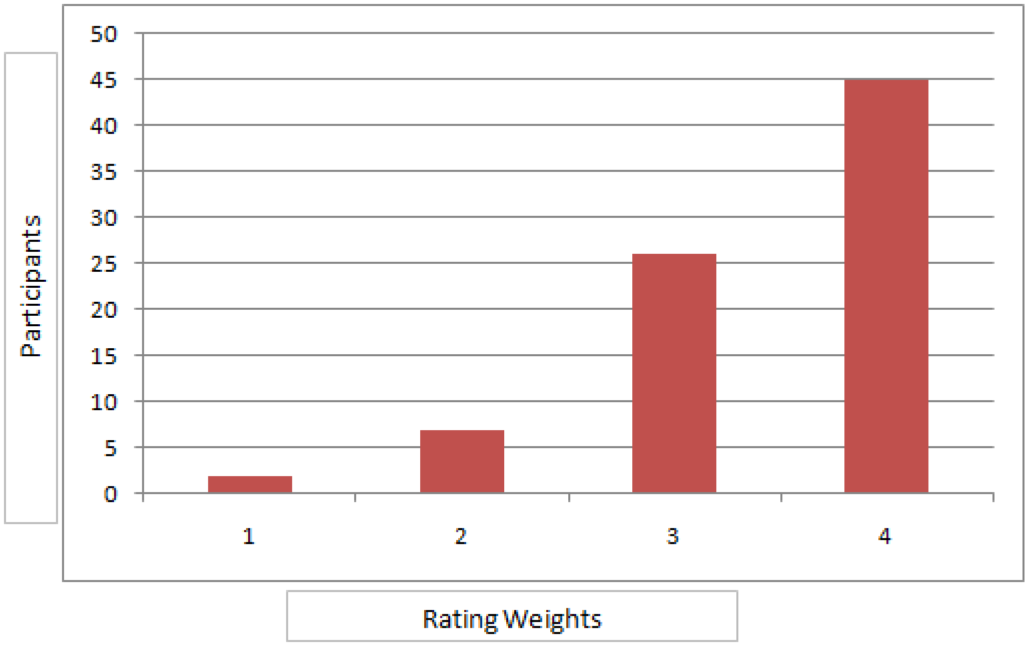

Figure 4.3: Popularity Of Information

4.1.2 Reputation Layer

Reputation is defined as "Overall quality as seen and judged by users" according to the Merriam-Webster's [34] online dictionary. The past behavior of the entities and opinion of others is utilized to find the reputation. The opinion is based on the past history of interactions.

4.1.2.1 Mathematical Model

Beta probability density function presents probability distribution of binary events [6].Posterior probability density function of binary events is discussed well in the textbook of statistics [105]. Unlike other expert techniques the proposed technique has firm basis of statistical theory. The beta family probability distribution is indexed by two parameters namely alpha and beta. The beta distribution function can be represented by:

$$f(p\|\alpha,\beta) = p(1{-}p) where 0 \leq p \leq 1, \alpha \geq 0,\ beta \geq with restriction p \neq 0 if \alpha \leq 1, p \neq 1 if \beta \leq 1 \tag{4.2}$$

$$f(p\|\alpha,\beta) = \frac{(\alpha+\beta)}{\alpha+\beta} \tag{4.3}$$

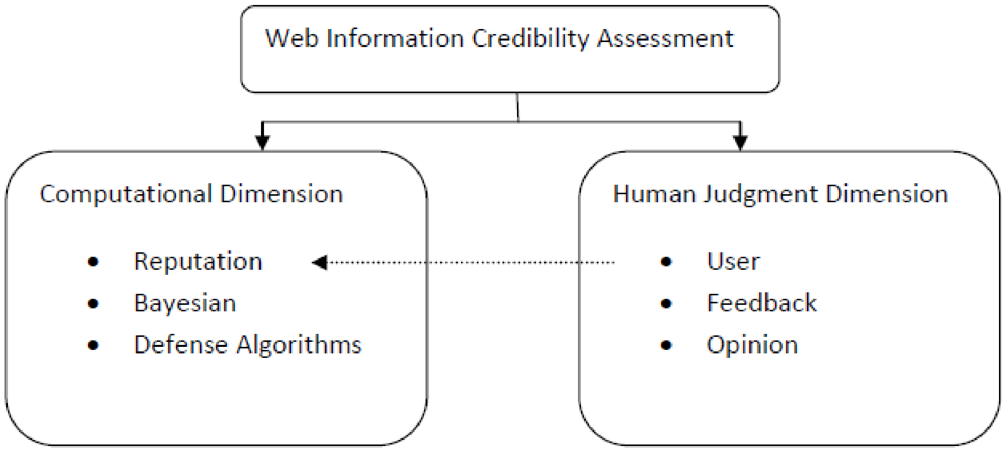

Figure 4.4: Web Information Credibility Assessment Dimensions

If the alpha and beta are of the same values then the beta pdf is 0.5. As the value of beta increases PDF(Probability Density Function) shifts towards right. Blei et al. [85] has proposed beta reputation systems since it can truly capture the feedback after an eCommerce transaction. The feedback is not a binary event, but they have categorized the feedback from the transactions as positive and negative representing satisfaction and dissatisfaction.

4.1.2.2 Expert Reputation

The numbers of positive and negative interactions are found. The values are then used to find the expected value. The expected value of all the interactions of a particular node is added to give the expert reputation of a node. The node with the maximum Expert Reputation value is considered to be the expert node in that particular context. If we add the Expert Reputation value of node i for all the contexts, we can get an overall expert level of the node. Beta Probability distribution [6] can be used to present the probability distribution of binary events. The beta family of probability distribution consists of two variables namely alpha and beta. The beta distribution is used to describe the subjective degree of belief. Thus it is most suited to model the relationship between the nodes in the network. The relationship shows the degree of belief of a node regarding the amount of trust it can place on the connecting node. The expected value of the distribution is given by

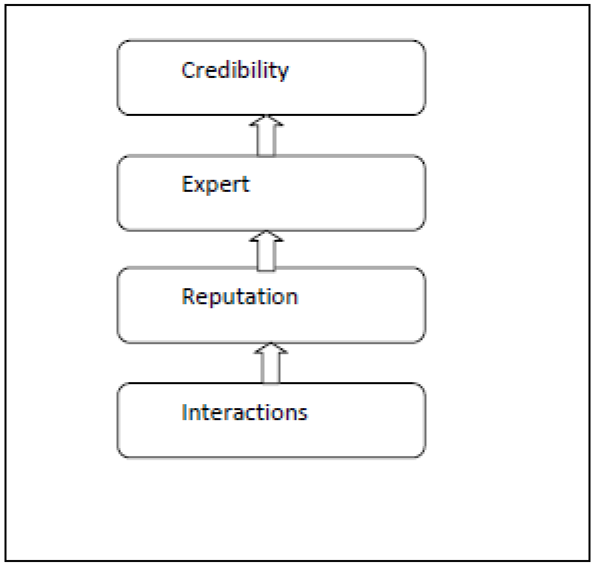

Figure 4.5: Layered Architecture

$$E(v) = \frac{\alpha}{\alpha + \beta} \tag{4.4}$$

A node in a social network might be having positive interactions in the form of a number of tweets or a number of comments received on a Facebook post in a particular context or subject. There can be negative interactions as well.

"Let the positive interactions be represented by alpha and negative interactions by beta. Let $'x'$ represents a particular context and$'A'$represents the number of activities in a particular context. $'M'$ is the total number of participants/nodes in the network. Thus we can compute the expert node in a particular context $'A'$ from equation 4.3, let's suppose y, y_1 represents the number of outcomes of alpha and beta respectively, that means after every y outcome we can expect y_1 outcome. In our framework, let p represent the observed number of outcomes for y and n represent the observed number of outcomes for y_1, then the following equations can be derived".

$$\alpha = y + 1 \tag{4.5}$$

$$\beta = y1 + 1 \tag{4.6}$$

$$E(v) = \frac{y+1}{y+y_1+2} \tag{4.7}$$

By Substituting the number of outcomes for y and y_1, the following equation is derived

$$E(v) = \frac{p+1}{p+n+2} \tag{4.8}$$

The Expert Reputation of a single node is then given by

$$E = \sum_{i=1}^{M-1} E(v) \tag{4.9}$$

here 'E' indicates the expected value(Reputation) of a node to behave in a particular context. Thus 'E' indicates the expected value of a node to behave in a particular context. The node with the highest 'E' value is referred to as the expert node in a certain context. Given in the next section is an algorithm based upon the above discussion. The input to the algorithm is the dataset that represents a graph of nodes. The algorithm evaluates the rank of each node, the node with the maximum reputation value is then regarded as the expert node. Algorithm 1 shows the part of the algorithm that categorizes the interactions into positive or negative based upon their type, which can be in the form of text, continuous values, or explicit values.

Algorithm 2 finds the reputation value of the node and then finally computes the expert node. Line number 2 and 3 computes the positive and negative interactions.Line 6 then calculates the expected value of the node. Finally Line 7 compares the expected value of the node to rest of the nodes, thereby bringing out the node with maximum expected value treated as expert node. The time complexity of this algorithm is exponential in nature i.e. $O(n^2)$. The interaction categorization algorithm is linear $O(n)$ in nature.

4.1.2.3 Time Based Analysis

The older interactions must be differentiated from recent interactions. In other words, if users had positive interactions quite a time ago, they must not be treated in the same manner as the positive interactions that took place more recently. By following Josang [106] definition we can incorporate this with positive and negative interactions in the following manner. Let's assume a node i had a sequence of interactions. Then the positive and negative interactions of node i are; $p = t$ and $n = t$.

Algorithm 1 Interaction Categorization

if $Interactions = Text$ **then**
  $result \leftarrow Senti(Text, i)$
  if $result = positive$ **then**
    $p \leftarrow p++;$
  else
    $n++$
  end if
else if $Interaction = value$ **then**
  Enter cutoff
  if $value > cutoff$ **then**
    $p++$
  else
    $n++$
  end if
else if $Interaction = explicit$ **then**

  if positive **then**
    $p++$
  else
    $n++$
  end if
end if

Algorithm 2 Reputation Based Expert Rank

Load Dataset
loop
  Compute no. of positive interactions p for node i
  Compute no. of negative interactions n for node i
  Compute Expected value for node i as
  $Ev \leftarrow p + 1/p + n + 2$
  Let T represent total number of nodes
  Compare Reputation of node i with T-i nodes
  $max \leftarrow i$
end loop
Compute max as Expert

Where't' is the time factor, if t is 1, it means all interactions are counted. While if 't' is 0 it means only the last interaction is kept.

4.1.3 Expert Layer

The expert layer of the framework is responsible for generating expert ranking according to the reputation score given by the Reputation Layer. Researchers [107, 108] proposed a technique whereby a user is ranked as an expert based on reputation information of co-existing users. The calculation structure adopted by this technique is a simple summation. Reputation mechanisms already employed for expert ranking are ad-hoc based. The proposed scheme is however based on a sound mathematical basis. The Bayesian-based reputation is the most effective mechanism in comparison to other techniques. Here the quality of interactions is utilized to find the reputation score of the user.

4.1.4 Credibility Layer

Using equation 4.1, this layer is responsible to associate the expert rank of the user with the content, that was either produced/reviewed/ edited by him. The user interacts with this layer to find the credibility of the information. Thus this layer can be considered equivalent to the user interface layer.

4.2 Experiment and Results

The authors conducted two experiments one experiment is carried out to evaluate the effect of categorizing the type of interactions upon rankings and its correlation with actual rankings. In the first experiment, a comparison is performed to the latest technique of content credibility [7]. The second experiment is conducted on a dataset extracted through a survey. The dataset is developed to have ground truth values that can be used for comparison purposes.

4.2.1 Performance Evaluation Metrics

The estimated reputation values generated through the proposed model are compared to the real values to evaluate the ability of the model in predicting rankings close to real rankings. Two performance indicators [109, 110] within this context are reported, the average absolute error between real and predicted values, and the linear correlation between real and predicted values. In addition to these indicators, Precision is also utilized. Precision is given by(tp stands for true positives, fp stands for false positives)

$$Precision = \frac{t_p}{t_p + f_p} \tag{4.10}$$

Precision means the probability with which the algorithm accurately generates the ranking i.e. it truly ranks an expert. Recall and F Measure is also calculated for the baseline and the proposed technique. The recall is given by

$$Recall = \frac{t_p}{t_p + f_n} \tag{4.11}$$

4.2.2 Experiment 1

We experimented on Dataset [111] that is of an intra-organizational network, where the interactions are weighted on a scale from 0-5 that defines the frequency of advice requested.

0 : I Do Not Know This Person

1 : Never

2 : Seldom

3 : Sometimes

4 : Often and

5 : Very Often

First, a ranked list of nodes was generated using only positive interactions. In the second instance using the cutoff, final calculations are carried out by using both categories of interactions i.e positive and negative. The two ranked lists differed from each other as can be seen in the given figures and the tables. The table 4.1 shows the ranked list through two methods in ascending order, the top nodes are at the bottom of the table. The variance of top3 nodes given in the table 4.1 was calculated to find how close they reflect original opinions. The results showed that the ranked list generated through both positive and negative interactions was closer to the mean as compared to the list generated through only positive interactions.

