Applying Hebbian Theory to Enhance Search Performance in Unstructured Social-Like Peer-to-Peer Networks

Chester S.J. Huang, Stephen J.H. Yang, and Addison Y.S. Su

Unstructured peer-to-peer (p2p) networks usually employ flooding search algorithms to locate resources. However, these algorithms often require a large storage overhead or generate massive network traffic. To address this issue, previous researchers explored the possibility of building efficient p2p networks by clustering peers into communities based on their social relationships, creating social-like p2p networks. This study proposes a social relationship p2p network that uses a measure based on Hebbian theory to create a social relation weight. The contribution of the study is twofold. First, using the social relation weight, the query peer stores and searches for the appropriate response peers in social-like p2p networks. Second, this study designs a novel knowledge index mechanism that dynamically adapts social relationship p2p networks. The results show that the proposed social relationship p2p network improves search performance significantly, compared with existing approaches.

Keywords: P2P query routing, social networks, semantic search.

Manuscript received: Sept. 17, 2011; revised Apr. 09, 2012; accepted Apr. 19, 2012.

This work was supported by the National Science Council, Taiwan, under grant numbers NSC98-2511-S-008-006-MY3, NSC98-2511-S-008-007-MY3, and NSC99-2511-S-008-006-MY3, and the Research Center for Science & Technology for Learning of the University System of Taiwan.

Chester S.J. Huang (phone: +886 3 4227151 ext 35326, shinjia.huang@gmail.com) and Stephen J.H. Yang (corresponding author, jhyang@csie.ncu.edu.tw) are with the Department of Computer Science & Information Engineering, National Central University, Taiwan.

Addison Y.S. Su (ncuaddison@gmail.com) is with the Advanced Communication Laboratory, National Central University, Taiwan.

http://dx.doi.org/10.4218/etrij.12.0111.0588

I. Introduction

In the past few years, peer-to-peer (p2p) networks have become popular for privacy, autonomy, and resource sharing [1]. Structured p2p networks include Chord [2], Can [3], and Pastry [4], whereas unstructured p2p networks include Gnutella [5], among others. Structured p2p networks use distributed hash tables to provide a natural platform for keyword-match searching [6]. This mechanism guarantees the retrieval of existing resources and provides an upper bound to retrieve costs. In contrast, unstructured p2p networks use flooding or random walk to visit all peers and thus consume significant amounts of bandwidth and decrease search performance.

To solve these problems, previous studies proposed social-like p2p networks that improve search performance by simulating the social interaction in unstructured p2p networks [7], [8]. Much akin to human social networks, in which people connect with each other through mutual interests, two peers in a social-like p2p network may connect if they are interested in each other's resources. Such social-like p2p networks allow peers to enrich their capabilities, build a local knowledge index, and use this index efficiently for peer selection. Although this mechanism elevates bandwidth consumption, it increases the efficiency of sending queries [9], [10]. Related studies show that current social-like p2p networks lack three critical factors to address poor performance in resource discovery.

The first factor is a lack of information on certain responding peers [11], [12]. Current studies propose that a peer use previous query responses from its neighbors to help determine how to send new query messages. Each peer in a social-like p2p network builds a local knowledge index to record query responses from its neighbors but does not examine whether each neighbor is a successfully responding peer. Thus, this method is relatively inefficient for collecting and storing relevant information on the content of other peers in social-like p2p networks.

The second factor is a lack of support for semantic search [13]. The keyword-match search mechanism employed in current social-like p2p networks is relatively inefficient when compared to sophisticated information retrieval techniques. This is because p2p networks lack useful global knowledge on popular resources and the relationships between keywords and resources, which are difficult to obtain in social-like p2p networks. Although existing networks support keyword-match searching, they do not support semantic searches. Thus, these networks can only find resources that have the exact keyword indicated in a query and have difficulty satisfying user requirements.

The third factor is the failure to maintain a consistent distributed knowledge index [14]. Related studies use the least recently used (LRU) algorithm to maintain a local knowledge index without duplicates. This approach places the most recently used keyword at the top and the least recently used keyword at the bottom. The local knowledge index then deletes the least recently used keyword after reaching its maximum. A major disadvantage of the LRU algorithm is that it considers only the problem of removing the least recently used keyword and thus cannot retain a useful keyword in the local knowledge index.

To solve these critical problems, the proposed social relationship p2p network uses a measure based on Hebbian theory [15] to create a social relation weight. Social relationship p2p networks use this social relation weight to select the appropriate respond peer. Each peer can adjust the social relation weight to improve search performance. We conduct comprehensive trace-driven simulations to evaluate this study. The results show that the proposed social relationship p2p network improves search performance significantly, compared with existing approaches. Our contributions are twofold. First, by using the social relation weight, the query peer stores and searches for the appropriate response peers in social-like p2p networks. Second, this study designs a novel knowledge index mechanism that dynamically adapts social relationship p2p networks.

The remainder of the paper is organized as follows: section II presents a brief summary of related works on searching in social-like p2p networks, section III presents the proposed social relationship p2p network to enhance search performance, section IV introduces the experimental method, section V discuss the evaluation results, and, lastly, section VI provides a conclusion and presents directions for future research.

II. Related Works

1. NeuroGrid

NeuroGrid uses the experience of previous queries to help peers make routing decisions. Each peer builds a knowledge index based on the results of previous queries and supports distributed searches through its knowledge index. When a peer receives a query, it passes that query to the peers in its knowledge index who are directly associated with the keywords in the query. If the query peer receives a correct response message, the query peer updates its knowledge index to associate the responding peers with the query keywords. If a query peer does not receive a response message or provides resources that do not match the query, this approach randomly forwards the query to other connected peers. Therefore, this system is only effective for previously queried keywords and is inappropriate for networks in which peers join and leave quickly.

