# Familiarity and Trust: Measuring Familiarity with a Web Site

Jie Zhang
Faculty of Computer Science
University of New Brunswick
Fredericton, NB, E3B 5A3, Canada
Email: jie.zhang@unb.ca

Ali. A. Ghorbani
Faculty of Computer Science
University of New Brunswick
Fredericton, NB, E3B 5A3, Canada
Email: ghorbani@unb.ca

***Abstract*— This work aims at measuring familiarity to contribute to the formalization of trust. Trust has always been bundled with familiarity to become a popular topic in the areas of psychology, sociology and computer science. Correlation between familiarity and trust has been explored and proved by many studies from different perspectives. A new model of trust has been proposed by Carter and Ghorbani to formalize the value-centric trust in agent societies. However, the measurement of familiarity in their work is roughly the similarity of values between two agents. Familiarity measurements proposed by other researchers are not convenient due to the instability and abstruseness of familiarity, or are useful only in certain circumstances and are quite problem-specific. We propose a convenient way of measuring familiarity with a Web site and continuously updating its value based on the exploration of factors that may affect familiarity. The five major factors include prior experience, repeated exposure, study duration, level of processing and forgetting rate. The human factors are mapped to the properties of Web application domain through a factors hierarchy. Experiments to evaluate the performance of the proposed measurement are discussed in the future work section.**

## I. Introduction

In the financial field, trust has always been a focus because greater trust is strongly related to better economic outcomes. Trust has always been bundled with familiarity to become a popular topic in the fields of psychology, sociology and computer science. Correlation between familiarity and trust has been explored and proved by many researchers from different perspectives. Through an experimental investigation involving an investment game and an ultimatum game, Barr demonstrated that people in resettled villages trust each other less than people in non-resettled villages due to lack of familiarity [1]. Many other researchers explored the relationship between trust, familiarity and investment. Individuals prefer familiar investments, and fear change and the unfamiliar [2]. This phenomenon shows the effects of familiarity on financial decisions through trust. Huberman summarized much evidence discovered by others, such as the evidence, discovered by Kilka and Weber, that business students are more optimistic about their home countries' stocks than other countries'; the evidence, discovered by Coval and Moskowitz, that U.S. investment managers prefer local companies; and so on [3]. After having listed many instances of investment in the familiar, he analyzed the geographic distribution of the shareholders of a Regional Bell Operating Company (RBOC) and related the amount of individuals' investment in the RBOCs to the typical U.S. household's net worth and stock holdings to offer the explanation of the home country bias: people simply prefer to invest in the familiar.

Web users crave familiarity, as pointed out by Nick Usborne. General Web users would love it if all our sites looked like Amazon.com. Unfamiliar things in a Web site will give a user reason to pause. As a result, the access and transaction rates will drop off. In addition, other conclusions such as that familiarity breeds trust online and familiarity breeds online spending have been mentioned in many financial reports. Customers' trust in an E-seller is considered to be an important factor for E-commerce success and increasing familiarity with the Web site is considered to be a process of building trust online.

Many definitions of trust have been summarized from different perspectives and properties of trust have been explored in the literature review part of Carter and Ghorbani's work [4], [5]. As they pointed out, trust is multidimensional in that it can be facilitated through familiarity. They further clarified the concept of trust by establishing a new model of trust: trust is a combination of self-esteem, reputation and familiarity. The relationship between familiarity and trust is clearer in the model. Trust has been further formalized through a concept graph map. Two major ingredients, reputation and self-esteem, are determined by roles based on the foundation of trust, values. Familiarity is roughly the similarity of values based on the argument that familiarity between two agents is a result of similarity in the underlying value-systems of the two individuals. The familiarity value is then determined by the Hamming distance of agent value hierarchies. We believe that familiarity is more knowledge-based and measurement of familiarity can be more particular between two agents in the Web application domain.

Measuring familiarity is a comparatively hard issue. Many researchers measure familiarity by questionnaires in their works. It is not convenient because the process has to be redone after each time when the familiarity changes and the familiarity changes continuously. Other proposed measurements are useful only in certain circumstances and are quite problem-specific. Examples of them are described in detail in Section

II (Related Work). In our work, we analyze major factors that may affect user familiarity with a Web site and come up with a reliable way to measure it by taking into account all the factors, including prior experience, repeated exposure, level of processing, study duration and forgetting rate.