Table 4.1: Ranking With(+ve/-ve) And Without(+ve) Proposed Reputation Scheme

S.No	+ve/-ve	Positive	S.No	+ve/-ve	Positive
1	30	13	24	14	36
2	24	14	25	27	42
3	3	3	26	25	27
4	46	20	27	6	30
5	15	39	28	7	6
6	42	28	29	11	16
7	16	4	30	13	43
8	32	1	31	37	41
9	10	12	32	1	31
10	22	22	33	19	44
11	43	8	34	29	5
12	40	38	35	23	11
13	41	33	36	26	25
14	36	26	37	31	37
15	39	46	38	18	35
16	4	45	39	45	2
17	21	17	40	38	18
18	35	10	41	33	19
19	44	23	42	28	7
20	9	34	43	8	29
21	17	21	44	34	32
22	12	40	45	2	24
23	5	15	46	20	9

Table 4.2: Mean Average Error(MAE) Of Top Nodes

Ev Set1 (+ve/-ve)	Ev Set2 (+ve)	MAE Set1	MAE Set2
3.94	3.60	0.14	0.20
3.93	3.00	0.12	0.80
3.89	3.72	0.07	0.09

Ev = Expected Value, Positive = +ve, Negative = -ve

Table 4.3: Comparison Of Ranked Lists MAE1 Experiment1

MAE[baseline1]	MAE [Proposed]
0.50	0.14
0.40	0.12
0.07	0.07

4.2.2.1 Comparison To Baseline

An experiment is performed on the dataset to compare the proposed technique with the web content credibility technique [7] treated as a baseline technique. The authors have also compared their technique with a reputation calculation structure based on PageRank [76, 84] considered as baseline2 and normal distribution based reputation model(NDR) as baseline3 [42]. The Metric for comparison is user opinion. Mean Average Error (MAE) of the proposed ranking to the actual user opinion was found. The experiment revealed the following results. The top 3 nodes obtained from the baseline showed MAE of 0.5, 0.4, and 0.07. A comparison to the proposed technique showed that the MAE is much lesser. This shows the proposed technique is more effective in presenting the actual opinion of the user.

4.2.3 Experiment 2

Using the information through the filled survey forms we developed a dataset of the students of a class in graph format. The survey was double blinded, it was conducted with students of the University of Engineering and Technology Peshawar. An edge shows the friendship relationship. The weight of the edges represents the interactions of the students. The weights are assigned based upon the ratings provided by the students for their friends regarding the interactions related to subject knowledge.The weights are scaled in the range of 1-5,i.e.

1 : Nil

2 : Fair

Table 4.4: Comparison Of Ranked Lists MAE2 Experiment1

MAE[baseline2]	MAE [Proposed]
0.30	0.14

Table 4.5: Comparison Of Ranked Lists MAE3 Experiment1

MAE[baseline3]	MAE [Proposed]
0.16	0.14

Table 4.6: Comparison w.r.t Precision ,Recall, Correlation Experiment 1

Metrics	Baseline1	Baseline3	Proposed
Precision	0.042	0.021	0.06
Recall	0.110	0.050	0.30
F Measure	0.060	0.029	0.10
Correlation	0.240	-0.076	0.39

3 : Good

4 : Very Good

5 : Excellent

This dataset is undirected where nodes represent the students and the edges represent the rated expert friendship relationship among them. This dataset has been taken as a special scenario where each node has three friends showing an equal number of links for each node. Such a special scenario would highlight the shortcomings of previous link based expert techniques based on the PageRank thereby showing the capability of the proposed technique. Given in the table 4.7 is the ranked list of students according to the proposed algorithm. The original dataset presents the ground truth. The neighbors or close relations possess the same level of knowledge. Thus the ratings given by students for their friends are considered the ground truths. The average mean is calculated for every student.

Table 4.7: Rank List2

Nodes	EV	Nodes	Avg	Nodes	EV	Nodes	Avg
13	6.60	13	0.11	21	1.7	40	0.035
2	5.570	2	0.099	4	1.69	3	0.034
37	5.000	8	0.087	30	1.69	12	0.030
17	4.99	6	0.08	3	1.66	21	0.030
32	4.90	17	0.08	12	1.66	30	0.029
8	4.200	32	0.08	20	1.66	26	0.027
6	4.150	37	0.077	34	1.66	28	0.026
25	4.060	10	0.058	26	1.65	34	0.025
38	4	7	0.056	28	1.65	27	0.023
14	3.33	38	0.054	16	1.63	31	0.023
10	3.30	5	0.05	23	1.63	16	0.020
7	3.26	14	0.05	27	1.63	20	0.020
33	3.26	25	0.05	31	1.63	23	0.020
40	3.10	33	0.044	19	0.85	19	0.017
35	2.52	35	0.04	9	0.833	9	0.0158
5	2.50	39	0.039	22	0.833	22	0.014
39	2.50	4	0.038	24	0.66	24	0.003
1	0	1	0	18	0	18	0
11	0	11	0	29	0	29	0
15	0	15	0	36	0	36	0

EV = Expected Value

4.2.3.1 Comparison with Baseline

The comparison is made to calculate precision-recall. Precision is given by (tp stands for true positives, fp stands for false positives) Precision means the probability with which the algorithm accurately generates the ranking i.e. it truly ranks an expert.

Precision @k is an important metric in the field of information retrieval to find the percentage of accurately discovered documents. K stands for a certain number of instances. For example, if k is 5 it implies the number of accurate discovered items in the top 5 entries in comparison to the ground truth values. We calculated P@3, which came as 0.66 . Overall precision is 0.32 while of baseline it is 0.05. Pearson Correlation is calculated of the two ranked lists, to find if the ranking produced by the proposed algorithm closely correlates to the human judgment. The correlation of top5 ranked list is 0.87 compared to the baseline technique where correlation is 0.55. The correlation result shows that the expert ranked nodes positively relates to human judgment. Figure

Table 4.8: Comparison Of Ranked Lists MAE3 Experiment2

MAE[baseline1]	**MAE [Proposed]**
1.20	0.20
0.20	1.20
0.86	0.46

Table 4.9: Comparison w.r.t. Precision, Correlation Experiment 2

Metrics	**Baseline1**	**Baseline3**	**Proposed**
Precision	0.05	0.05	0.32
Recall	0.02	0.09	0.40
F Measure	0.028	0.064	0.36
Correlation	0.55	-0.13	0.87

3.8 shows this correlation, whereby the rank list generated though the proposed algorithm closely relates to the rank list according to average weights. Experiment on the dataset to compare the proposed technique with the web content credibility technique [19] treated as baseline technique was carried out. The Metric for comparison is user opinion. Mean Average Error(MAE) of the proposed ranking to the actual user opinion was calculated. The experiment revealed results shown in the table 4.1. The top 3 nodes obtained from the baseline showed Variance of 1.2, 0.2, and 0.86. The comparison showed that the variance is much lesser for the proposed technique except for the second ranked node. This shows the proposed technique is more effective in presenting the actual opinion of the user. Thus the ranking of the proposed technique is more accurate in showing the credibility level of the content associated with them. Figure 4.6 and figure 4.7 show the comparison of techniques with respect to correlation, precision, and MAE of the two experiments.

4.2.4 Analysis

The results from the two datasets show that the mean average error for the proposed technique is lesser as compared to the previous baseline models. In Experiment 1, the working of the proposed model is evaluated and it was found that the inclusion of both positive and negative types of interactions yields lesser mean average error as compared to the scenario when only positive interactions are utilized as is done by the previous technique where only positive interactions are counted towards a user's popularity/engagement. The second experiment conducted on a dataset compiled through a survey showed again that the mean average error for the proposed technique

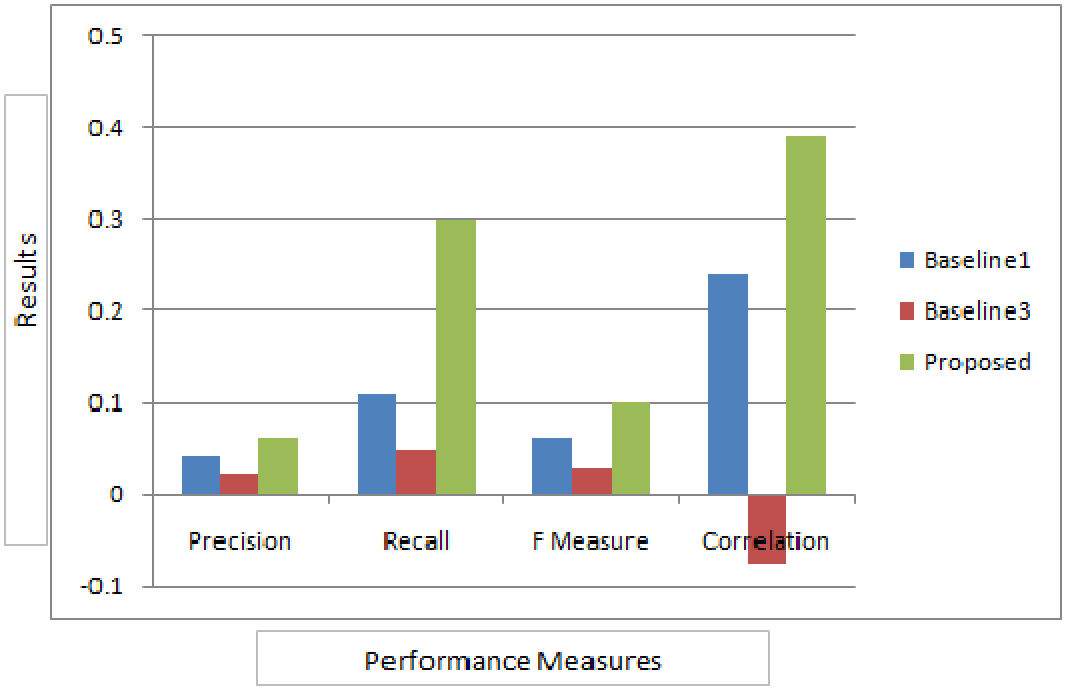

Figure 4.6: Comparison w.r.t. Precision, Correlation Experiment1

is lesser, compared to the baseline techniques. In these experiments, Precision and Correlation results also supported the effectiveness of the proposed technique. The results of the techniques are however not promising in experiment 1, but the performance of the proposed technique is still better. The dataset from experiment 1 is dense as compared to the second dataset that is sparse. The percentage difference of precision of the baseline and proposed technique is 27 % as compared to experiment 1 where it is 18%, giving us an insight that incase of sparse data, the proposed algorithm has almost 27 % more precise results while in case of the dense dataset it 18%r more precise only. A similar trend is found for Recall values. The performance of baseline 2 in terms of precision and correlation is too poor to report. In the case of baseline3 again the results for the proposed techniques are better following almost the same trend. The proposed technique works better with sparse datasets, that is the case in the problem scenario, where it is not necessary that a particular node might be having interactions with every other node and vice versa. Baseline2 which is PageRank based looks at the number of connections regardless of the quality of the connections. Baseline 3 that is NDR based also performs less, since it produces less accurate results for sparse data than for dense [112]. Observing results from Experiment1 alone, it is evident that results from NDR based technique are less then baseline1, while for Experiment2 performance of NDR(baseline3) is better than baseline1, due to the fact that although the experiment1 dataset is

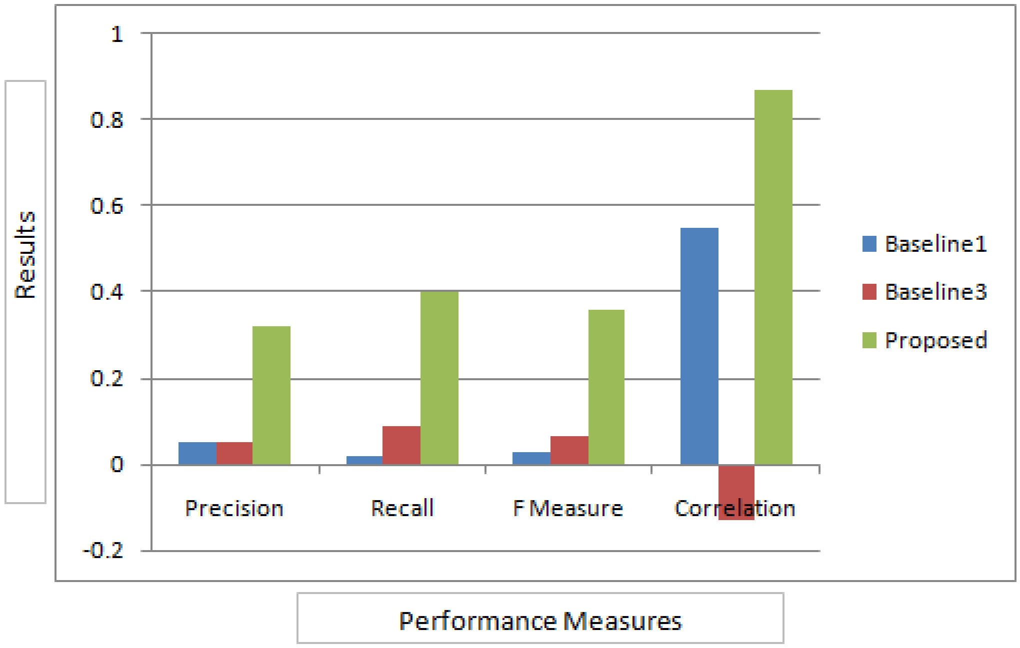

Figure 4.7: Comparison w.r.t Precision, Correlation Experiment2

dense, the ratings are distributed normally as compared to the dataset from Experiment2 where ratings are on the higher or lower end and not normally distributed. Thus we can summarize the findings that the proposed technique produces better results regardless of the density of the dataset and pattern of ratings, while the other two techniques appear to be dependent on these two factors. The PageRank based baseline has already produced poor results due to the inability of taking the quality of interactions and negative referrals into account. We have also calculated the correlation for the baseline techniques and the proposed against the null hypothesis i.e. “There is a linear relationship between original data and predicted data for all the four techniques”. The null hypothesis stands true for the three techniques the baselines(baseline1, baseline3) and the proposed technique. However, the hypothesis was rejected for the PageRank(baseline2) based technique. The significance value for baseline3 is negative as shown in figure 4.6 that shows a negative relationship whereby if original data is showing ranking in highest to lowest order the predicted data is mostly in opposite direction. The results of the proposed technique are significant with a value of 0.39 for experiment 1 and 0.87 for experiment 2. The correlation in experiment 2 is also negative thereby indicating the opposite relation between original and predicted data as shown in figure 4.7.