III. Social Relationship P2P Networks

We propose a social relationship p2p network that uses a measure based on Hebbian theory to create the social relation weight. This measure consists of six rules: 1) semantic similarity rule, 2) answer relation rule, 3) recommend relation rule, 4) degradation rule, 5) new peer rule, and 6) dynamic expansion rule. The proposed rules affect the social relation weight of each peer. The following sections detail each rule.

1. Semantic Similarity Rule

When sending a query message, a peer starts the semantic similarity rule and finds possible response peers from its knowledge index. Initially, this study used an ontology-based Open Directory Project (ODP) (otherwise known as DMOZ) hierarchy to generate the domain knowledge structure and the interest topic profile in the peer knowledge index. Based on the DMOZ hierarchy, the semantic similarity (sim_{Topic}) between the keywords of a query message q_t and the keywords of neighbor peers s_t stored in the knowledge index is calculated by

$$sim_{\text{Topic}}(q_t, s_t) = \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & \text{if } q_t \neq s_t, \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

If the height of q_t in the DMOZ hierarchy is represented by $h1$ and the height of s_t in the DMOZ hierarchy is represented by $h2$, the h is determined by the lesser of the respective values of $h1$ and $h2$, and the l is determined by the shortest path between q_t and s_t. Additionally, α and β represent weight control parameters for l and h, respectively. In line with the research conducted by Li and others [13], this study sets the parameters as α=0.2 and β=0.6.

Based on (1), this study obtains the semantic similarity sim_{Topic} and determines the social relation weight (*SRW*) of neighbor peers using the knowledge index. Then the rank value of each peer using semantic similarity and social relation weight, while using variables a and b to achieve a flexible search result [14], is calculated by

$$Rank = \frac{a * SRW + b *}{(a+b)} \text{ or } \frac{SRW + sim_{\text{Topic}}}{2}. \tag{2}$$

Variable a represents the proportion of social relation weight (set at 0.2 in the experiment), while variable b represents the proportion of semantic similarity (set at 0.2 in the experiment). The results are used to rank the value of each peer. According to the level of value, the rule is adjusted to arrange the order of query messages. For example, peer A uses (1) to obtain the semantic similarity and the social relation weight of neighboring peers. Using (2) to calculate the rank value of each peer, peer A only sends a query message to each of the three highest-ranking peers, B, C, and D.

2. Answer Relation Rule

According to the previous rule, if peer D responds to peer A, peer A receives the response and begins the answer relation rule. The answer relation rule allows peer A to judge the search result of the response message.

The social relation weight $w_{A\text{D}}$ and $\Delta w_{A\text{D}}$ are calculated by

$$w_{A\text{D}}' = w_{A\text{D}} + \Delta w_{A\text{D}}, \quad \Delta w_{A\text{D}} = \gamma \cdot x_A \cdot x_{\text{D}}, \tag{3}$$

where $w_{A\text{D}}$ is the social relation weight of peers A and D, as determined by previous search results, while $\Delta w_{A\text{D}}$ is the social relation weight of peers A and D, as determined by current search results. The product of $\Delta w_{A\text{D}}$ is multiplied by γ, x_A, and x_D, where γ is the learning rate, x_A is the parameter of appropriate resources of peer A, and x_D is the parameter of response messages of peer D. If the search time of peer A transpires in a specific time period, then x_A is 1. Otherwise, x_D is 0. If the response time of peer D transpires in a specific time period, then x_D is 1. Otherwise, x_A is 0. The learning rate γ is calculated by

$$\gamma = \begin{Bmatrix} q \cdot \left(1 - \dfrac{\tau}{T+1}\right) \cdot \sigma, & \tau \le T \\ 0, & \tau > T \end{Bmatrix}. \tag{4}$$

The term q represents the value of the resource provided. The term σ represents the learning parameter. According to Ghanea-Hercock and others [14], the learning parameter σ of peer A and peer D is 1 when forming social-like p2p networks. If the network exists for a period of time, the learning parameter of peer A and peer D is 0. Otherwise, the learning parameter σ is 1. The term τ represents the response time in the current search process. The term T represents the maximum search time in the current search process. The functions of parameters τ and T help avoid a lengthy record search and response time for peers. In (3), the social relation weight $\Delta w_{A\text{D}}$ for the current search results plus the social relation weight $w_{A\text{D}}$ for the previous search results equals the new social relation weight $w_{A\text{D}}'$. Peer A records the new social relation weight $w_{A\text{D}}'$ in the knowledge index and employs the limited value Lv (set at 0.001 in experiments in this study). The knowledge index of peer A only stores peer D and the social relation weight $w_{A\text{D}}'$ that exceeds the limited value Lv.

3. Recommend Relation Rule

In the semantic similarity rule, peer A sends a query message to each of the highest-ranking peers B, C, and D from the knowledge index. However, peers B, C, and D may return incorrect search results to peer A. To solve this problem, peer D initiates the recommend relation rule. Initially, peer D uses the semantic similarity rule to find the social relation weights w_{DE}, w_{DF}, and w_{DG} and uses (2) to sort these weights. Based on this ranking, peer D selects the three highest-ranking peers E, F, and G and sends each of them a query message.

When peer D receives search results from peer G, (3) is used to calculate the new social relation weight w_{DG}'. Additionally, peer D stores the w_{DG}' in the knowledge index. Thereafter, peer D returns correct search results to peer A. Peer A uses the recommend relation rule to adjust the social relation weight w_{AD}' because the search result is closer to the demand of peer A. The term φf represents the adjustment parameters of the recommend relation rule. The φf value (set at 0.5 in this experiment) multiplied by the previous social relation weight $w_{A\text{G}}$ of peer G equals the new social relation weight w_{AG}', that is,

$$w_{AG}' = \varphi f \cdot w_{AG}. \tag{5}$$

Subsequently, the new social relation weight w_{AG}' plus the current social relation weight w_{AD} of peer D equals the social relation weight w_{AD}', that is,