The rest of the paper is organized as follows. Related work of measuring familiarity is reviewed in Section II. Section III describes in detail all the five major factors affecting familiarity. The way of measuring and updating familiarity is proposed in Section IV. Conclusions and future work will be presented in Section V.

## II. Related Work

Different approaches have been proposed by many researchers to measure familiarity, including distributing questionnaires to customers to measure their familiarity with products, evaluating users' information search behavior (including searching efficacy and reading time) to measure their topic familiarity, and using Web frequency (the number of pages containing a given keyword) to measure cultural familiarity.

Magnus Söderlund measures familiarity by providing open-ended questions to customers [6], [7], [8]. The measurement is based on customers' experience. The first questionnaire is distributed to customers, and the second questionnaire will be distributed to only the ones who return the first questionnaire. The question may be, for example, "how many times did the customer make trips with the airline to domestic and international destinations during the past 12 months?". Presumably, it is a question that invites vagueness. In order to deal with the problem, the zero-order correlation of the two items appearing in both questionnaires is calculated to evaluate the reliability of the two answers. If the answers are reasonable and reliable, the answer from the second questionnaires will be used as the familiarity value. Some researchers [9], [10] have carried out experiments to test the effects of familiarity on object naming. They measured familiarity by a Likert-type scale (for example from 1 to 5 in which 1 indicated Very Familiar and 5 indicated Not Familiar) according to the answers provided by customers for their profile sheets with a wide range of values.

As indicated in [11], topic familiarity is an important factor influencing information seeking. It is possible to infer topic familiarity from information search behavior including reading time and searching efficacy. The familiarity value associated with each search topic is assessed on a 1 to 5 point Likert-type scale, where 1 was not at all, 3 was somewhat and 5 was extremely familiar. Reading time was derived from the search logs and search efficacy was measured as the ratio of the number of saved documents to the total number of viewed documents. Experimental results involving 36 volunteers indicate that searching efficacy increases and reading time decreases with topic familiarity.

User's cultural familiarity matches Web frequency. Web frequency may be used as a measure of cultural familiarity [12]. A tool called the Association Engine has been built to predict familiarity in a way that strongly correlates with human judgments by using only the frequency with which a word is found in Google's index. Experiments were carried out to measure the familiarity based on the subject's answers for the two questions: how often she sees the term and how she rates her level of understanding of its meaning. Both ratings were on a 5 point Likert-type scale (5 being see a lot and very familiar, 1 being never seen and not familiar).

Based on analysis done by many researchers' work in the field of psychology and sociology, we explore a variety of human factors that may affect user familiarity with a Web site. By building the hierarchy of all the factors, we map them to the properties in the Web application domain. A method of initializing familiarity value and a formula of updating it are proposed as well. Experiments will be carried out in the future work to evaluate the performance of the proposed familiarity measurement.

## III. Factors Affecting Familiarity

Human factors psychologically affecting familiarity have to be found in order to measure familiarity. As discovered by many researchers, it may be concluded that the major factors include prior experience, repeated exposure, level of processing, study duration and forgetting rate. A mapping from the human factors to the properties in the Web application domain will be clarified as well.

### *A. Exploration of Factors*

A review of 30 years of research is given for the purpose of distinguishing recollection and familiarity [13]. Although aging does not significantly affect familiarity because familiarity is different from recollection, some factors discovered in the review are empirical findings, such as study duration, forgetting rates, level of processing and so on. Perceptual matching is one factor mentioned in [13]. Changing the modalities of an object leads to decrease in familiarity. However, the relationship between familiarity and implicit memory is also mentioned in [13]. A lot of research shows that familiarity is functionally dissociable from performance on perceptual implicit memory tasks. Therefore, we do not take changes made on Web sites into account. Whittlesea [14] suggested that feelings of familiarity can be aroused even without prior experience if the perceptual processing of the stimulus is fluent. However, we are not interested in the fluency of the processing of the stimulus as long as we believe that the understanding or learning of services provided by a Web site is not fluent. On the other hand, Whittlesea did point out that prior experience of a stimulus can produce the feeling of familiarity. Two experiments were carried out in [15] to explore the relationship between familiarity and similarity. Note that the factors of properties of different objects that will affect familiarity are not included because only different levels of familiarity with the (roughly) same object is analyzed in our work. The factor of slight and incremental change to the Web site will be taken into account when we predicate the formula of updating familiarity value. Exploration of each factor is further described separately as follows.