4.2.5 Identifying Features For Categorization

In this section, the author has discussed the properties of the blogs that could be utilized for feature selection of the text posted on blogs and social networking sites. For this, a question was asked from the participants to rank the following in order of their importance.

- User who answers most frequently
- User who posts lengthy answers
- Users who post some reference link
- Users who answer immediately

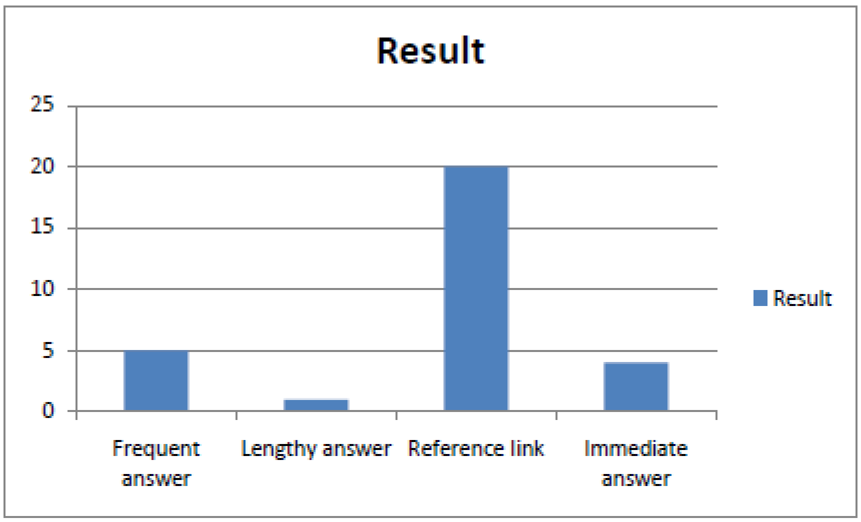

Figure 4.8: Comparison Of Features

The survey showed the above ranking scheme for the features. Most of the users were of the opinion that if an answer contains a reference link then it can be more reliable. The second rank is for frequent answerer as is utilized by some of the earlier works. The above ranking can be utilized as a weighing factor while categorizing the interactions as positive or negative.

$$Feature = \sum_{i=1}^{n} Ui, f * Wt \tag{4.12}$$

Where U stands for the user, f for a feature and Wt the respective weight of the feature. This information is utilized for feature selection of interaction categorization. The rating scheme derived from the survey results is helpful in assigning importance to different features. For example, the text of the post contains reference links and if timely answered could be categorized in a positive domain. The threshold of these features is not decided.

In scenarios where ratings are given on a particular scale as is the case of the dataset developed through this survey, there is a need to identify a cutoff point to categorize the ratings to positive or negative domains. In this survey, the author has used the average as cutoff according to the participant's opinion expressed through the survey question. The average value was 3.1 thus, values above 3 are considered to be positive while below 3 are considered negative. Such a form of binary categorization varies with scenarios and also depends on historical data.

4.3 Chapter Summary

The chapter discusses the proposed approach towards content credibility assessment. The hypothesis is evaluated through a survey study. The results show that an expert opinion is a useful tool in the assessment of the credibility of the content. The hypothesis is further verified by the application of the proposed scheme to the dataset developed through the survey study. The experiment is conducted on two different datasets, the results are shown in the form of graphs and tables. The results show that opinion generated through experts is related to credibility, this is proved through Precision, MAE, Correlation. The Analysis of the results builds an argument towards the effectiveness of the results. It compares the results from two experiments revealing important insight that the proposed scheme is independent of the pattern of ratings and density of data. The last part of the chapter also highlights the survey results related to different features that can be utilized for interaction categorization.

Chapter 5

Reputation Based Expert Rank

The expert ranking is a problem of the World Wide Web and is approached through different techniques by researchers as discussed in the literature review section. Reputation based systems utilize information of the past history of an entity for its reputation rank. This information can be directly obtained or is obtained from other peers. Since the problem under discussion involves the identification of an expert that is not well known, reputation information can be an efficient and effective tool for identification and ranking. This is the sub problem of the major problem of content credibility of this thesis, the proposed technique of the major problem is a solution of this problem as well with the specific issue of ranking employees in large organizations. In this chapter, the author has given further details in the context of expert ranking for organizations referred to as EER i.e. enterprise expert ranking. Thus the objective of this chapter is to discuss the proposed technique that can rank employees of large organizations where the expertise of an employee is tacit or undocumented. In such a scenario the graph or document based techniques does not serve the purpose; therefore reputation based technique that utilizes different kinds of interactions is proposed. The proposed architecture includes two basic modules. One is responsible for the categorization of past interactions into positive or negative categories. While the other module utilizes the beta probability density function to calculate the rank. The experimental results show a comparison to existing expert ranking techniques. Given below are the key modules of the expert ranking.

5.1 Problem Formulation

In this section authors have defined basic terminologies utilized in EER followed by a formal definition of the problem of Expert Employee Ranking.
Interactions Categorization: Interactions are categorized as alpha or beta, whereby alpha represents all kinds of positive relations, feedbacks, ratings, for example in the case of social media they can be the number of followings. While beta represents all kinds of negative interactions, relations, feedbacks, ratings. For example number of unfollowings represent beta in a scenario of social networks for instance.

Expert Employee: An employee is considered as an expert if he possesses the highest level of skills pertaining to his subject area while Tacit Expertise refers to experts with the highest level of skills for which there is no documentary proof.

Definition (Enterprise Expert ranking using Employee Reputation): Let $E = \{e1, e2, e3, \ldots, en\}$ be the set of expert users. Ev represents the expected value. Let I be the set of all interactions in which user$U$ has participated, where $I = \{i1, i2, i3, \ldots, im\}$ and categorization results in $i = \{alpha/beta\}$ and $U = \{u1, u2, u3, ..., un\}$ such that Ui, has some documented and undocumented skills. Authors assume that a userUi is a member of E if he has participated in interaction Ii , if and only if the $Rep((Ev), Ui, I)$ is maximum.Where $Rep()$ utilizes beta probability density function to compute the expected value for set I and Ui.

5.2 EER: Enterprise Expert Reputation

The employee expert ranking system, consists of two major modules including a module that categorizes the interaction while the other that utilizes the beta probability function as a reputation calculation structure for ranking, as shown in figure 5.1.

5.2.1 Categorization Of Negative And Positive Interactions

Since interactions can be either positive or negative, they can be categorized based on various parameters, in addition to the sentiment analysis technique. If the interactions present a scaled opinion set of continuous values as in the case of datasets DS1, DS2, DS3 an optimum cutoff point needs to be decided for binary classification. Those above the cutoff can be considered as positive and the lesser ones as negative interactions. There are certain other techniques[104]that can be employed for identification of the cutoff point with accuracy, but most of them require ground

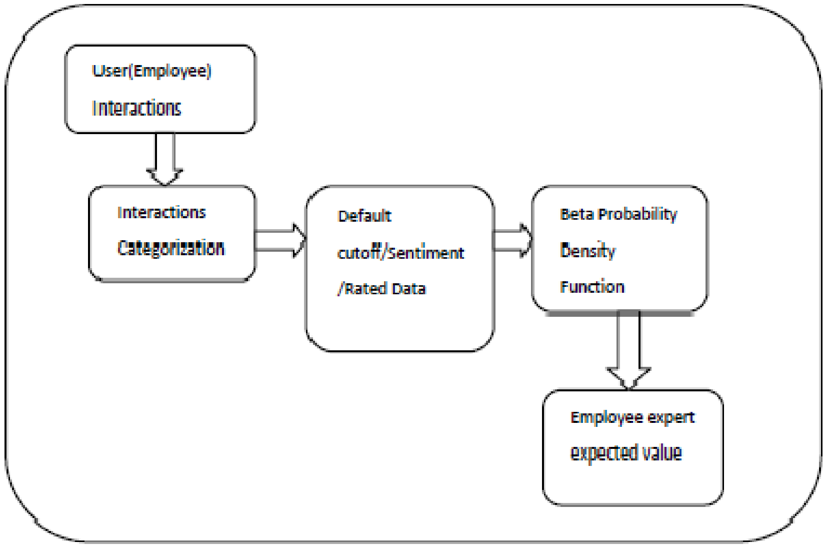

Figure 5.1: EER Architecture

truth values. However 0.5 and mean probability are also satisfactory techniques for cutoff points of the datasets. In certain scenarios, the interaction is already weighed as positive or negative for example stack overflow microblogs and etc.

5.2.1.1 Sentiment Analysis

Sentiment analysis is the technique that captures the hidden meaning of the sentences. It is the branch of opinion mining. Generally, it classifies web comments into positive, negative, and neutral categories. For this mostly naive Bayes algorithm is employed since it produces almost 75% accurate predictions [113, 114]. In the first step, filtration is done to remove any unwanted text from the messages. In the second step, training is performed with the bag of words, and in the last step, the prediction is carried out. WordNet method is the classical one, whereby the presence of positive or negative words is used to find the sentiment of a sentence. For training purposes, certain positive and negative words are used. There are certain datasets [113] available as well that have already highlighted the positive and negative interactions through sentiment analysis.

5.2.1.2 Other Methods

This method depends upon the scenario, for example in the case of social networks or micro-blogs, as discussed in [113] mention links, answer count, answer length, mention quotes can be utilized as the attributes. However, defining the threshold of those attributes requires certain more methodologies for example fuzzy logic, etc. In the case of a large organization where employees are allowed to rate each other for the expertise level on a certain scale for example subject knowledge, practical exposure, project, etc can be employed as the candidate attributes. For example, the number of replies is one such parameter. The minimum number of replies can identify the threshold. Replies above the threshold can be considered positive while lesser can be considered as negative interactions.

5.3 Experiments And Results

Three different experiments are carried out on the datasets, to verify the functionality of the proposed technique. The tests are aimed at verifying the precision of the rankings, the ability to evaluate the quality of interactions, and adaptation to dynamism against the baselines.

5.3.1 Datasets

The framework was tested on three different datasets; the detail of these datasets is given in the table below. The datasets are available freely for research purposes [111].

DS1: In the first network, ties are differentiated in terms of the value placed on the information or advice received ("For each person in the list below, please show how strongly you agree or disagree with the following statement: In general, this person has expertise in areas that are important in the kind of work I do."). The weights in this network is also based on a scale from 0 to 5.

0 : I Do Not Know This Person

1 : Strongly Disagree

2 : Disagree

3 : Neutral

4 : Agree

5 : Strongly Agree

DS2: In the second dataset, the interactions among the researchers are differentiated in terms of advice ("Please indicate the extent to which the people listed below provide you with information you use to accomplish your work"). The weights are based on the following scale:

0 : I Do Not Know This Person/I Have Never Met this Person

1 : Very Infrequently

2 : Infrequently

3 : Somewhat Infrequently

4 : Somewhat Frequently

5 : Frequently

6 : Very Frequently

DS3: The third dataset is based on the employees' awareness of each others' knowledge and skills ("I understand this person's knowledge and skills. This does not necessarily mean that I have these skills or am knowledgeable in these domains but that I understand what skills this person has and domains they are knowledgeable in"). The weight scale in this network is:

0 : I Do Not Know This Person/I Have Never Met this Person

1 : Strongly Disagree

2 : Disagree

3 : Somewhat Disagree

4 : Somewhat Agree

5 : Agree

6 : Strongly Agree

5.3.2 Baseline Techniques

Comparison to a recent graph based technique based on PageRank i.e. expert rank[12] and worker rank[17]technique specifically designed to rank employees in an enterprise according to expertise. The technique is unique in a way that unlike other ranking techniques they have utilized reputation information that has a basis on normal distribution reputation. Another reputation based technique[16]utilizing Hidden Markov Model(HMM) as a reputation calculation structure is also treated as the baseline. These techniques are treated as the baseline for comparison purposes. The document based techniques could not be utilized for comparison purposes due to the initial hypothesis of the research that states the absence of documentary proof for the problem under discussion.

5.3.3 Experimental Setup

The experiments were performed on standard i5@ 2.5 GHz machine. RAM 4.0 GB. Windows 7 operating system is used. Programming of the algorithms is done in Matlab 2014.