$$w_{AD}' = w_{AD} \cdot w_{AG}'. \tag{6}$$

4. Degradation Rule

Peer A initially uses the semantic similarity rule and the answer relation rule to send query messages to neighboring peers. However, social relationship p2p networks include peer variable characteristics; the social relation weights of peers can change at any time. For instance, if peer A previously sent query messages to peers B, C, and D often but has not sent query messages to them in a long time, then the searching performance of peer A might decrease, in which case the social relation weights of peers B, C, and D must be adjusted. Thus, we have designed a mechanism that manages the social relation weight of peer A and adjusts the social relation weights of peers B, C, and D. The degradation rule refers to the time-to-live (TTL) of peers and previous search activities between peers to adjust the social relation weight. If peer A has not searched the knowledge index of peer D for a long time, then peer A adjusts the social relation weight of peer D. In the process of adjusting the social relation weight, peer A and peer D may search for each other again and their social relation weights are gradually reduced by

$$w_{ad}(s+1) = w_{ad}(s) \cdot \exp(-\frac{s}{\varphi d}). \tag{7}$$

In (7), the term s represents the time since the last update, whereas φd (set at 0.8 in this experiment) represents a constant between 0 and 1. The term φd represents the degradation parameter of the degradation rule. This rule is simply expressed as

$$w_{ad}(s+1) = w_{ad}(s) \cdot \varphi d. \tag{8}$$

If peer A has not recently searched for peer D, it gradually adjusts the limited value Lv. Peer A then removes peer D from its knowledge index.

5. New Peer Rule

The proposed semantic similarity rule, answer relation rule, and recommend relation rule establish social relation weights between peers to improve overall searching efficiency. These rules gradually form the social relation weight after a long search period. However, this approach is unsuitable for new peers because they have no search history or a relation to any peers. Therefore, new peers must use a flooding search algorithm to search for resources. Thus, if many new peers join the network at once, they decrease the overall search performance. The proposed new peer rule solves this problem. The new peer rule defines a bootstrapping value, which indicates the extent of peer search and response. The bootstrapping value can be calculated by

$$Bootstrapping = (1 + Indegree) * (1 + Outdegree) . \tag{9}$$

The outdegree represents the number of peers within the knowledge index. During the search process, the query peer attaches the bootstrapping value, enabling the next peer to record it in the knowledge index after receiving the query message. Thus, when the new peer K joins the network, it randomly contacts peers. If peer F receives request messages, then peer F returns peer G with the highest bootstrapping value to new peer K. Finally, peer K connects to peer G, and it future-searches relevant peers and information.

6. Dynamic Expansion Rule

The main function of the degradation rule is to reduce the social relation weight between peer A and peers B, C, and D in the knowledge index if peer A has not frequently searched peers B, C, and D with a high social relation weight or peer A has not searched for a long time. In this case, the social relation weight between peer A and peers B, C, and D gradually decreases in the knowledge index. If the social relation weight of peer A is less than the limited value Lv, then peers B, C, and D are removed from the knowledge index of peer A. However, this can create problems when peers delete too much from the knowledge index, which affects overall search performance negatively. Therefore, this study presents a design of a dynamic expansion rule that dynamically links peers with high bootstrapping values and establishes social relation weights to enable an effective search process. In this approach, peer A finds high pre-bootstrapping values of n peers in the knowledge index, assuming that $n = 3$.

Peer A sends a query message to peers L, M, and N, all of whom have high bootstrapping values. When peers L, M, and N receive this message, they find the m peers with high bootstrapping values in the knowledge index. If these m peers with high bootstrapping values are peers O, P, and Q, then peers L, M, and N recommend these higher bootstrapping peers to peer A. Peer A then establishes social relation weights for peers O, P, and Q. Based on this description, the dynamic expansion rule solves the problem of the degradation rule of deleting too many peers from the knowledge index.

IV. Simulation Settings

The simulations in this study are conducted in a realistic environment to evaluate the performance of social relationship p2p networks. A NeuroGrid simulator [16] is used to realize the simulation in a social relation p2p environment and to adjust parameter optimization settings. NeuroGrid provides a framework for finding resources within a social relation p2p environment and is based on the idea of automating the process. NeuroGrid, which is an open-source simulator and an unstructured p2p network environment, uses a semantic searching (semantic routing) platform to build the experimental environment.

1. Content Creation and Distribution

The ODP, or DMOZ, is an ontology-based data set, and it is a popular and authoritative classification of content distributed among a community of content editors. Hasse and others [17] demonstrated that the DMOZ is suitable for their p2p system. To approximate our simulation to the real world data structure, we apply this data set to a random generation of a domain knowledge structure and data (including topics, documents, and keywords). The data set is so large that, for our simulation, we select a subset consisting of the first three levels of the DMOZ.

Previous studies [18] have shown that the distribution of documents and keywords in the domain knowledge structure can be approximated by Zipf's law in the form of $\sim 1/x^{\alpha}$, where y is the frequency, x is the rank, and α is the constant. The study simulation follows this estimated distribution to generate documents and keyword distribution in *dmoz.org*. Each simulation generates 1,000 keywords and distributes them to 2,000 documents, and each document is randomly assigned three keywords.

Previous measurement studies [19] indicate that the distribution of shared documents in p2p networks is unbalanced. Some peers in existing p2p networks tend to download many documents but share few documents or none at all. To avoid this situation, each peer in this study has several interest topics and shares a number of network documents and keywords through a probabilistic method. Most shared documents and keywords are relevant to the interest topics of a peer, with a probability of 90%, but are occasionally irrelevant to this area. For each relation, at least one of the keywords in each document should match the interest topics of the query peer. Based on the DMOZ hierarchy, we set 40 interest topics that are generated and each cover 40 keywords in this experiment.

2. Query Message Generation and Delivery

Each simulation in this study chooses a random peer as the query peer and starts a search on a keyword. We further define the interest topic profile in the peer knowledge index to describe each peer interest shift, as follows:

Definition 1: The interest topic profile consists of three attributes to describe primitive attributes of a peer interest, namely, topic index, interest topic, and the number of keywords. Each topic is assigned a unique identifier. Each peer has his or her own interest topics. In addition, each interest topic includes a number of keywords.