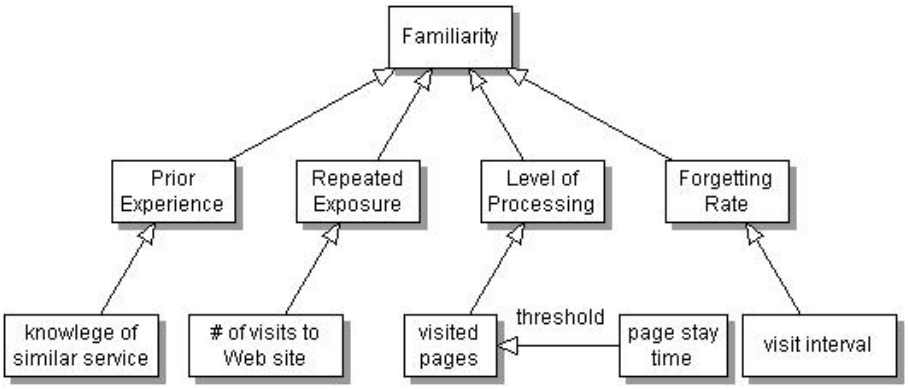


Fig. 1. Mapping from Human Factors to Properties in the Web Application Domain

*Prior experience* produces feelings of familiarity. The source of prior experience is not necessarily the object itself, but the meaning of it or an object which semantically relates to the current object. According to [15], similarity has an effect on the feeling of familiarity as well. Prior experience is based on the Web sites providing similar services, of course, and will have an effect on the feeling of familiarity. User's prior experience information can be provided directly by the user through a questionnaire.

The experiments carried out in [15] show that *repeated exposure* will affect the feeling of familiarity. The feeling of familiarity will increase after each stimulus. The more times the user visits the Web site and uses the services, the more familiar she will be with the Web site.

*Level of processing* is associated with how much familiarity can be gained [13]. Deep processing (processing the meaning) leads to an greater increase in familiarity than shallow processing (processing the perceptual aspects). Deep processing and shallow processing produce different feelings of familiarity. The difference between deep processing and shallow processing here is between processing the meaning and processing the perceptual aspects. In our work, we treat the difference as how deeply the user uses the Web services. For instance, visiting different numbers of Web pages and Web pages with different importance produces different feelings of familiarity.

An increase in *study duration* leads to corresponding increases in familiarity [13]. Study duration in our work is defined as the stay time on a Web site. The longer a user stays on a Web site, the more pages she may read, and therefore, the greater feeling of familiarity with the services she may gain. However, the stay time on individual page must be taken into account because of the possibility that, for instance, one may come to a Web site and open a Web page, but then go away for a cup of coffee. Stay time on a page is also used to evaluate the relevance of the page to the level of processing.

Both immediate delays and long-term delays decreases familiarity. *Forgetting rate* is determined by the interval between two times of visits to the Web site. The longer the interval between visits, the greater the decrease in the user's feeling of familiarity.

### B. Factors Hierarchy

As explored above, familiarity is affected by the five major factors: prior experience, repeated exposure, level of processing, study duration and forgetting rate. The factors are further determined by user's interaction with the Web site. A mapping from the factors to the properties in the Web application domain is shown in Figure 1. Prior experience is determined by knowledge of the Web sites providing similar services. Repeated exposure is represented by how many times the user has been to the Web site and used the services. Level of processing is determined by the page requests that the user made during each visit. Page stay time is the threshold to evaluate the relevance of the visited page. Forgetting rate is calculated by the interval between the last visit and the current visit, and the factor of the Web site itself described later. Note that study duration is not included in the hierarchy because the number of visited pages and page stay time together can represent study duration.

## IV. Measuring Familiarity

Once a new user comes to the Web site, her familiarity value will be initialized based on her prior experience with the same or similar services provided by other Web sites. Her feeling of familiarity with a Web site will be updated after each time that she visits the Web site again. Her familiarity value will be decreased or increased based on all the factors explored earlier, including repeated exposure, level of processing, study duration and forgetting rate.