5.3.4 Experiment 1: Rank Match Test

This experiment is aimed at finding if the proposed technique is successful in the representation of the ratings given by the nodes. The tests are conducted to do the comparison against baseline techniques, that is firstly a graph based approach that is based on page rank algorithm [37]called as BL1. Second baseline [17]is the recent reputation calculation structure adopted by researchers in generating the expert ranking of workers in an enterprise referred to as BL2 in text. The researcher adopted a weighted average technique based on NDR i.e. normal distribution reputation [42]. Comparison to document based techniques stands invalid since the problem under discussion has the hypothesis of finding an expert when no documentary evidence exists or supports the decision. The authors conducted tests on DS1, DS2, and DS3. The estimated reputation values generated through the proposed model are compared to the real values so as to evaluate the ability of the model in predicting rankings close to real rankings.

Performance indicators including the average absolute error [110] and correlation between real and predicted values are reported in the literature [109].The result are shown in Table 5.1 and Fig 5.2.

Table 5.1: Comparison Of Ranked Lists MAE3 Experiment1

Datasets	**Avg Weight**	**MAE[EER]**	**MAE [BL1]**	**MAE[BL2]**
DS1	3.80	0.10	0.30	1.60
DS2	2.76	0.07	0.56	1.59
DS3	4.50	1.14	2.30	3.30

Table 5.2: Comparison Of Techniques w.r.t Precision

Datasets	**BL2(Worker Rank)**	**BL1**	**EER**
DS1	0.0	0.0	0.06
DS2	0.1	0.15	0.20
DS3	0.0	0.10	0.20

The results from precision, mean average error metrics show that the proposed EER technique is more accurate in representation of the expert level of a node.

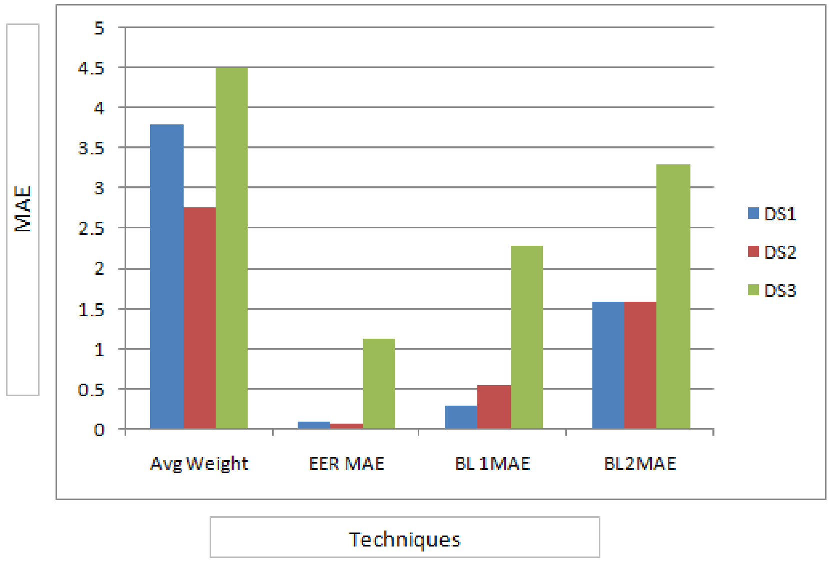

Figure 5.2: Comparison Of Techniques w.r.t MAE

5.3.4.1 Analysis

The MAE of the proposed technique EER is lesser as compared to baseline techniques for all of the three datasets. Similarly, the precision results show better performance of the proposed technique, however, the precision results of baseline2 are almost zero, the weights given in BL2 are based on Normal distribution. The performance of BL1 i.e. a graph based technique is better than BL1. However it also yields zero results for DS1, this is because in DS1 every node has an equal number of connections. The BL1 technique that is based on PageRank is unable to handle such cases, even the weighted version is also incapable to produce encouraging results. An important insight is the fact that the mean average error for dataset DS3 is greater for all the three techniques including the proposed technique as compared to the other two datasets. Careful observation led the authors to the conclusion that the ratings in DS3 are almost average or around the average value. Therefore the Mean Average Error in the case of DS3 for the three techniques is a bit more than the rest. Thus previous techniques and the proposed technique observe the same behavior for the DS3 kind of dataset.

5.3.5 Experiment 2:Interaction Categorization Test

The second test was carried out to demonstrate the new ability of the proposed EER in the identification of the difference between negative and positive interactions thereby solving the issue of negative referrals of graph based techniques. The experiment was conducted on DS1 and DS2.

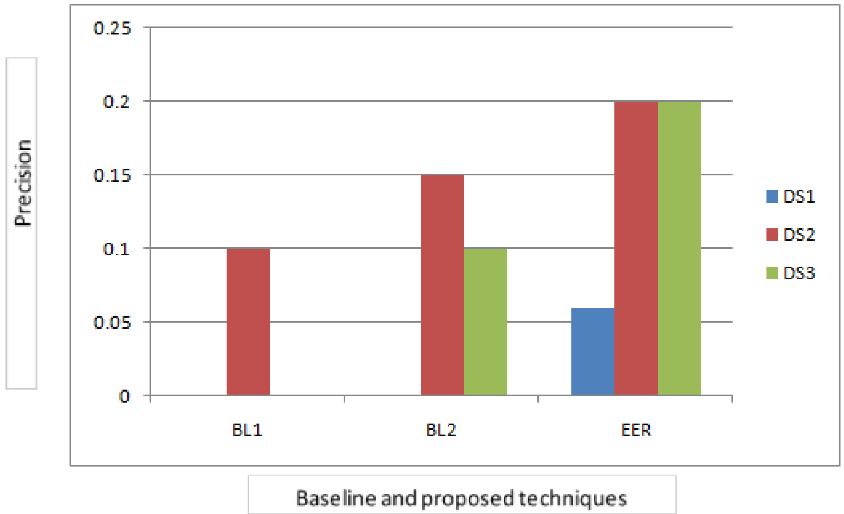

Figure 5.3: Comparison Of Techniques w.r.t Precision

First, a ranked list of nodes was generated using only positive interactions. In the second instance using the cutoff, both positive and negative interactions were utilized in calculations. The variances of top3 nodes were calculated to find how close they reflect original opinions. The results showed that the ranked list generated through both positive and negative interactions was closer to the mean as compared to the list generated through only positive interactions. The table 5.3 shows the MAE of the ranked list of dataset DS1 through two methods. The figure 5.4 shows that both ranked lists are not overlapping with very few instances.

Table 5.3: MAE Of Top Nodes

Average Weight set1(+ve)	**Avg Weight set2(+ve,-ve)**	**MAE[Rank Set1]**	**MAE [Rank Set2**
3.94	3.60	0.20	0.14
3.93	3.00	0.80	0.12
3.89	3.72	0.09	0.07

The same experiment was conducted on the dataset DS2, in this dataset, there is almost 60% overlap. For this manual checking of the dataset was carried out. It resulted that most of the nodes had a rating 3 or above thus most of them had positive interactions. Thus, the difference in the result was minimum. The graphical representation in figure 5.5 shows the comparison and overlap of two lists

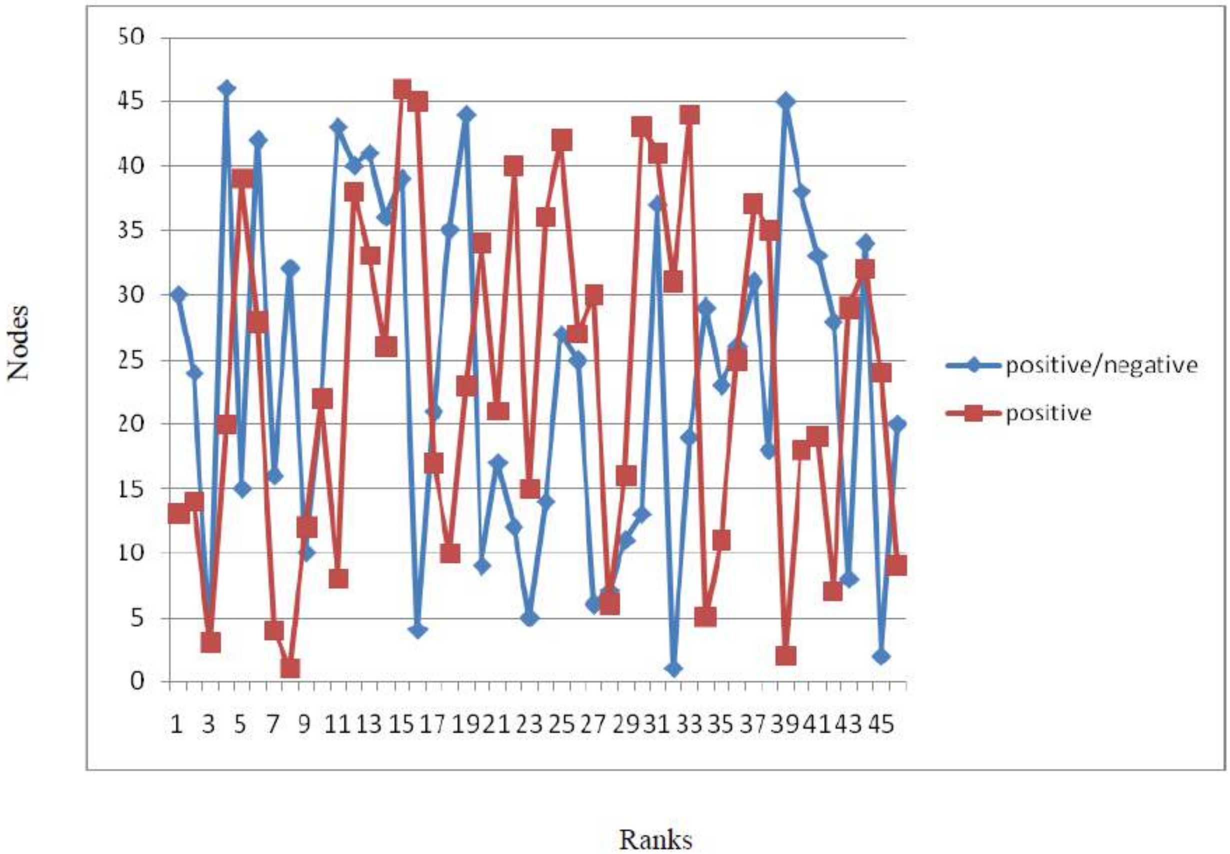

Figure 5.4: Interaction Overlap Graph

5.3.5.1 Interaction Test

In order to verify the novice ability of the algorithm to compute the expert rank based on different types of interactions, the author took the number of interactions of the expert node 27 from dataset1. The author divided these interactions into half and assumed half of them as positive and half as negative interactions. The Expected value was 0.99. Then we took 2/3 of the interactions and assumed them to be positive and 1/3 as negative interactions and again calculated the expected value that was 0.66. In the third test, we utilized 2/3 of interactions as negative and 1/3 as positive interactions, this time expected value was calculated to be 0.33. When compared with other latest techniques[11, 12] to identify the expert node in a social network, expert identification based on different types of interactions cannot be identified. The beta distribution due to its characteristic alpha and beta parameters can differentiate both types of interactions in order to identify the expert node. Fig 5.6 shows the variation in the probability distribution with the change in the ratio of

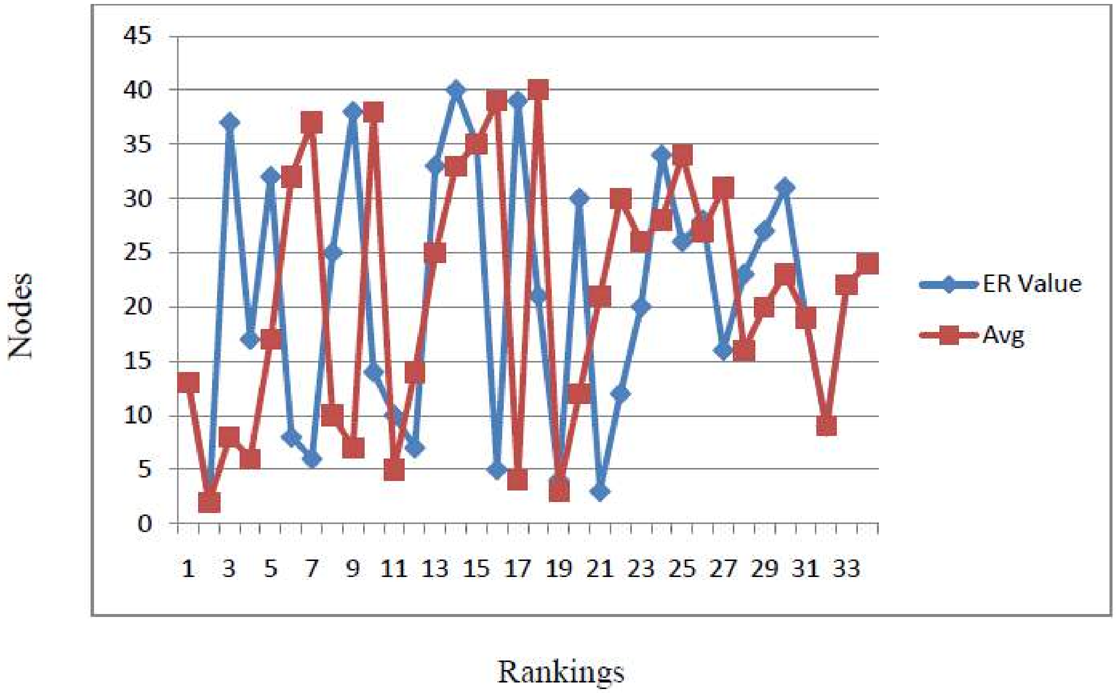

Figure 5.5: Interaction Overlap Graph

positive and negative interactions.