Ren and others [20] agreed that peer interest shift is a vital factor in p2p networks. To consider this factor, 1% of peers randomly shift interests in the experiment, and their subsequent major query messages follow the new interest. According to the peer interest shift factor, each simulation in the experiment chooses a random peer as the query peer and starts a search on a query message. Each query message is randomly selected from the interest topics of the query peer with a probability p (p = 90%) or occasionally from a random area with a probability of $1 - p$. Each query message is tagged with a TTL to limit the lifetime of a message to four hops in the experiment. The control parameter $K_{\max}$ is also added to the query message. Liu and others [8] defined the control parameter $K_{\max}$, which is the maximum number of hops. Following the suggestion in [8], the TTL sets 4 and the control parameter $K_{\max}$ sets 3 in this experiment.

3. Network Initialization

Saroiu [21] suggests that certain p2p networks are scale-free networks in which the connectivity of peers follows a power-law distribution: $p(k) = \alpha \cdot k^{-\gamma}$. The probability $p(k)$ that a peer in the network connects with k other peers is proportional to k^{-r}. Because various factors affect the distribution of connectivity, it is unreasonable to generate a random power-law distribution of connectivity in the simulation, irrespective of other peer characteristics. To observe the evolution of network topology, the simulations start from a small random network. Each peer randomly and bidirectionally connects to four peers to generate a random topology. Each peer maintains approximately eight links at the beginning of each simulation. At the beginning, because no interaction occurs between peers, each peer keeps an empty knowledge index capable of containing a maximum of 80 keywords and associated peer addresses (if no other size is specified).

4. Network Evolution

Media reports [8] indicate that p2p networks are growing rapidly on the Internet. However, [18] indicates that the size of some p2p networks have remained constant. Most p2p simulations do not consider the phenomenon of consistent rapid growth. Therefore, we simulate a growing network that starts with a small set of 700 peers. In this simulation, one hundred peers join the network in every loop until the population reaches 1,400 peers.

5. Simulation Parameter Settings

The simulations in this study use a round-based simulation framework based on the parameter settings listed in Table 1. Each simulation generates 1,000 keywords and distributes them to 2,000 documents. Each document is randomly assigned three keywords. Each simulation generates 40 interest

Table 1. Simulation parameter settings.

Parameter name	Value
Documents	2,000
Keywords	1,000
Peer number	700
Query number	30,000
Time to live (TTL)	4
Limited value (Lv)	0.001
Maximum number of hops (K_{max})	3
Interest topics (N)	40
Total of peers joined the network every loop (T)	100
Learning parameter (σ)	1
Adjustment parameter (φf)	0.5
Degradation parameter (φd)	0.8
Proportion of social relation weight (a)	0.2
Proportion of semantic similarity (b)	0.2

topics, and each topic covers 40 keywords. One hundred peers join the network in every loop until the population reaches 1,400 peers [8]. The researchers [8] suggested setting the TTL to 4 and the control parameter K_{max} to 3 for this type of experiment. Our simulation shows the optimal value of the limited value Lv (set at 0.001). According to Ghanea-Hercock and others [14], the learning parameter σ of each peer is 1 when beginning to form a social-like p2p network. Variable a represents the proportion of social relation weight (set at 0.2 in the experiment), and variable b represents the proportion of semantic similarity (set at 0.2 in the experiment). Furthermore, [14] suggests that the adjustment parameter φf and the degradation parameter φd are in the range of 0 and 1. In our past experience with this simulation, φf and φd were set at 0.5 and 0.8, respectively.

6. Hypotheses

This study evaluates the proposed social relationship p2p network through several experiments to compare its performance with that of related approaches. Before presenting the final evaluation results, the following list summarizes the major hypotheses under investigation:

Hypothesis 1. Social relationship p2p networks have greater search precision than NeuroGrid and Gnutella.

Hypothesis 2. Social relationship p2p networks have higher recall than NeuroGrid and Gnutella.

Hypothesis 3. Social relationship p2p networks have higher F-measures than NeuroGrid and Gnutella.

Hypothesis 4. The semantic similarity mechanism enhances the search precision of social relationship p2p networks.

Hypothesis 5. The weight mechanism enhances the search precision of social relationship p2p networks.

Hypothesis 6. The degradation mechanism enhances the search precision of social relationship p2p networks.

Hypothesis 7. The dynamic expansion mechanism enhances the search precision of social relationship p2p networks.

Hypothesis 8. Social relationship p2p networks have higher message efficiency than NeuroGrid and Gnutella.

Hypothesis 9. Social relationship p2p networks have more messages per query than NeuroGrid and Gnutella.

V. Experiment Results

Figure 1 shows that social relationship p2p networks achieve greater precision by retrieving documents more quickly and accurately than others, proving Hypothesis 1. Because Gnutella does not own the social network concept, its search precision is low and is close to 0. In contrast, NeuroGrid uses the concept of social networks to achieve high search precision. We use 10,000 query messages as the reference point. For 10,000 query messages, the search precision of NeuroGrid is 0.21. At 22,000 query messages, the search precision is 0.43. However, the proposed social relationship p2p network has higher search precision than NeuroGrid, with an initial search precision of 0.25. For 10,000 query messages, the social relationship p2p networks increase search precision to 0.54 and further increase to 0.7 for 22,000 messages.

NeuroGrid and social relationship p2p networks exhibit significant differences because of the social relation weight mechanism and semantic similarity mechanism. If a system exhibits semantic similarity, it can recommend response peers to enhance the hit rate of query keywords. For instance, the knowledge index of peer C has the keyword “Algorithm” for peer A and the keyword “Graph Theory” for peer B. The weight of “Algorithm” is 0.8, and the weight of “Graph Theory” is 0.6. If peer C sends a query message including the keyword “Algorithm,” peer A receives a high ranking value based on weight rankings. However, peer B has a high ranking value based on the semantic similarity mechanism. The combination of the semantic similarity mechanism and the social relation weight mechanism enables the possibility to search for various keywords and find similar semantic peers.