### A. Initializing Familiarity Value

Prior experience is determined by how much experience the user has with similar Web sites. As Söderlund suggested [7], the familiarity can be measured through two questionnaires, which are carefully constructed through two processes: defining candidate questions regarding the five major factors (presented earlier in the paper), and validating the questions. The first questionnaire is distributed to viewers, and the second questionnaire will be distributed to only the ones who return the first questionnaire. The zero-order correlation of two sets of

answers is used to deal with the problem of vagueness. A well-defined questionnaire is very important to determine the user's prior experience with other similar Web sites. A 100 point scale for 10 questions about visiting Web sites and using the Internet may be specified. The general questions may be, for example: "How often do you use the Internet?", "How often do you use the same or similar Web services?", "How familiar do you think you are with the same or similar Web services?", and so on. The answers from the second questionnaire or the average of two answers may become the initial familiarity value($F_1$ in the later formulas). For the later use, the prior knowledge ($S_1$ in the later formulas) needs to be calculated as follows:

$$S_1 = -\ln(1 - \frac{1}{F_1}) \tag{1}$$

### B. Updating Familiarity from Knowledge

The value of user familiarity with the Web site can be calculated from the user's knowledge about the Web site as follows:

$$F_k = \frac{1}{1 + e^{-S_k}} \tag{2}$$

where $F_k$ and $S_k$ represent the familiarity value and the knowledge value of the $k$th (current) visit to the Web site, respectively.

Since the familiarity value is affected by the previous level of processing and the forgetting rate, and it is determined by the user's knowledge, a simple formula of updating user's knowledge may be as follows:

$$S_k = S_{k-1} + L_{k-1} - R_{k-1} \qquad (k \geq 2) \tag{3}$$

where $S_{k-1}$ represents the user's knowledge value of the $(k-1)$th visit (the last visit) to the Web site, $L_{k-1}$ is the level of processing of the last visit to the Web site, and $R_{k-1}$ represents forgetting value since the last visit. The initial value of user's knowledge, $S_1$, has been determined by Equation 1.

According to Bahrick's work [16], the learning curve is similar to an exponential curve. It is affected by the pre-knowledge that the user has. Thus, the formula to calculate the current level of processing may be assumed as follows:

$$L_{k-1} = S_{k-1}(1 - e^{-Q_{k-1}/l}) \qquad (k \geq 2) \tag{4}$$

where $Q_{k-1}$ represents the previous learning quality and $l$ represents the learning coefficient. The value of $Q_{k-1}$ can be adjusted based on the domain and the particular circumstance. It is affected by how many and which Web pages have been visited by the user during the previous session. The value of $l$ may differ for different domains with different learning materials.

After a previous visit to the Web site, the user left and started forgetting. The forgetting value is, of course, based on the knowledge that the user has of the Web site up to the moment when she left. Thus, the forgetting value $R_{k-1}$ can be calculated as follows:

$$R_{k-1} = (S_{k-1} + S_{k-1}(1 - e^{-Q_{k-1}/l}))(1 - r_{k-1}) \tag{5}$$

where $r_{k-1}$ is the retention rate for the last visit to the Web site. As discovered by Hermann Ebbinghaus in 1885 [17], forgetting has an exponential nature. Thus, the retention rate can be roughly described by the following formula:

$$r = e^{-\triangle t/S} \tag{6}$$

In Equation 5, $S$ is the relative memory strength. $\triangle t$ represents the time difference between the current visit and the last visit to the Web site. It is very interesting that the basal forgetting rate differs little between individuals. Therefore, $S$ slightly changes for different individuals. We replace it by another symbol, $m$, which represents the memory coefficient. It may differ largely for different domains with different structural characteristics. The formula for calculating retention rate is finally as follows:

$$r_{k-1} = e^{-\triangle t_{k-1}/m} \qquad (k \geq 2) \tag{7}$$

and the formula to calculate the forgetting value becomes as follows:

$$R_{k-1} = S_{k-1}(2 - e^{-Q_{k-1}/l})(1 - e^{-\triangle t_{k-1}/m}) \tag{8}$$

Overall, the current user's knowledge about the Web site can be calculated by the formula as follows:

$$S_k = S_{k-1}(2 - e^{-Q_{k-1}/l})e^{-\triangle t_{k-1}/m} \qquad (k \geq 2) \tag{9}$$

One case that must be taken into account is that the user might visit other Web sites that provide the same or similar services during her visit to the Web site. Thus, a factor $s$ is introduced to take both cases into account. The formula for calculating the current user's knowledge becomes

$$S_k = S_{k-1}(2 - e^{-Q_{k-1}/l})e^{-\triangle t_{k-1}/m} + s \qquad (k \geq 2) \tag{10}$$

The value of $s$ can be determined approximately from experiments that will be introduced in Section V.