Another test was carried out to demonstrate that the EER technique can differentiate interactions,i.e. positive and negative interactions thus highlighting this new feature of the proposed EER. For this authors utilized DS2. The file was manipulated and the weights of the expert node identified through the EER technique that is node1 were manipulated, while the expert node identified through the BL1 technique was Node3. Proposed and baseline(BL) algorithms were run on the manipulated dataset. Since the proposed technique is able to differentiate both positive and negative interactions, therefore a new node is identified as an expert node i.e. Node2 in place of Node1. The BL1 algorithm ranked the same Node3 as an expert node even after manipulation, this is because PageRank cannot estimate the negative connections and referrals. This leads to the conclusion that the proposed technique can identify both types of interactions and can rank/identify the experts accordingly while the baseline techniques are unable to address this and only takes into account the positive interactions.

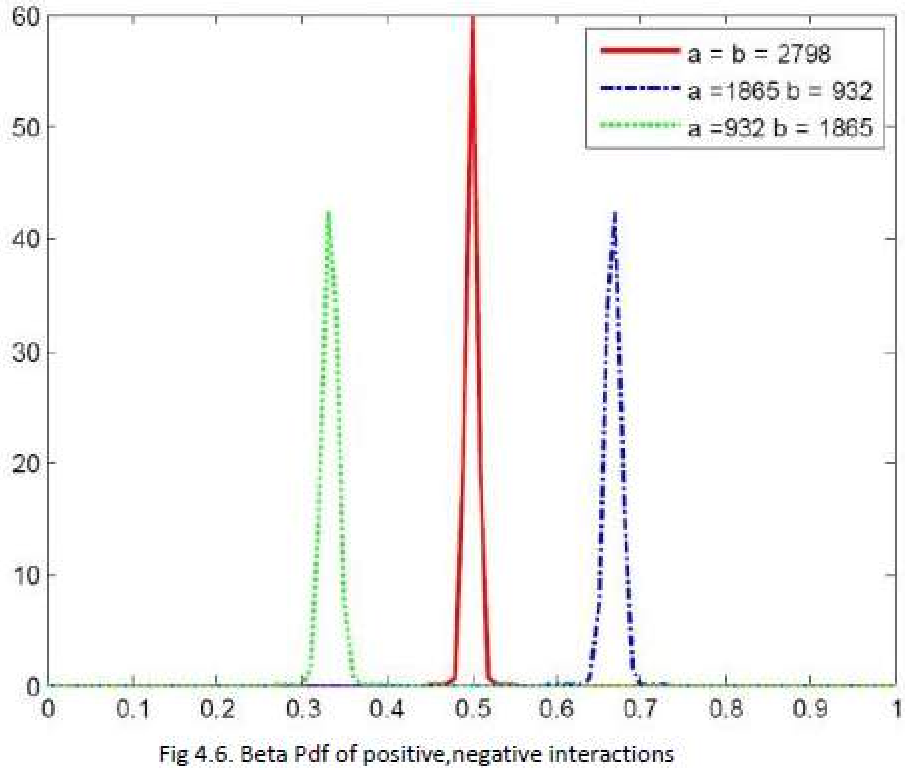

Figure 5.6: Beta Pdf Of Positive, Negative Interactions

5.3.5.2 Evaluating Quality Of Interaction

Quality of interactions between the social network nodes can be judged by employing either certain sentiment analysis techniques or by using certain cutoff techniques of binary classification for continuous values. The datasets 2,3,4 provides ratings given in the continuous form. The problem with such a dataset is to classify them into either positive or negative class. The author researched [115] various techniques for this purpose, the most suitable technique is either using a default parameter of 0.5 or by using mean probability. Mean probability is the calculation of mean value from the range of the values. In the given datasets the range values are 0-6. The mean is 3 thus the cutoff. In the default method, the cutoff would be 3.5. Thus the values above 3 are considered to be of positive class while the values below 3.5 are considered to be of negative class.

5.3.6 Experiment 3: Dynamic Behavior Test

In order to evaluate whether the proposed technique can adapt itself according to time and history, a test was carried out to verify the performance against the third baseline that utilizes HMM for repu-

Table 5.4: Expected Value (observation probabilities) With Varying History

Interaction History	**Expected Value(EER)**	**Expected Value (HMM)**
All History	0.40	0.45
Latest	0.3	0.20
Latest 3	0.40	0.40
Latest 5	0.42	0.50
Latest 7	0.50	0.60

tation calculation. The literature revealed that HMM outperforms beta probability based reputation since it is able to address the issue of reputation staticity. However it is to be noted that HMM is a generalized version of the beta probability distribution, also the HMM is limited by the duration of the state that follows a geometric pattern usually unsuitable for real life scenarios. HMM-based technique becomes more complex if combined with learning techniques, also the complexity of this model increases with parameter estimation [116]. Comparatively beta probability based reputation techniques are simple to interpret for daily life examples since it is parameterized by only two variables.

The basic thing to compare between the two techniques is the in terms of the response to change in behavior by the two systems. Thus authors carried out a specific scenario of a node with a history of 10 interactions with different other nodes. Using the proposed technique with a time factor of t = 0, the expected value of 0.4 is calculated against the time factor value of t=1, the expected value is 0.3. where t=0 means usage of all interactions of the history while t=1 means utilization of only recent interactions. Figure 5.7 and table 5.4 shows the expected value from proposed and HMM-based techniques in case of variable history.

The result shows that both models respond to change in history however the behavior of the HMM-based technique is steeper when compared. The scenario under discussion in this section of the thesis is the identification of tacit expertise, that does not suffer a rapid change thus HMM technique can be an expensive choice in terms of the limitations discussed above. HMM is suitable for finding out deviation of behavior from the previous well established behavior.

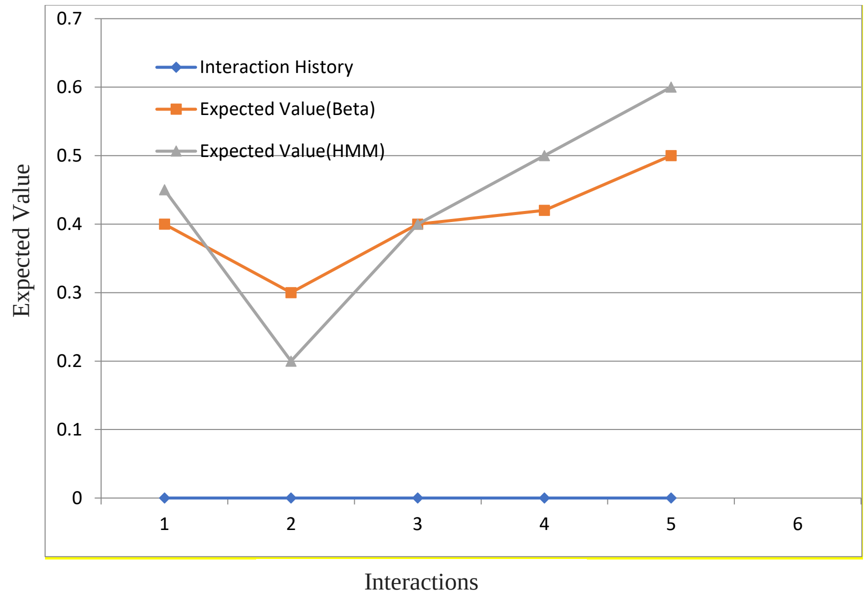

Figure 5.7: The HMM Model & Proposed Model With Different Observation Probabilities.

5.4 Chapter Summary

This chapter has given details of the proposed expert rank algorithm along with details of the basis of the algorithm i.e. the beta probability density function. Each module of the algorithm is explained that includes categorization of interactions, calculation of positive and negative interaction, and finally generation of rank values. The experiment and related datasets are also discussed. The algorithm is tested through interaction categorization, ranking match, and dynamic behavior test. The results are compared against PageRank inspired expert ranking, normal distribution based models, and HMM models. The results show that the proposal is able to differentiate between negative or positive interactions, unlike page rank where negative referrals cannot be handled. The results also show that the ranking closely reflects the opinion of users. The results are independent of the density of data and can adjust rank according to the time factor.

Chapter 6

Attack And Defense Mechanisms

Attacks on reputation systems are targeted towards the goal of the system. The goal of reputation systems is the ability to accurately represent the ranking of participants according to the interaction of the participant. However, if participants launch certain attacks, the system does not produce accurate rankings that promote service provisioning to dishonest participants while bares away the honest participants. The attack model is carried out by insiders who are authenticated to the system. The attacks are classified as follows.

6.1 Self Promotion Or Sybil Attack

The user can augment his/ her reputation with false information. This attack is coupled with the Sybil attack. Systems that consider only positive feedbacks are more prone to this attack. The attacker fabricates false-positive interaction about itself. Self Promotion is achieved through a Sybil attack whereby a user can start creating fake accounts in order to give positive feedbacks to his own account for increasing reputation [19]. The term Sybil emerged from a book about a woman who had a dissociative identity problem. A Sybil account can create malicious accounts to subvert the reputation or ranking of a legal user, these days internet bullying, etc all fall into this type of attack. Recently these attacks have been under discussion in the context of social media specifically twitter [117]. The authors of this work have utilized a regression model to predict the user's profile. Recently researchers utilized game theoretic model [118] to handle the Sybil attacks specifically addressing the issue of trust value updation in a dynamic context.

6.1.1 White Washing

This attack is also called a self-service attack. In this attack, the malicious node starts its original behavior after gaining a good reputation initially. The systems that rely on long historical data are more prone to this type of attack. The malicious behavior can be of a Sybil or slandering. Systems like Sybil Guard [119], SumUp [120], Sybil Limit [121] have proposed techniques to reduce the number of attacks. However, they are not specifically designed for an expert ranking scenario. SybilGuard is based on the "social network" among user identities, where an edge between two identities indicates a human-established trust relationship. Malicious users can create many identities but few trust relationships. One research towards fighting against Sybil attacks in expert ranking systems MHITS [122] utilizes SumUp algorithm, in this system, the nodes are removed through the SumUp strategy before the ranking process. SumUp is a Sybil resilient online content rating system that uses the trust network among users to defend against Sybil attacks. It uses the concept of max-flow. Another stream of research in the field of defense mechanism is related to getting truthful feedback from the users is the peer prediction method [123, 124]. They provide a proper reward system to agents who provide truthful reports of other agents in a nash equilibrium manner. These systems have theoretically tried to address the Sybil attacks by monitoring the agent behavior related to incurring cost by giving opinions.

6.1.2 Slandering

Malicious users could give negative feedback for the users who are positive, thereby affecting the reputation of deserving users. The effect of a single slandering node is less, however, it can have an impact when nodes collude to damage the positive reputation of a node. A typical defense mechanism has a penalty mechanism once a slandering node is identified. Attaching a node with authentic transaction/interaction can also act as a preventive measure.

6.2 Defense Algorithms

The following section gives details of the proposed defense algorithms against the attacks already discussed.

The proposed scheme based upon the Bayesian reputation system is able to prevent the Sybil attack, by evaluating the reputation value of the raters. The scheme ranks every user in the network according to the reputation value. The reputation value is calculated by the feedback given by all the rest of the members. Thus if a Sybil attack is launched with fake ids, not all of the users of the network would be given good feedback about them, or in other words, they might encounter isolation. This holds true for feedback in the form of opinions or through text.

The whitewash attack is countered by having a time factor with the reputation value, thus older feedback is given less importance than the recent feedback values. In the case of a slandering attack, if the user is giving false negative feedbacks, after a certain number of interactions, this reflects that the feedback is malicious. Since a true user will not have any further interaction after negative interactions. Such users can also be filtered out thus, preventing the slandering attack. The attack resistant reputation algorithms are given below.

6.2.1 Slandering Attack Defense Algorithm

In the case of a slandering attack if the user is giving false negative feedback, after a certain amount of interaction this reflects that the feedback is malicious. Since a true user will not have any further interaction, after the negative interactions. Such users can also be filtered out thus preventing the slandering attack. Thus the attack resistant reputation algorithm 3 is given below.

Algorithm 3 Slandering Attack Defense

loop
    if $interaction == negative$ **then**
        $N++$
    else
        $P++$
    end if
end loop
loop
    $Let X \leftarrow i$
    if $X.N > 10$ **then**
        Let s represent slandering node thus $s \leftarrow X$
        Filter out s
    end if
end loop

6.2.2 Sybil Attack Defense Algorithm

The scheme ranks every user in the network according to the reputation value. The reputation value is calculated by the feedback given by all the rest of the members. Thus if a Sybil attack is launched with fake ids, not all of the users of the network would give a good feedback about them, or in other words, they might encounter isolation. This also holds true for feedback in the form of opinions or through text. The proposed scheme weighs the reputation of the node according to the reputation of the interacting node, lines 10–11 of algorithm 4.