To prove Hypothesis 2, this study provides a comparison of the search recall of the social relationship p2p networks NeuroGrid and Gnutella. Figure 2 plots the average recall of the social relationship p2p networks in each run. Gnutella has higher recall than NeuroGrid and the proposed social relationship p2p network. Gnutella uses flooding with limited

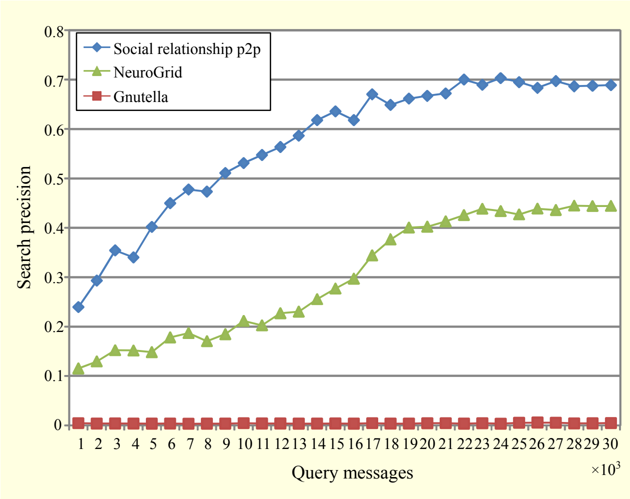

Fig. 1. Search precision.

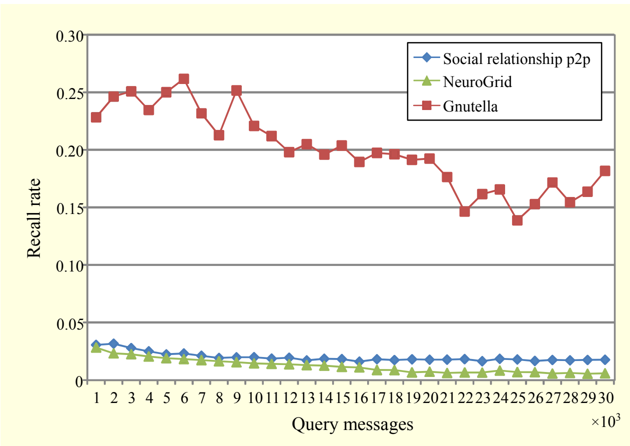

Fig. 2. Search recall.

TTL, but flooding produces too many search messages in the network. Thus, we only compare the search recall of social relationship p2p networks and NeuroGrid. We use the number of query messages (21,000) as the reference point, and the search recall of NeuroGrid and Gnutella is 0.02 and 0.17, respectively, whereas the search recall of the proposed method is 0.03. As the number of query messages increases, the search recall of the proposed social relationship p2p network exceeds that of NeuroGrid. This means that social relationship p2p networks provide more relevant information. This not only improves search precision but also provides users with more choice.

To prove Hypothesis 3, this study provides a comparison of the F-measure of social relationship p2p networks, NeuroGrid, and Gnutella. Figure 3 shows that social relationship p2p networks outperform NeuroGrid and Gnutella. This simulation uses the number of query messages (21,000) as the reference point. The F-measures of NeuroGrid and Gnutella are 0.012

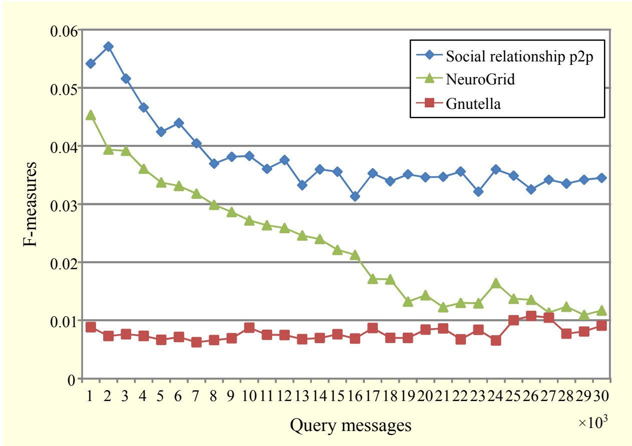

Fig. 3. F-measure.

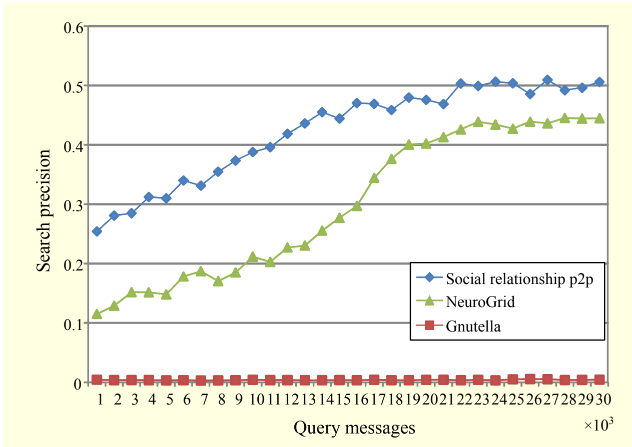

Fig. 4. Semantic similarity mechanism.

and 0.009, respectively, whereas the F-measure of the social relationship p2p networks is 0.035. The F-measure of social relationship p2p networks outperforms that of NeuroGrid and Gnutella while increasing per query message ($\times 10^3$). This means that the search precision, or the number of received messages, is more balanced than that in Gnutella or NeuroGrid.