### C. Measuring the Learning Quality

The current level of processing is determined by the pre-knowledge that the user has about the Web site and the current learning quality. The current learning quality is further affected by how many and which Web pages have been visited by the user during the current user session. In order to calculate the current learning quality, each page needs to be assigned by a knowledge importance value. The current learning quality is then the sum of the knowledge importance values of all the visited pages in the current user session. The knowledge importance of a page from the user's point of view is equivalent to the importance of a page from the page's point of view because the user can gain more knowledge from the more important pages. As pointed out by Brin and Page [18], the importance of a Web page can be judged by the number of links pointing to it from other Web pages (within the Web site, in our case). They invented the very popular PageRank algorithm to determine the importance of a page. An extended PageRank algorithm, called the Weighted PageRank (WPR) algorithm [19], is applied to determine the knowledge

importance of a page in our work. The WPR algorithm takes into account not only the importance of the inlinks (links to a page) but also the outlinks (links from a page) , and distributes rank scores based on the popularity of the pages. Unlike the PageRank algorithm dividing the rank value of a page evenly over the pages to which it links, the WPR algorithm assigns larger rank values to more important pages. Based on the idea of using popularity from the number of inlinks and outlinks, the WPR formula to calculate the importance of page is represented as

$$PR(u) = (1-d) + d\sum PR(v)W^{in}_{(v,u)}W^{out}_{(v,u)}$$

where $d$ is a dampening factor that is usually set to 0.85. $PR(u)$ and $PR(v)$ are importance of page $u$ and $v$, respectively. $W^{in}_{(v,u)}$ and $W^{out}_{(v,u)}$ are weights of $link(v,u)$ calculated differently.

## V. Conclusions and Future Work

We summarized the relationship between familiarity and trust in the literature review. The relationship has been further clarified through a new model of trust. Measurement of familiarity may contribute to the formalization of trust. We explored the factors mainly affecting familiarity. The five factors include prior experience, repeated exposure, level of processing, study duration and forgetting rate. However, the feeling of familiarity may be affected by many other factors such as aging, user's interests, user's major, user's education level and so on. We do not consider them here because they are not comparatively important. The human factors were mapped to the properties of the Web application domain. Finally, we came up with a convenient way to measure and update familiarity value. The performance of the measurement will be evaluated by future experiments described in detail as follows.

In the future work, experiments will be carried out to determine the values of the learning coefficient $l$ and the memory coefficient $m$, to adjust the knowledge value $s$ from other similar Web sites, and to evaluate the accuracy of the familiarity measuring algorithm. Participants, materials and experimental procedure will be discussed below.

A large number of participants from the University of New Brunswick will be involved in the study. They will come from different faculties including Computer Science, Business Administration, Chemical Engineering and Electrical and Computer Engineering. The selection of all the participants will cover different study fields, different education levels and different ages.

The Web site of Saint Thomas University will be chosen as material in the experiments. The reasons for choosing the Saint Thomas University Web site include: a) the Web site contains very rich hyperlinks in order to satisfy the WPR algorithm's needs; b) normally people from the University of New Brunswick do not visit it, but they do have prior knowledge about university Web sites; and c) the Web site does provide particular services that people are concerned with, including admission, programs, student services, course registration, housing and so on. Four major processes will be applied to assign knowledge importance to each page, including building a Web map, finding the root set, finding the base set and applying the WPR algorithm.

Before the participants start visiting the Saint Thomas University Web site, a well-defined questionnaire will be distributed to them to evaluate their prior knowledge of university Web sites. The same questionnaire will be distributed to the participants when they return the first one. The zero-order correlation of two sets of answers will be used to deal with the problem of vagueness. The average value from two questionnaires will be used as the prior knowledge value ($S_1$ in the formulas presented earlier in the paper).