Algorithm 4 Sybil Attack Defense

1: Load Dataset
2: **loop**
3: Compute no. of positive interactions p for node i
4: Let z represent node having positive interactions with node i
5: Compute no. of positive and negative interactions of node z
6: Compute Expected value for node z as
7: $Ev(z) \leftarrow p + 1/p + n + 2$
8: Compute no. of negative interactions n for node i
9: Compute Expected value for node i as
10: $Ev(i) \leftarrow p + 1/p + n + 2$
11: $Ev(i) \leftarrow Ev(i) * Ev(z)$
12: **end loop**

6.2.3 White Wash Attack Defense Algorithm

The whitewash attack is countered by having a time factor with the reputation value, thus older feedback is given less importance than the recent feedback values. A related concept has also been discussed in the dishonesty detector based on historical information [87].

Algorithm 5 White Wash Attack Defense

Find time t node i interacting node n-i
if $t = 0$ **then**
Latest interaction only node i,n-i
else
All interactions node i,n-i
end if

6.3 Results And Analysis

The experiments are conducted to evaluate the defense capability of the proposed mechanism. The experiments are performed in three scenarios. The results are shown in terms of the performance metrics appropriate for the domain.

6.3.1 Experiment setup

The experiments are performed on three different scenarios on the data set. Dataset utilized in scenario1 is taken from [111] freely available pool of datasets that are related to intra-organizational interactions The second dataset lists the interactions of 40 nodes. The interactions are rated on a scale of 1-5, where 1 is the lowest and 5 being the highest. These ratings are the interaction feedback values after utilizing the service. This data set is developed as part of a study carried out with the students of the University of Engineering and Technology Peshawar. The study was unbiased carried out by a third party. In order to evaluate the performance of the defense mechanism, the author utilized precision, recall, FMeasure as the performance evaluation metrics.

6.3.2 First Scenario

We have performed a simulation to test the defense mechanism against the Sybil attack. For this, we introduced some new nodes(fake nodes) in the dataset [111]. These fake nodes were involved in giving positive feedback for node 8. Reputation rank calculated by the proposed algorithm without defense mechanism gives it a high rank. In order to verify the defense mechanism, the reputation rank of a fake node is found. Since the fake nodes did not interact with other nodes thus, the resulting rank was low. The new rank of node 8 is then calculated by weighing it by the reputation of the fake nodes and real nodes as is discussed in equation 2. The blue colored series shows node 8 rank due to fake nodes, while the red colored series shows node 8 after utilizing the defense mechanism in the proposed algorithm. This clearly shows that the node loses its fake reputation due to the proposed defense mechanism. To check the functionality against the slandering attack, we manipulated the interactions of node 8 towards a specific node 1. This is done by inducing more negative interactions. As a threshold, if the number of negative interactions exceeds 10, the node 8 is blacklisted and its interactions towards node 1 and other nodes are not recorded.

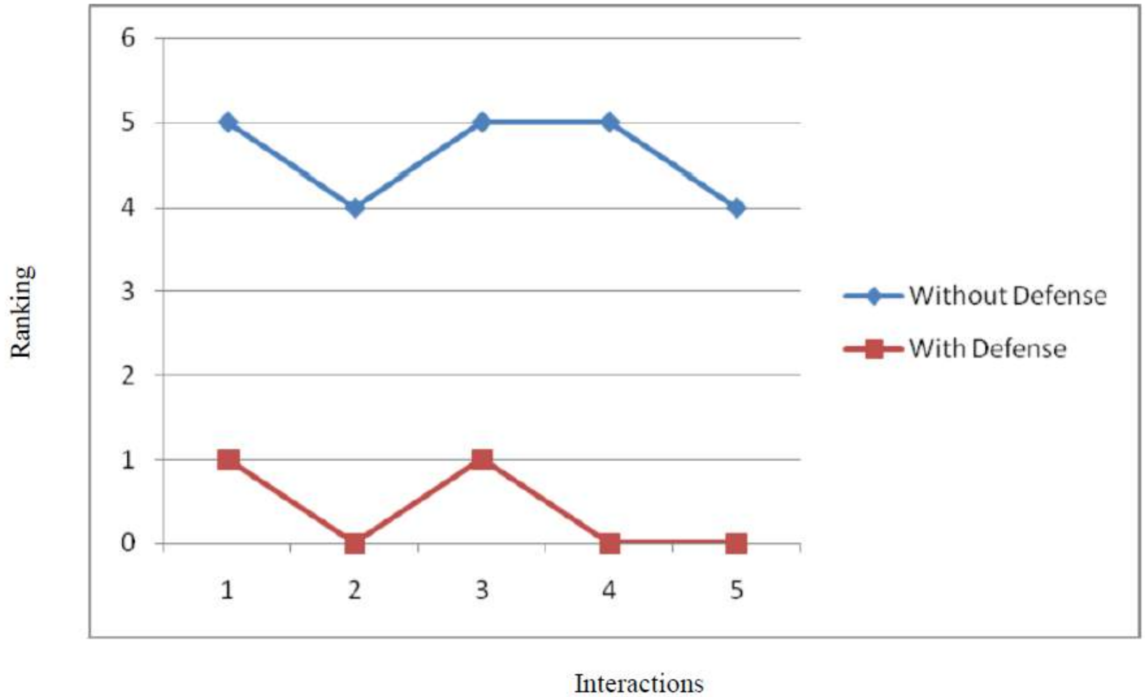

Figure 6.1: Reputation Rank In Case Of Sybil Attack

6.3.3 Second Scenario

A rank list is generated from the data set with the basic reputation rank algorithm. In order to prevent the Sybil attack, a rank list is generated according to the Sybil algorithm as discussed above. The top three nodes are node 17, 2, 37. In order to verify if Sybil can be prevented the data set is manipulated and new nodes are introduced with the intention to raise the reputation rank of a particular node i.e. node 13. The results, however, show that 13 node does not attain any significant advantage, thus the attack is effectively prevented. In this particular case, the newly introduced nodes had no interaction with other nodes so their rank is zero. However, in another case, the reputation rank of these newly introduced nodes is fabricated so as to find the implications of the attack. In such a case, the rank of node 13 got improved from the previous case. Launching the same attack on other content credibility systems i.e PageRank the rank of target nodes gets incremented since more connections are now developed. To the author's knowledge, this issue is not addressed in any of the versions of the PageRank. Similarly, NDR based system could not prevent it since more feedback is now added thereby increasing the overall average of the reputation value of the target node. Another baseline [7] that is based upon summation reputation calculation, whereby the number of followers, likes are treated as input to calculate the reputation rank easily

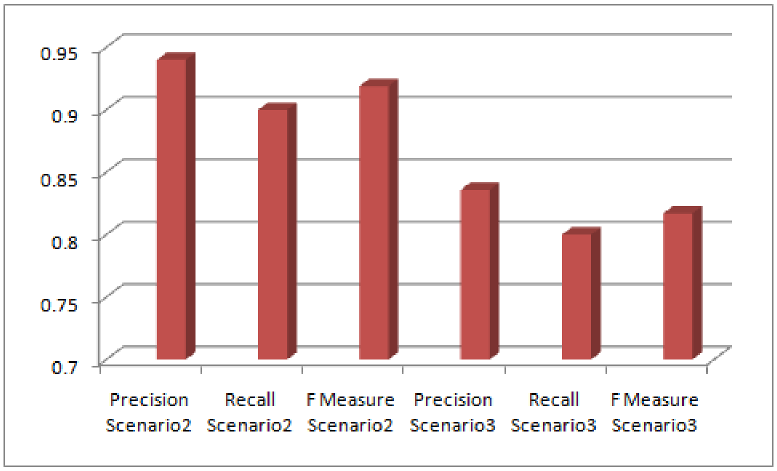

Figure 6.2: Comparative Analysis

suffer this attack.

After clear observation, it was found that node13 gained reputation and its rank improved to 4th place from 7th place in the ranking. Although the Sybils had no prior interactions, thus their impact should have been negligible but the nodes that were rated by node 13 also experience change in their reputation values thereby changing their ranks. Thus in the final sorting node, 13 may have got certain improvement in rank. The authors calculated the Precision that measures the number of nodes that gained a new improved rank. Thus in this scenario Precision is 0.94, while Recall is 0.90.

6.3.4 Third Scenario

Scenario3 is of the Slandering attack, similar to Sybil but here instead of positive feedback, the attackers intend to decrease the reputation rank of the target node. Thus in the above setup instead of newly introduced nodes giving positive feedback, they instead give poor feedback so as to malign the target node. The analysis of the results reveals that ranking after an attack is quite similar to the ranking after the Sybil attack. However few nodes lost their rank and went down in the rank list. The target node 13 didn't lose its rank, the reason being the nodes that are referred by node 13 also face change in rank, thereby in overall calculation node 13 still retains its position, however, its reputation value drops considerably. the Precision calculation of this scenario yields a value of 0.86 and recall is also 0.8 approx.

6.3.5 Analysis

Three different scenarios are discussed in this article. The results reveal that performance of the defense module in the case of the Sybil attack is better than in the case of the Slandering attack. That shows that the nodes gaining ranks are monitored in a better manner. It is to be noted that the defense module is preventive in nature. The precision values verify this in both scenario2 and scenario3. The results also reveal that datasets that are dense in nature can have a behavior similar to above whereby decreasing or increasing the rank of a node can have a ripple effect on many others. While in the case of sparse datasets where the connectivity is loose, prevention against these attacks can have promising outcomes. Since in scenario2 although the attacking nodes had no reputation rank they still succeeded to raise rank of target node 13 due to the density of connections. Thus we can predict that attackers could gain an advantage if the attacking nodes have already gained a reputation. The comparative figure 6.2 shows the Precision, Recall, and FMeasure in the case of scenario2 and scenario3. The FMeasure clearly shows that the proposed mechanism is good enough in prevention against these attacks. The existing web content credibility systems utilizing reputation [7] do not have a defense module to prevent such attacks, which can affect the credibility evaluation of the system. Thus comparative analysis to these can not be reported. However, it is pertinent to mention that certain mechanisms have been proposed to detect the Sybil attack in other application domains.

6.4 Chapter Summary

This chapter has discussed various attacks on the reputation systems concerning web content credibility and expert rank systems. The attack model is briefly discussed. The proposed defense algorithm is discussed and its efficacy is presented through different scenarios. The algorithm weighs the reputation rank of a node by the reputation rank of its raters to prevent the Sybil attack, limits the number of negative interactions to ward off the Slander attack, and controls the history of interactions to prevent the Whitewash attack. The scenarios of Sybil, Slander, and Whitewash attacks are evaluated against the proposed mechanism. The results show that the proposed mechanism has the capability of preventing these attacks.

Chapter 7

Conclusion And Future Work

Trust Judgment of web information is a new field approached in different ways by researchers. Some researchers have placed more emphasis on, site design and information representation as a factor towards the declaration of web information being trustworthy, while others look at this problem from cognitive theories. This thesis has laid down a formal relationship of trust to credibility. It was also observed from the literature that trust is achieved through credibility. The problems addressed in this thesis are concerned with the credibility of web content. Web content exists in various forms such as information on web pages, microblogs, email communication. Since information on the web is shared by people from different expertise levels so the thesis also focused on addressing problems related to existing expert ranking techniques. The proposed approach is based upon a reputation model, thus various reputation attacks in the context of web content credibility and expert ranking are also handled. An architecture for content credibility is proposed. The architecture consists of algorithms that relate an expert with the credibility of the content thereby offering the following advantages including;

- Inclusion of negative and passive along with positive and active interactions in the calculation of the reputation rank.
- Negative referral and quality of interactions in expert ranks.
- An expert rank technique specifically for large organizations when the expertise of employee is tacit and there is no documentary proof about it.
- Preventive capability of the defense algorithms against attacks on reputation based content credibility systems.

- Significant improvement in precision, recall, and correlation compared to baselines.

Key findings of this research in web content credibility, expert ranks, and their performance evaluations are presented in this chapter. In the end, suggestions for future work are presented.

7.1 Summary Of Key Findings

The research objectives are discussed in chapter 1 of the thesis, thus here we have summarised the key findings accordingly. The thesis has proposed a theoretical analysis of the relationship of trust to credibility. Thus it is found that trust is achieved through credibility. Credibility can be checked by finding the source or content of information. The thesis is, however, considering the situation when incorrect information flows from credible sources. Thus in this case content needs to be evaluated by certain applicable techniques to guarantee its credibility. The survey study showed that out of various factors, expert opinion is an important factor that can regard a piece of information as credible. The proposed technique that has a strong basis of the Bayesian algorithm has shown promising results in comparison to the various baselines. The results from the experiments showed that the proposed technique can differentiate types of interaction. The results are independent of the pattern and density of data that limit the performance of the baselines. The results in terms of Precision show 27% and 18% improvement for the two experiments with a significant correlation. The credibility framework has provided a generic technique applicable to various application domains.

The advent of knowledge economies and internet based business markets, has highlighted the need for expert ranking techniques. Most of the researchers have explored these techniques in the field of micro-blogs and few in the field of large organizations and enterprises. However, the authors identified that with the rise in the online enterprises where employees from diverse backgrounds join in, expert ranking is the need for the managers, especially for the cases when documented proof of expert knowledge is not present. Thus this issue is addressed with a newly proposed technique that is based upon reputation information rather than graph or documents based information. The reputation information is gathered through the past interactions of the user. The technique has a statistical basis and utilizes types of interactions to rank the nodes in comparison to previous techniques that either utilized the number of links or documents produced by the user to find its rank. The technique is also capable of solving the negative referral problem and collusion faced by previous techniques. The experimental results and comparison to previous baseline techniques of expert ranking provide enough evidence regarding the better performance of the proposed technique. The performance metric of Precision shows about 7% improvement on average for the experiments conducted on three different datasets. We have also proposed the time factor in the technique, whereby older interactions are given lesser importance than the newer ones. Comparison to baselines demonstrated that the proposed technique is more appropriate to model dynamic environment when the changes are not rapid. It was also revealed from the results that inclusion of both types of interactions in calculating the reputation rank of a node has lesser MAE as compared to the case when only positive interactions are utilized. The analysis brought an insight that the results of baselines are dependent upon nature and rating trend of the datasets. The proposed algorithm however showed independence from these. Detailed graphical figures regarding response of the proposed technique to change in history of information, along with comparison of results among three datasets in terms of MAE and Precision appropriately explain the results.