To prove Hypothesis 4, this study provides a comparison of the search precision of social relationship p2p networks with NeuroGrid and Gnutella. This simulation tests the semantic similarity mechanism of social relationship p2p networks. Figure 4 shows that the semantic similarity mechanism of social relationship p2p networks improves the overall search precision effectively. This simulation uses 20,000 query messages as the reference point. Because Gnutella does not include semantic similarity mechanisms, it has to adopt a flooding approach to send query messages and use the keyword matching approach to match information. This causes Gnutella's search precision to remain at 0. NeuroGrid uses a knowledge index to resolve Gnutella's poor search precision; each peer uses keyword matching to match the peer who

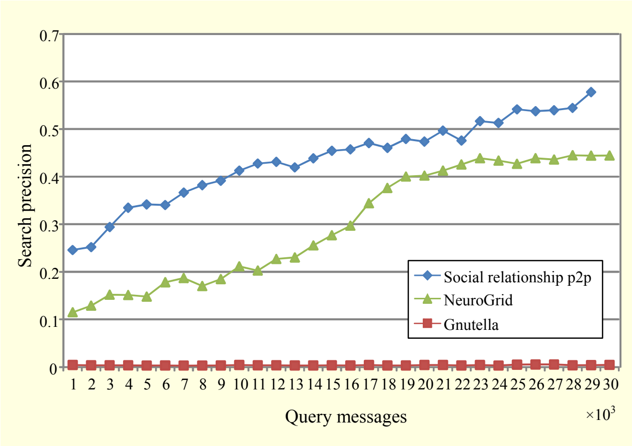

Fig. 5. Social relation weight mechanism.

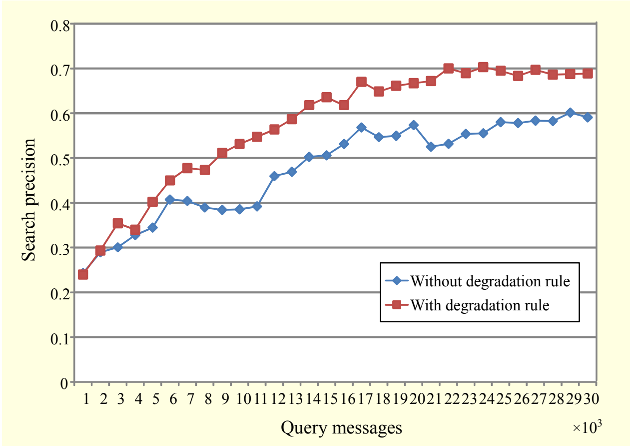

Fig. 6. Using degradation mechanism for search precision.

provides information. Neuogrid differs from Gnutella because each peer's knowledge index records the peers that respond successfully and then searches for these peers first in the next search. However, the keyword matching approach cannot identify similar keywords. For example, if a query peer wants to find the "Graph Theory" keyword, most peers find "Graph Theory" information. However, if the query peer wants to search for "Graph Summary" in the next search process, it can only find "Graph Theory" information peers in the knowledge index. Thus, it cannot extend search results, which decreases search precision. Therefore, the search precision of NeuroGrid remains at 0.1, initially. For 20,000 query messages, the search precision reaches 0.4. Thus, NeuroGrid has higher search precision than Gnutella. However, the proposed social relationship p2p network further improves the keyword match method in NeuroGrid by using the semantic similarity mechanism. The proposed approach can send a query message to peers with similar semantic search keywords and is not limited to knowledge index records. Therefore, social relationship p2p networks had a search precision of 0.25. At 20,000 query messages, the search precision of the proposed social relationship p2p network exceeds that of NeuroGrid.

To prove Hypothesis 5, this study provides a comparison of the search precision of the social relationship p2p networks, NeuroGrid, and Gnutella. This simulation tests the social relation weight mechanism of social relationship p2p networks. Figure 5 shows that Gnutella cannot use the concept of social networks, and it has a low search precision of 0, with no chance of improvement. NeuroGrid has higher search precision than Gnutella. This simulation uses 21,000 query messages as the reference point. Although NeuroGrid initially has a search precision of 0.1, this increases to 0.4 for 21,000 query messages. For 27,000 query messages, the search precision of NeuroGrid increases to 0.44. However, the search precision of the social relationship p2p network is still higher than that of NeuroGrid. Starting from an initial search precision of 0.25, the search precision of the social relation network increases to 0.5 for 21,000 query messages. For 27,000 query messages, the search precision increases to 0.55, revealing significant differences between search precision in the NeuroGrid and that in the social relationship p2p network. If a query peer finds an uninterested peer in the NeuroGrid, it records that peer in the knowledge index, and the search process may be able to find this peer in a subsequent search. In the same situation, the social relationship p2p network sorts peers based on social relation weight. Each query peer records useful peers and provides them with a positive point. Conversely, query peers do not record useless peers, and they assign them negative scores.

The higher search precision in the social relationship p2p network compared to that in Neurogrid is also due to the degradation mechanisms. To prove Hypothesis 6, this simulation retains only the degradation mechanism to determine whether the degradation mechanism affects search precision performance. Figure 6 shows that search precision of the social relationship p2p networks range from 0.2 to 0.3 with and without the degradation mechanism. However, as the number of query messages increases, the search precision of social relationship p2p networks without the degradation mechanism increases to 0.59. The social relationship p2p network with the degradation mechanism achieves a search precision of up to nearly 0.7. Using 21,000 query messages as the reference point, the social relationship p2p networks with and without the degradation mechanism exhibit a 0.15 (0.68 to 0.53) variation in search precision. These results show that the degradation mechanism for the social relationship p2p networks affects search precision significantly.

The social relationship p2p network has higher search precision than NeuroGrid because the social relationship p2p network includes the dynamic expansion mechanism. To prove

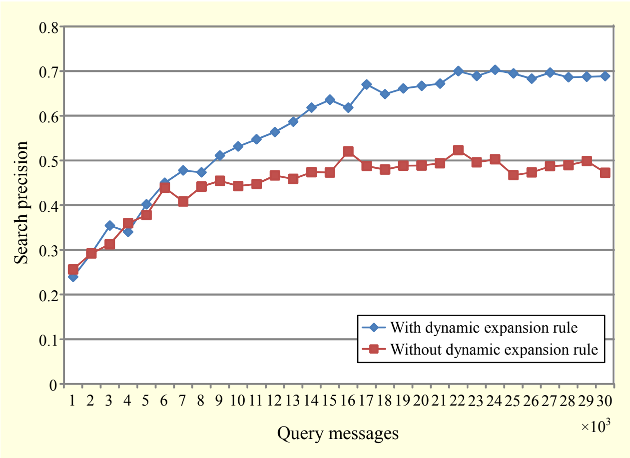

Fig. 7. Using dynamic expansion mechanism for search precision.