When the participants are visiting the Web site, the visited pages will be recorded. Learning quality value $Q$ will be calculated for each participant's visit according to the discovered importance of the visited pages. After each visit, the participants will be asked to answer the questionnaire again to measure their current familiarity. After a period of time (a few days, for example), the same process will be repeated and the time difference will be recorded as well.

Partial results will be used to determine the values of $l$, $m$ and $s$. Others will be used for testing. The approach of K-fold cross validation will be used to evaluate the performance of the proposed familiarity measurement.

## VI. Acknowledgments

This work was funded by the Atlantic Canada Opportunity Agency (ACOA) through the Atlantic Innovation Fund (AIF) and through grant RGPIN 227441-00 from the National Science and Engineering Research Council of Canada (NSERC) to Dr. Ali A. Ghorbani.

## References

[1] A. Barr, "Familiarity and trust: An experimental investigation," *Working Paper. Oxford: Centre for the Study of African Economies*, 1999.

[2] H. H. Cao, D. Hirshleifer, and H. H. Zhang, "Fear of the unknown: The effects of familiarity on financial decisions," in *WFA*, 2002.

[3] G. Huberman, "Familiarity breeds investment," *Journal of Review of Financial Studies*, vol. 14, no. 3, pp. 659–80, 2001.

[4] J. Carter and A. A. Ghorbani, "Value centric trust in multi-agent systems," in *Proceedings of IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 2003.

[5] ——, "Towards a formalization of value-centric trust in agent societies," *to appear, Journal of Web Intelligence and Agent Systems*, 2004.

[6] J. Nordfalt and M. Soderlund, "Customer experience and its effects on attitudes and future behavior," in *Edvardsson B, Brown SW, Johnston R and Scheuing EE (eds.), Service Quality in the New Economy: Interdisciplinary and International Dimensions*, 2000.

[7] S. Magnus, "Customer familiarity and its effects on post-purchase evaluations," in *Breivik E, Falkenberg AW and Gronhaug K (Eds), Rethinking European Marketing, Proceedings from the 30th EMAC Conference, NHH, Bergen*, 2001.

[8] S. Magnus and J. Gunnarsson, "Customer familiarity and its effects on satisfaction and dissatisfaction," in *SSE/EFI Working Paper Series in Business Administration*, 2000.

[9] S. A. Himmanen, K. Gentles, and K. Sailor, "Rated familiarity, visual complexity, and image agreement and their relation to naming difficulty for items from the boston naming test," *Journal of Clinical and Experimental Neuropsychology*, vol. 25, no. 8, pp. 1178–1185, 2003.

[10] J. Snodgrass and M. Vanderwart, "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, pp. 174–215, 1980.

[11] D. Kelly and C. Cool, “The effects of topic familiarity on information search,” in *In Proceedings of the Second ACM/IEEE Joint Conference on Digital Libraries (JCDL 2002), Portland, OR.*, 2002.
[12] D. A. Shamma, S. Owsley, S. Bradshaw, S. Sood, and K. J. Hammond, “Using web frequency as a measure of cultural familiarity,” *ACL 2004*, 2004.
[13] A. P. Yonelinas, “The nature of recollection and familiarity: A review of 30 years of research,” *Journal of Memory and Language*, vol. 46, pp. 441–517, 2002.
[14] B. W. A. Whittlesea, “Illusions of familiarity,” *Journal of Experimental Psychology: Human Learning and Memory*, vol. 19, no. 6, pp. 1235–1253, Nov 1993.
[15] R. L. Moreland and R. B. Zajonc, “Exposure effects in person perception: Familiarity, similarity, and attraction,” *Journal of Experimental Social Psychology*, vol. 18, no. 5, pp. 395–415, Sep 1982.
[16] H. P. Bahrick, “Semantic memory content in permastore: Fifty years of memory for spanish learned in school,” *Journal of Experimental Psychology: General*, vol. 113, pp. 1–37, 1984.
[17] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology*. New York by Teachers College, Columbia University, 1885.
[18] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” in *Computer Networks and ISDN Systems, 30(1-7)*, 1998, pp. 107–117.
[19] W. Xin and A. Ghorbani, “Weighted pagerank algorithm,” in *Proceedings of the 2nd Annual Conference on Communication Networks and Services Research (CNSR 2004)*, vol. 1, 2004, pp. 305–314.