Finally, the thesis has also discussed reputation attacks. The reputation attacks in the form of Sybil, Slandering, and Whitewash are discussed and their defense algorithms are proposed. These attacks target to falsely increase or decrease the reputation rank of a node. These attacks are not addressed in previous content credibility techniques. The proposed defense mechanism is demonstrated in three different scenarios thereby showing attack preventive capability of the mechanism.

The results from different experiments are shown in detail in the form of figures and tables. The above discussion thus, leads us to conclude that reputation based approach with the basis of the Bayesian calculation structure provides better credibility analysis of the content.

7.2 Contributions

The contribution of the research is the web content credibility framework that is reputation based and can be implemented with minimum overhead. The proposed technique can differentiate between types of interactions, as positive, negative, active, and passive. The third contribution is the proposed technique that can overcome the shortcomings of previous expert ranking techniques that are mostly link-based and employed the Page Rank Algorithm. The shortcomings include the ability to handle negative referrals, and avoidance of collusion; this includes its application to a special scenario of large organizations. Another important contribution of the thesis is the proposal of a defense mechanism to fight against various threats to reputation based systems. These contributions are targeted towards a framework that can assess the trustworthiness of web content. Generally, we can highlight the contributions as

- Trust judgment framework for Web Information, that can guide the users to assess the credibility of the information present on the web in the form of social networks, websites.

- Content credibility framework based on reputation information with the ability to take into account the positive, negative, active, passive behavior.
- An expert rank technique based on the quality of interactions.
- A domain-specific expert rank technique for large enterprise, to find the expertise of the employees with tacit expertise, or undocumented expertise.
- Defense Mechanism against Threats to Reputation Systems in the form of Sybil, Slandering, and Whitewash attacks in the context of web credibility systems.
- A dataset is created from the results of the survey study. The dataset shows interactions among nodes in a friendship relationship, where the edges are ranked according to expert knowledge of each other.

7.3 Suggestions For Future Work

The proposed framework has a wide range of applications to the domains of healthcare, finances, education, economy. One of the suggested areas is the application of the proposed framework to healthcare content present on the web. However, other future areas can also be explored. Some suggestions are given to provide a direction.

First, there is a need to develop a real time application that includes expert opinion for recommending services and products.

Secondly in future work, research on other features that could be utilized for interaction categorizations needs to be carried out. In this thesis, we have discussed direct ranking or sentiment based categorizations. Other feature sets might include text length, the presence of hyperlinks and time to name a few. The threshold for categorization against these features is also an area of research.

Lastly but an open area that can be explored is the countermeasures against the attacks on the reputation model in terms of detection. We also aim to further verify the proposed defense mechanism that is preventive, by its application to a real scenario.

Appendix A

Survey Study Questionnaire

Note: This questionnaire is for research purpose only, all data will be kept confidential. Name:

Q.1 What are the things you take into account while considering the credibility of answers in a blog?

- User who answers most frequently
- User who posts lengthy answers
- Users who post some reference link.
- Users who answer immediately
- All of above

In case of all of above rank in order of importance.

Q.2 Rank following in order of their importance for evaluating the reliability of information present on web.(put 1 for most important , then 2, 3 and 4 for least important)

- Source of information
- Popularity of information
- Expert analysis of information
- Recommendation by others.

Q.3 Given a rating scale 0-6(0 being lowest rating, 6 being highest), which value would you consider as cutoff such that values above it are considered positive and values below the cutoff are considered negative.

Q.4 List 3 of your friends (from this class only) along with their subject knowledge on scale of 1-5(1: nil, 2: fair, 3: good, 4: very good, 5: excellent)

- 1:nil
- 2:fair
- 3:good
- 4:very good
- 5:excellent

Bibliography

[1] Y. Kato, S. Kurohashi, and K. Inui, "Classifying information sender of web documents," *Internet Research*, vol. 18, no. 2, pp. 191–203, 2008.

[2] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[3] J. Huang, "Knowledge provenance: An approach to modeling and maintaining the evolution and validity of knowledge," Ph.D. dissertation, 2008.

[4] Y. Gil and D. Artz, "Towards content trust of web resources," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 5, no. 4, pp. 227–239, 2007.

[5] T. Lappas, K. Liu, and E. Terzi, "A survey of algorithms and systems for expert location in social networks," in *Social network data analytics*. Springer, 2011, pp. 215–241.

[6] B. E. Commerce, A. Jøsang, and R. Ismail, "The beta reputation system," in *In proceedings of the 15th bled electronic commerce conference*. Citeseer, 2002.

[7] M. Alrubaian, M. Al-Qurishi, M. Al-Rakhami, M. M. Hassan, and A. Alamri, "Reputation-based credibility analysis of twitter social network users," *Concurrency and Computation: Practice and Experience*, vol. 29, no. 7, p. e3873, 2017.

[8] E. H. Jung, K. Walsh-Childers, and H.-S. Kim, "Factors influencing the perceived credibility of diet-nutrition information web sites," *Computers in Human Behavior*, vol. 58, pp. 37–47, 2016.

[9] A. A. Shah, S. D. Ravana, S. Hamid, and M. A. Ismail, "Web credibility assessment: affecting factors and assessment techniques," *Information research*, vol. 20, no. 1, pp. 20–1, 2015.

[10] S. Ghosh, N. Sharma, F. Benevenuto, N. Ganguly, and K. Gummadi, "Cognos: crowdsourcing search for topic experts in microblogs," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2012, pp. 575–590.

[11] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 261–270.

[12] G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, "Expertrank: A topic-aware expert finding algorithm for online knowledge communities," *Decision Support Systems*, vol. 54, no. 3, pp. 1442–1451, 2013.

[13] L. Zhang, X.-Y. Li, J. Lei, J. Sun, and Y. Liu, "Mechanism design for finding experts using locally constructed social referral web," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 8, pp. 2316–2326, 2014.

[14] D. P. David, M. M. Keupp, and A. Mermoud, "Knowledge absorption for cyber-security: The role of human beliefs," *Computers in Human Behavior*, vol. 106, p. 106255, 2020.

[15] R. Borges, M. Bernardi, and R. Petrin, "Cross-country findings on tacit knowledge sharing: evidence from the brazilian and indonesian it workers," *Journal of Knowledge Management*, 2019.

[16] M. Kokkodis, "Reputation deflation through dynamic expertise assessment in online labor markets," in *The World Wide Web Conference*, 2019, pp. 896–905.

[17] M. Daltayanni, L. de Alfaro, and P. Papadimitriou, "Workerrank: Using employer implicit judgements to infer worker reputation," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 263–272.

[18] X. Liu, R. Nielek, A. Wierzbicki, and K. Aberer, "Defending imitating attacks in web credibility evaluation systems," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 1115–1122.

[19] K. Hoffman, D. Zage, and C. Nita-Rotaru, "A survey of attack and defense techniques for reputation systems," *ACM Computing Surveys (CSUR)*, vol. 42, no. 1, p. 1, 2009.

[20] M. S. Ackerman and T. W. Malone, *Answer Garden: A tool for growing organizational memory*. ACM, 1990, vol. 11, no. 2-3.

[21] B. Krulwich, C. Burkey, and A. Consulting, “The contactfinder agent: Answering bulletin board questions with referrals,” in *AAAI/IAAI, Vol. 1*, 1996, pp. 10–15.

[22] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, “Wtf: The who to follow service at twitter,” in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 505–514.

[23] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, “Tweetcred: Real-time credibility assessment of content on twitter,” in *International Conference on Social Informatics*. Springer, 2014, pp. 228–243.

[24] G. Ferreira, A. C. Traeger, G. Machado, M. O’Keeffe, and C. G. Maher, “Credibility, accuracy, and comprehensiveness of internet-based information about low back pain: A systematic review,” *Journal of medical Internet research*, vol. 21, no. 5, p. e13357, 2019.

[25] A. Nabozny, B. Balcerzak, and A. Wierzbicki, “Automatic credibility assessment of popular medical articles available online,” in *International Conference on Social Informatics*. Springer, 2018, pp. 215–223.

[26] T. Grandison and M. Sloman, “A survey of trust in internet applications,” *IEEE Communications Surveys & Tutorials*, vol. 3, no. 4, pp. 2–16, 2000.

[27] A. Abdul-Rahman and S. Hailes, “A distributed trust model,” in *Proceedings of the 1997 workshop on New security paradigms*. ACM, 1998, pp. 48–60.

[28] Abdul-Rahman, Alfarez and Hailes, Stephen, “Using recommendations for managing trust in distributed systems,” in *Proceedings IEEE Malaysia International Conference on Communication*, vol. 97, 1997.

[29] M. Carbone, M. Nielsen, and V. Sassone, “A formal model for trust in dynamic networks,” in *First International Conference onSoftware Engineering and Formal Methods, 2003. Proceedings.* IEEE, 2003, pp. 54–61.

[30] P. Resnick and R. Zeckhauser, “Trust among strangers in internet transactions: Empirical analysis of ebay’s reputation system.” *The Economics of the Internet and E-commerce*, vol. 11, no. 2, pp. 23–25, 2002.

[31] J. Sabater and C. Sierra, “Regret: reputation in gregarious societies,” in *Agents*, vol. 1, 2001, pp. 194–195.

[32] W. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, “Travos: Trust and reputation in the

context of inaccurate information sources," *Autonomous Agents and Multi-Agent Systems*, vol. 12, no. 2, pp. 183–198, 2006.

[33] T. Dong-Huynha, N. Jennings, and N. Shadbolt, "Fire: An integrated trust and reputation model for open multi-agent systems," in *ECAI 2004: 16th European Conference on Artificial Intelligence, August 22-27, 2004, Valencia, Spain: including Prestigious Applicants [sic] of Intelligent Systems (PAIS 2004): proceedings*, vol. 110, 2004, p. 18.

[34] F. C. Mish, *Merriam-Webster's collegiate dictionary*. Merriam-Webster, 2004.

[35] B. Fogg and H. Tseng, "The elements of computer credibility," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1999, pp. 80–87.

[36] A. Josang and W. Quattrociocchi, "Advanced features in bayesian reputation systems," in *International Conference on Trust, Privacy and Security in Digital Business*. Springer, 2009, pp. 105–114.

[37] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." 1999.

[38] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 221–230.

[39] C. L. Clarke, G. V. Cormack, and T. R. Lynam, "Exploiting redundancy in question answering," in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2001, pp. 358–365.

[40] T. M. Ciolek, "The six quests for the electronic grail: Current approaches to information quality in www resources," *Revue Informatique et Statistique dans les Sciences Humaines*, vol. 32, no. 1-4, p. 176, 1996.

[41] L. Ding, P. Kolari, T. Finin, A. Joshi, Y. Peng, Y. Yesha *et al.*, "On homeland security and the semantic web: A provenance and trust aware inference framework," in *Proceedings of the AAAI Spring Symposium on AI Technologies for Homeland Security*, 2005.

[42] A. Abdel-Hafez, Y. Xu, and A. Jøsang, "A normal-distribution based reputation model," in *International Conference on Trust, Privacy and Security in Digital Business*. Springer, 2014, pp. 144–155.

[43] J. Golbeck, "Combining provenance with trust in social networks for semantic web content

filtering," in *International Provenance and Annotation Workshop*. Springer, 2006, pp. 101–108.

[44] J. Golbeck, J. Hendler *et al.*, "Filmtrust: Movie recommendations using trust in web-based social networks," in *Proceedings of the IEEE Consumer communications and networking conference*, vol. 96, no. 1. Citeseer, 2006, pp. 282–286.

[45] S. Nepal, W. Sherchan, and C. Paris, "Strust: A trust model for social networks," in *2011ieee 10th international conference on trust, security and privacy in computing and communications*. IEEE, 2011, pp. 841–846.

[46] J. Huang and M. S. Fox, "Trust judgment in knowledge provenance," in *16th International Workshop on Database and Expert Systems Applications (DEXA'05)*. IEEE, 2005, pp. 524–528.

[47] E. Chang, T. S. Dillon, and F. K. Hussain, "Trust and reputation relationships in service-oriented environments," in *Third International Conference on Information Technology and Applications (ICITA'05)*, vol. 1. IEEE, 2005, pp. 4–14.

[48] P. Avesani, P. Massa, and R. Tiella, "A trust-enhanced recommender system application: Moleskiing," in *Proceedings of the 2005 ACM symposium on Applied computing*. ACM, 2005, pp. 1589–1593.

[49] R. Levien, "Advogato trust metric," 2003.

[50] C.-N. Ziegler and G. Lausen, "Spreading activation models for trust propagation," in *IEEE International Conference on e-Technology, eCommerce and eService, 2004. EEE'04. 2004*. IEEE, 2004, pp. 83–97.