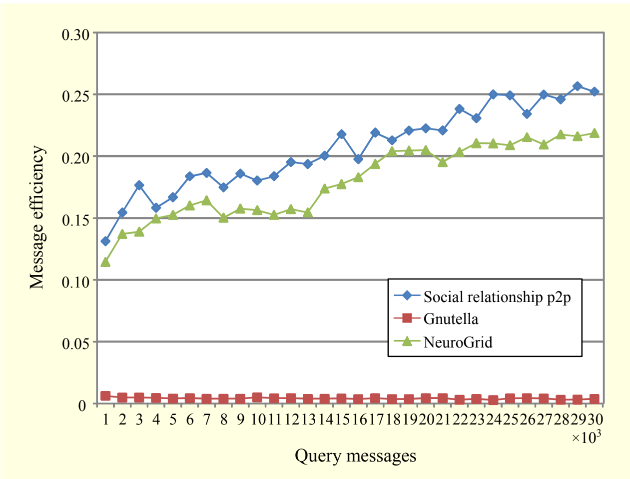

Fig. 8. Message efficiency.

Hypothesis 7, the social relation p2p only retains the dynamic expansion mechanism to determine if that affects search precision performance. Figure 7 shows that the search precision of the social relationship p2p networks range from 0.25 to 0.26 with and without the dynamic expansion mechanism. However, as the number of query messages increases, the search precision of the social relationship p2p networks without the dynamic expansion mechanism increases 0.5. This does not reach the level achieved by the same network with the dynamic expansion mechanism, which is 0.7. At a reference point of 23,000 query messages, this simulation shows a 0.2 (0.7 to 0.5) variation in search precision with and without the dynamic expansion mechanism. These results show that the dynamic expansion mechanism for the social relationship p2p networks affects search precision performance significantly.

To prove Hypothesis 8, Fig. 8 shows the contrast of message efficiency in the social relationship p2p networks, Gnutella,

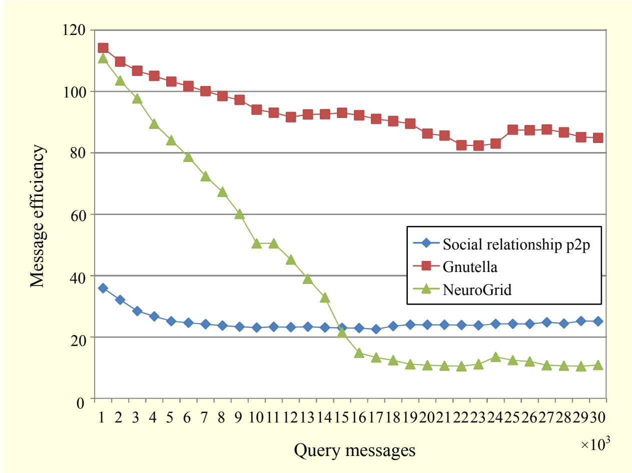

Fig. 9. Messages per query.

and NeuroGrid, indicating that social relationship p2p networks outperform NeuroGrid and Gnutella substantially. This simulation uses 21,000 query messages as the reference point. The message efficiency of Gnutella is approximately 0, and the message efficiency of NeuroGrid is 0.19, whereas the message efficiency of the proposed approach is 0.23. The statistical message efficiency of the social relationship p2p network increases by 0.25 and the statistical message efficiency of NeuroGrid increases by 0.23 as the number of query messages increases. The statistical message efficiency of Gnutella remains at 0. This means that the proposed approach increases search efficiency similarly to NeuroGrid.

To prove Hypothesis 9, the comparison illustrated in Fig. 9 shows the messages per query of the social relationship p2p networks, NeuroGrid, and Gnutella, indicating that the social relationship p2p network outperforms Gnutella and is close to NeuroGrid. We use 15,000 query messages as the reference point. The messages per query of Gnutella is 90, and the messages per query of NeuroGrid is 20, whereas the messages per query of the social relationship p2p network is nearly 20. Because the social relationship p2p network can accumulate partially useful information, the average messages per search query decreases to nearly 1. Furthermore, the query messages increase rapidly in the early stages of simulation. The average messages per query of the social relationship p2p network is only slightly smaller than that of NeuroGrid.

VI. Conclusion

This paper demonstrated that the characteristics of social networks can help improve search performance in p2p networks because of similarities between social networks and p2p networks. Based on Hebbian theory, we designed a peer selection method, called a social relationship p2p network, which identifies the social relation weight of peers. This social relation weight indicates the number of answers with which peers can correctly respond to a query peer. A strong social relation weight indicates that a peer can respond to the query, whereas a weak social relation weight indicates that a peer cannot respond. This approach stores the peer of strong social relation weight in the local knowledge index. A peer with a weak social relation weight is eventually dropped from the local knowledge index. Each peer who successfully connects with other peers gradually increases its social relation weight. Consequently, peers with high social relation weights can connect with each other easily.

We evaluated the search performance of social relationship p2p networks by using various performance metrics, including search precision, search recall, F-measure, message efficiency, and messages per query. The simulation experiments were performed in a near-realistic environment to collect convincing results for evaluation. This simulation environment included the network topology, degree of distribution, and peers joining and leaving the network. Simulation results and analysis indicated that the social relationship p2p network achieved enhanced search precision and located resources quickly. In addition, the simulation results showed that social relationship p2p networks outperformed Gnutella and NeuroGrid significantly in search performance. Extra simulations showed that the application of the social relationship p2p network increased message efficiency, compared to that of Gnutella and NeuroGrid.