[51] Y. Gil and V. Ratnakar, "Trellis: An interactive tool for capturing information analysis and decision making," in *International Conference on Knowledge Engineering and Knowledge Management*. Springer, 2002, pp. 37–42.

[52] H. Berghel, "Lies, damn lies, and fake news," *Computer*, vol. 50, no. 2, pp. 80–85, 2017.

[53] emergentinfo.tumblr. (2014) emergentinfo.tumblr. [Online]. Available: emergentinfo.tumblr.com.

[54] T. A. Schwandt, Y. S. Lincoln, and E. G. Guba, "Judging interpretations: But is it rigorous? trustworthiness and authenticity in naturalistic evaluation," *New directions for evaluation*, vol. 2007, no. 114, pp. 11–25, 2007.

[55] M. Alrubaian, M. Al-Qurishi, A. Alamri, M. Al-Rakhami, M. M. Hassan, and G. Fortino, "Credibility in online social networks: A survey," *IEEE Access*, vol. 7, pp. 2828–2855, 2018.

[56] M. J. Metzger, "Making sense of credibility on the web: Models for evaluating online information and recommendations for future research," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 13, pp. 2078–2091, 2007.

[57] M. J. Metzger and A. J. Flanagin, "Credibility and trust of information in online environments: The use of cognitive heuristics," *Journal of Pragmatics*, vol. 59, pp. 210–220, 2013.

[58] R. Savolainen, "Judging the quality and credibility of information in internet discussion forums," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 7, pp. 1243–1256, 2011.

[59] N. Asokan, V. Shoup, and M. Waidner, "Optimistic fair exchange of digital signatures," *IEEE Journal on Selected Areas in communications*, vol. 18, no. 4, pp. 593–610, 2000.

[60] K. Cho and C. D. Schunn, "Scaffolded writing and rewriting in the discipline: A web-based reciprocal peer review system," *Computers & Education*, vol. 48, no. 3, pp. 409–426, 2007.

[61] A. Olteanu, S. Peshterliev, X. Liu, and K. Aberer, "Web credibility: Features exploration and credibility prediction," in *European conference on information retrieval*. Springer, 2013, pp. 557–568.

[62] M. Kakol, R. Nielek, and A. Wierzbicki, "Understanding and predicting web content credibility using the content credibility corpus," *Information Processing & Management*, vol. 53, no. 5, pp. 1043–1061, 2017.

[63] R. El Ballouli, W. El-Hajj, A. Ghandour, S. Elbassuoni, H. Hajj, and K. Shaban, "Cat: Credibility analysis of arabic content on twitter," in *Proceedings of the Third Arabic Natural Language Processing Workshop*, 2017, pp. 62–71.

[64] N. Y. Hassan, W. H. Gomaa, G. A. Khoriba, and M. H. Haggag, "Supervised learning approach for twitter credibility detection," in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. IEEE, 2018, pp. 196–201.

[65] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 675–684.

[66] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 729–736.

[67] J. Fairbanks, N. Fitch, N. Knauf, and E. Briscoe, "Credibility assessment in the news: Do we need to read," in *Proc. of the MIS2 Workshop held in conjuction with 11th Int'l Conf. on Web Search and Data Mining*, 2018, pp. 799–800.

[68] Y. Gao, X. Li, J. Li, Y. Gao, and S. Y. Philip, "Info-trust: A multi-criteria and adaptive trustworthiness calculation mechanism for information sources," *IEEE Access*, vol. 7, pp. 13 999–14 012, 2019.

[69] C.-N. Ziegler, "Semantic web recommender systems," in *International Conference on Extending Database Technology*. Springer, 2004, pp. 78–89.

[70] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, "Dublin core metadata for resource discovery," *Internet Engineering Task Force RFC*, vol. 2413, no. 222, p. 132, 1998.

[71] A. V. Pantola, S. Pancho-Festin, and F. Salvador, "Rating the raters: a reputation system for wiki-like domains," in *Proceedings of the 3rd international conference on Security of information and networks*. ACM, 2010, pp. 71–80.

[72] K. Aberer and Z. Despotovic, "Managing trust in a peer-2-peer information system," in *Proceedings of the tenth international conference on Information and knowledge management*. ACM, 2001, pp. 310–317.

[73] F. Cornelli, E. Damiani, S. D. C. Di Vimercati, S. Paraboschi, and P. Samarati, "Choosing reputable servents in a p2p network," in *Proceedings of the 11th international conference on World Wide Web*. ACM, 2002, pp. 376–386.

[74] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 417–424.

[75] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt, "Ratings in distributed systems: A bayesian approach," in *Proceedings of the Workshop on Information Technologies and Systems (WITS)*, 2001, pp. 1–7.

[76] C. Lodigiani and M. Melchiori, "A pagerank-based reputation model for vgi data," *Procedia Computer Science*, vol. 98, pp. 566–571, 2016.

[77] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in p2p networks," in *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003, pp. 640–651.

[78] M. Ravi, "Trust and uncertainty in distributed environments: application to the management of data and data sources quality in m2m (machine to machine) systems." Ph.D. dissertation, Université Grenoble Alpes, 2016.

[79] J. Wang, M. D. Molina, and S. S. Sundar, "When expert recommendation contradicts peer opinion: Relative social influence of valence, group identity and artificial intelligence," *Computers in Human Behavior*, vol. 107, p. 106278, 2020.

[80] T. Amjad, A. Daud, and N. R. Aljohani, "Ranking authors in academic social networks: a survey," *Library Hi Tech*, 2018.

[81] A. Al-Barakati and A. Daud, "Venue-influence language models for expert finding in bibliometric networks," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 14, no. 3, pp. 184–201, 2018.

[82] A. Daud, M. Song, M. K. Hayat, T. Amjad, R. A. Abbasi, H. Dawood, A. Ghani *et al.*, "Finding rising stars in bibliometric networks," *Scientometrics*, pp. 1–29, 2020.

[83] K. Wu, Z. Noorian, J. Vassileva, and I. Adaji, "How buyers perceive the credibility of advisors in online marketplace: review balance, review count and misattribution," *Journal of Trust Management*, vol. 2, no. 1, p. 2, 2015.

[84] E. Yan, Y. Ding, and C. R. Sugimoto, "P-rank: An indicator measuring prestige in heterogeneous scholarly networks," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 3, pp. 467–477, 2011.

[85] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[86] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 45–54.

[87] G. A. Wang, J. Jiao, A. S. Abrahams, W. Fan, and Z. Zhang, "Expertrank: A topic-aware expert finding algorithm for online knowledge communities," *Decision Support Systems*, vol. 54, no. 3, pp. 1442–1451, 2013.

[88] R. Divya, C. A. Kumar, S. Saijanani, and M. Priyadharshini, "Deceiving communication links on an organization email corpus," *Malaysian Journal of Computer Science*, vol. 24, no. 1, pp. 17–33, 2017.

[89] T. Hecking, A. Harrer, and H. U. Hoppe, "Discovery of structural and temporal patterns in mooc discussion forums," in *Prediction and inference from social networks and social media*. Springer, 2017, pp. 171–198.

[90] M. Anandarajan, C. Hill, and T. Nolan, "Probabilistic topic models," in *Practical Text Analytics*. Springer, 2019, pp. 117–130.

[91] B. Xu, H. Lin, Y. Lin, and Y. Guan, "Integrating social annotations into topic models for personalized document retrieval," *Soft Computing*, vol. 24, no. 3, pp. 1707–1716, 2020.

[92] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom, "Expertise identification using email communications," in *Proceedings of the twelfth international conference on Information and knowledge management*. ACM, 2003, pp. 528–531.

[93] Y. Fu, R. Xiang, Y. Liu, M. Zhang, and S. Ma, "Finding experts using social network analysis," in *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE Computer Society, 2007, pp. 77–80.

[94] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007, pp. 221–230.

[95] J. Wang, J. Sun, H. Lin, H. Dong, and S. Zhang, "Convolutional neural networks for expert recommendation in community question answering," *Science China Information Sciences*, vol. 60, no. 11, p. 110102, 2017.

[96] M. Faisal, A. Daud, and A. Akram, "Expert ranking using reputation and answer quality of co-existing users." *International Arab Journal of Information Technology (IAJIT)*, vol. 14, no. 1, 2017.

[97] A. Daud, J. Li, L. Zhou, and F. Muhammad, "Temporal expert finding through generalized time topic modeling," *Knowledge-Based Systems*, vol. 23, no. 6, pp. 615–625, 2010.

[98] Z. Liu and B. J. Jansen, "Identifying and predicting the desire to help in social question and answering," *Information Processing & Management*, vol. 53, no. 2, pp. 490–504, 2017.

[99] Liu, Zhe and Jansen, Bernard J, "Questioner or question: Predicting the response rate in

social question and answering on sina weibo," *Information Processing & Management*, vol. 54, no. 2, pp. 159–174, 2018.

[100] H. Yu, B. Zhou, M. Deng, and F. Hu, "Tag recommendation method in folksonomy based on user tagging status," *Journal of Intelligent Information Systems*, vol. 50, no. 3, pp. 479–500, 2018.

[101] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom, "Expertise identification using email communications," in *Proceedings of the twelfth international conference on Information and knowledge management*. ACM, 2003, pp. 528–531.

[102] B. Dom, I. Eiron, A. Cozzi, and Y. Zhang, "Graph-based ranking algorithms for e-mail expertise analysis," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, pp. 42–48.

[103] M. S. Faisal, A. Daud, A. U. Akram, R. A. Abbasi, N. R. Aljohani, and I. Mehmood, "Expert ranking techniques for online rated forums," *Computers in Human Behavior*, vol. 100, pp. 168–176, 2019.

[104] F. Calefato, F. Lanubile, M. C. Marasciulo, and N. Novielli, "Mining successful answers in stack overflow," in *Proceedings of the 12th Working Conference on Mining Software Repositories*. IEEE Press, 2015, pp. 430–433.

[105] G. Casella and R. L. Berger, "Statistical inference vol. 70," 1990.

[106] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision support systems*, vol. 43, no. 2, pp. 618–644, 2007.

[107] M. Faisal, A. Daud, and A. Akram, "Expert ranking using reputation and answer quality of co-existing users." *International Arab Journal of Information Technology (IAJIT)*, vol. 14, no. 1, 2017.

[108] T. Amjad, A. Daud, A. Akram, and F. Muhammed, "Impact of mutual influence while ranking authors in a co-authorship network," *Kuwait Journal of Science*, vol. 43, no. 3, 2016.

[109] F. Fouss, Y. Achbany, and M. Saerens, "A probabilistic reputation model based on transaction ratings," *Information Sciences*, vol. 180, no. 11, pp. 2095–2123, 2010.

[110] Y. Liu, U. S. Chitawa, G. Guo, X. Wang, Z. Tan, and S. Wang, "A reputation model for aggregating ratings based on beta distribution function," in *Proceedings of the 2nd International Conference on Crowd Science and Engineering*. ACM, 2017, pp. 77–81.

[111] R. L. Cross, R. L. Cross, and A. Parker, *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business Press, 2004.

[112] F. Loll and N. Pinkwart, “Using collaborative filtering algorithms as elearning tools,” in *2009 42nd Hawaii International Conference on System Sciences*. IEEE, 2009, pp. 1–10.

[113] A. Ashari, I. Paryudi, and A. M. Tjoa, “Performance comparison between naive bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool,” *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no. 11, 2013.

[114] V. A. Kharde and S. Sonawane, “Sentiment analysis of twitter data: A survey of techniques,” *International Journal of Computer Applications*, vol. 975, p. 8887.

[115] E. A. Freeman and G. G. Moisen, “A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa,” *Ecological Modelling*, vol. 217, no. 1-2, pp. 48–58, 2008.

[116] M. E. Moe, B. E. Helvik, and S. J. Knapskog, “Comparison of the beta and the hidden markov models of trust in dynamic environments,” in *IFIP International Conference on Trust Management*. Springer, 2009, pp. 283–297.

[117] M. Al-Qurishi, M. Alrubaian, S. M. M. Rahman, A. Alamri, and M. M. Hassan, “A prediction system of sybil attack in social network using deep-regression model,” *Future Generation Computer Systems*, vol. 87, pp. 743–753, 2018.

[118] B. Kumar and B. Bhuyan, “Game theoretical defense mechanism against reputation based sybil attacks,” *Procedia Computer Science*, vol. 167, pp. 2465–2477, 2020.

[119] A. B. Potey and A. B. Raut, “Combating sybil attacks using sybilguard,” *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, no. 2, 2013.

[120] D. N. Tran, B. Min, J. Li, and L. Subramanian, “Sybil-resilient online content voting.” in *NSDI*, vol. 9, no. 1, 2009, pp. 15–28.

[121] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, “Sybillimit: A near-optimal social network defense against sybil attacks,” in *2008 IEEE Symposium on Security and Privacy (sp 2008)*. IEEE, 2008, pp. 3–17.

[122] K. A. Rashed, C. Balasoiu, and R. Klamma, "Robust expert ranking in online communities-fighting sybil attacks," in *8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*. IEEE, 2012, pp. 426–434.

[123] N. Miller, P. Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," *Management Science*, vol. 51, no. 9, pp. 1359–1373, 2005.

[124] A. Dasgupta and A. Ghosh, "Crowdsourced judgement elicitation with endogenous proficiency," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 319–330.