In the future, we will implement social relationship p2p networks on certain open platforms for real-world system design. We will use valuable statistical instruments, which may demonstrate significant differences among NeuroGrid, Gnutella, and the proposed social relationship p2p network. We will investigate how to better adjust experimental parameters to enhance the search performance of the proposed social relationship p2p network.

References

[1] D.S. Milojicic et al., "Peer-to-Peer Computing," Technical Report HPL-2002-57R1, HP Laboratories, 2003.

[2] I. Stoica et al., "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," *Proc. ACM SIGCOMM*, 2001, pp. 149-160.

[3] S. Ratnasamy et al., "A Scalable Content-Addressable Network," *Proc. ACM SIGCOMM*, 2001, pp. 161-172.

[4] A. Rowstron and P. Druschel, "Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems," *Proc. IFIP/ACM Int. Conf. Distrib. Syst. Platforms*, 2001, pp. 329-350.

[5] M. Ripeanu, A. Iamnitchi, and I. Foster, "Mapping the Gnutella Network," *J. IEEE Internet Comput.*, vol. 6, no. 1, 2002, pp. 50-57.

[6] S. Rhea et al., "Handling Churn in a DHT," *Proc. USENIX ATC*, 2004, pp. 127-140.

[7] S.J.H. Yang and I.Y.L. Chen, "A Social Network-Based System for Supporting Interactive Collaboration in Knowledge Sharing over Peer-to-Peer Network," *Int. J. Human-Computer Studies*, vol. 66, no. 1, 2008, pp. 36-50.

[8] L. Liu, N. Antonpoulos, and S. Mackin, "Managing Peer-to-Peer Networks with Human Tactics in Social Interactions," *J. Supercomputing*, vol. 44, no. 3, 2008, pp. 217-236.

[9] S. Ciraci, I. Kopeoglu, and O. Ulusoy, "Reducing Query Overhead through Route Learning in Unstructured Peer-to-Peer Networks," *J. Netw. Computer Appl.*, vol. 32, no. 3, 2009, pp. 550-567.

[10] W.S. Yang and J.B. Dia, "Discovering Cohesive Subgroups from Social Networks for Targeted Advertising," *Expert Syst. Appl.*, vol. 34, no. 3, 2008, pp. 2029-2038.

[11] Y.M. Li and C.P. Kao, "TREPPS: A Trust-Based Recommender System for Peer Production Services," *Expert Syst. Appl.*, vol. 36, no. 2, 2009, pp. 3263-3277.

[12] J.C. Wang and C.C. Chiu, "Recommending Trusted Online Auction Sellers Using Social Network Analysis," *Expert Syst. Appl.*, vol. 34, no. 3, 2008, pp. 1666-1679.

[13] Y. Li, Z. Bandar, and D. McLean, "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources," *IEEE Trans. Knowledge Data Eng.*, vol. 15, no. 4, 2003, pp. 871-882.

[14] R.A. Ghanea-Hercock, F. Wang, and Y. Sun, "Self-Organizing and Adaptive Peer-to-Peer Network," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 36, no. 6, 2006, pp. 1230-1236.

[15] O.D. Hebb, *The Organization of Behavior: A Neuropsychological Theory*, New York: Wiley-Interscience, 1949.

[16] S. Joseph, "NeuroGrid: Semantically Routing queries in Peer-to-Peer Networks," *Proc. Int. Workshop P2P Comput.*, 2002, pp. 202-214.

[17] P. Haase et al., "Bibster-a Semantics-Based Bibliographic Peer-to-Peer System," *LNCS*, 2004, pp. 122-136.

[18] A. Löser, S. Staab, and C. Tempich, "Semantic Social Overlay Networks," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 1, 2007, pp. 5-14.

[19] C. Paui and M. Shepperd, "An Empirical Investigation into P2P File-Sharing User Behavior," *Proc. Americas Conf. Inf. Syst.*, Omaha, 2005.

[20] Y. Ren et al., "Explore the Small World Phenomena in Pure P2P Information Sharing System," *Proc. Int. Symp. CCGrid*, 2003, pp. 232-239.

[21] S. Saroiu, "A Measurement Study of Peer-to-Peer Files Sharing Systems," *Proc. Int. Conf. Multimedia Netw. Computing*, 2002.

Chester S.J. Huang received his BS in the Department of Management Information Systems from Leader University, Tainan, Taiwan and his MS in the Department of Management Information Systems from National Pingtung University of Science and Technology (NPUST), Pingtung, Taiwan. He is currently pursuing a PhD in the Department of Computer Science and Information Engineering at National Central University (NCU), Taoyuan, Taiwan. His research interests include peer trust in peer-to-peer networks and social relationships in peer-to-peer networks.

Stephen J.H. Yang is a distinguished professor of the Department of Computer Science and Information Engineering and the associate dean of Academic Affairs at the National Central University, Taiwan. Dr. Yang received his PhD in electrical engineering & computer science from the University of Illinois at Chicago, Chicago, IL, USA, in 1995. Dr. Yang has published over 60 journal papers and received the 2010 Outstanding Research Award from the National Science Council, Taiwan. His research interests include creative learning, 3D virtual worlds, app software, and cloud services. Dr. Yang is very active in academic services. He is currently the editor-in-chief of the *International Journal of Knowledge Management & E-Learning* and the associate editor of the *International Journal of Systems and Service-Oriented Engineering*. Dr. Yang also served as the program co-chair of APTEL 2011, ICCE 2010, TELearn 2009, ICCE 2009, IEEE SUTC2008, ICCE 2008, IEEE ISM2008, SDPS 2008, IEEE W2ME2007, IEEE CAUL2006, and IEEE MSE2003.

Addison Y.S. Su received his PhD in computer science and information engineering from the Department of Computer Science and Information Engineering at the National Central University (NCU), Taoyuan, Taiwan, in 2010. His current research interests include network packet analyzers, attack and defense of computers and smartphones, GPU distributed parallel computing technologies, and e-learning.