Abstract—Random walk (RW) has been widely used as a strategy for searching in peer-to-peer networks. The boom of social network applications introduces new impact to the classical algorithms on the Internet. In this paper, we model the random walk algorithm in peer-to-peer networks when social information is available. We define the social relationship between two nodes as the knowledge about the resources the other node possesses. We mathematically show that the social information can benefit the searching by extending the existing random walk search model.

I. INTRODUCTION

Peer-to-peer (P2P) networks attracted much attention in recent years. Random walk, noted for its locality, simplicity, low-overhead and robustness to structural changes, is a popular mechanism widely used in locating a resource or service efficiently in the peer-to-peer networks [1], [2], [3], [4]. However, peers in those models lack the capabilities to route queries efficiently due to limited information.

In order to enhance the search performance, social networks have been utilized to improve P2P applications through the cooperation between peers [5], [6], and the P2P structures based on social networks [7], [8].

In sociology, social relationships have been studied by sociologists for many years [9], [10]. With the increasing popularity of the Internet and the growing user participation, computer networks also exhibit similar social behaviors, as demonstrated in [11], [12]. Increasingly complex interpersonal on-line relationships have been formed with the assistance of various kinds of social network applications, Facebook [13], Myspace [14], etc.

Recently, researchers in the computer science and sociology are studying these on-line relationships [15], [12], [16], [17]. Large scale measurements are conducted to exploit the information behind large amount of users’ online activities. A recent project, based on an analysis of 25 billion chat sessions [12], reveals that people who chat with each other are more likely to share the same interests. Besides, the more frequently and the longer they talk, the closer the relationship is. Different from the conventional measurements of social networks (i.e., the degrees and links), [12], [16] improve the measurements with finer details, which open an era to measure the strength of on-line relationships.

As in one of the most widely accepted social network applications, Facebook [13], user’s recently updated information will be sent to his/her top friends, who are selected from the friends list based on the users’ preference. Generally speaking, users have closer relationships and more interactions with their top friends. Thus, we can expect users to have more knowledge/information about their top friends in terms of interests, communities they join, resources they share and so on. Among various social properties existing in a social relationship, we capture the most common and fundamental property, the knowledge about resources possessed by friends, in order to improve the search performance.

To the best of our knowledge, although it has been shown that search algorithms employing social information will yield better performance, quantitative analysis that quantifies the effects of social relationships on the performance improvement of random walk search algorithms in P2P networks has not been previously discussed.

By introducing the strength of social relationships, we model such social information and further apply it into the existing random walk search models [2], [3] in P2P networks. The strength of social relationships indicates the knowledge about the resources possessed by others, upon which we quantitatively analyze the effects of employing such social information and further propose a random walk search model when social information is available. We demonstrate the effectiveness of our proposed algorithm in terms of the search success rate in a given number of steps through both mathematical analysis and simulations.

The rest of this paper is organized as follows. Section II briefly introduces different kinds of searching algorithms in peer-to-peer networks and the recent progress in social networks studies. Detailed descriptions of the network model and assumptions are presented in Section III. Our mathematical results are provided in Section IV. Simulation results are presented in Section V.

II. RELATED WORK

There have been multiple studies on applying the social network information to enhance the performance of applications in P2P networks, such as [5], [7], [8].

The authors in [5] developed a model to increase cooperation among P2P peers based on a social network structure. They measured the strength of the relationship between a pair of peers by the amount of services they provided to and received from each other. Unlike [5], instead of considering...
the amount of mutual contribution, we focus on the knowledge about resources possessed by neighbors.

The authors in [7] described a P2P system using rules and concepts inspired by human behaviors and relationship dynamics. They presented a social links classification, Acquaintance-Link, Temporary Semantic-Link, Full Semantic-Link, which are similar to our approach. However, we further analyze the property of social relationships quantitatively with a detailed designed measurement of strength in our work.

In [8], authors presented a new social-like P2P algorithm for resource discovery by mimicking human interactions in social networks. They argued that no overheads was required to obtain any additional information of neighbor nodes. This scheme relies on the mechanism to measure the similarity of query’s interests. However, the forwarding strategy in our work relies on the knowledge about neighbors, which is featured in many on-line social network applications, and we are more interested in the quantitative analysis of the performance improvement with the presence of social information.

III. SYSTEM MODEL

A. Problem Description

Random walk is widely used for a node to locate a resource. When a node wants to get a resource that it does not possess, it sends a query to one of its neighboring nodes. When a node receives a query, it checks whether it has the requested resource. It replies to the query and terminates the search if it can answer the request; otherwise, it continues the process by forwarding the query to one of its neighboring nodes. To avoid the query circulating among the nodes indefinitely, the query also carries a time-to-live (TTL) counter to restrict the lifetime.

In the following discussion, the search efficiency, denoted as $E(T)$, is used as the performance benchmark of a random walk search model, defined as

$$E(T) = \frac{P_{\text{suc}}(T)}{T},$$

where $P_{\text{suc}}(T)$ denotes the probability that the search can succeed within $T$ hops. $E(T)$ better illustrates the performance of the search algorithms because it presents the comprehensive effects of the success rate [2], [3] and the number of hops. We analyze the efficiency of the proposed random walk search algorithm under different strength levels of social relationships.

B. Network Model

In the considered network model, there are $m$ resource items and $n$ peers in the network. Each peer possesses a different set of resource items. We use a matrix $D \in \{0, 1\}^{n \times m}$ as in [18] to describe the possessions of different peers, or mathematically as $D_{ij} = 1$ iff peer $i$ possesses resource item $j$.

The support set of item $j$ ($1 \leq j \leq m$) is a set of the peers that possess $j$, defined as $S_j = \{i | D_{ij} = 1\}$. Furthermore, the resource popularity of item $j$ is defined as $p_j = \frac{s_j}{n}$, where $s_j = |S_j|$ represents the number of peers that possess item $j$. In other words, it represents the number of replicas of item $j$ present in the network.

After establishing the basic network structure, we hereby introduce the neighbor-item matrix $F^{(i)}$ of the peer $i$. The matrix $F^{(i)}$ is a $f_i \times m$ matrix that contains the knowledge states with respect to peer $i$’s neighbors, where $f_i$ represents the number of peer $i$’s neighbors. The entry on the $l$th row and the $j$th column represents the knowledge state with regard to peer $i$’s $l$th neighbor ($1 \leq l \leq f_i$) and resource item $j$ ($1 \leq j \leq m$). Each knowledge state has three possible values 0, 1, Null, corresponding to different types of knowledge peer $i$ has. For instance, $F^{(i)}_{ij} = 1$ iff peer $i$ knows its $l$th neighbor has item $j$; $F^{(i)}_{ij} = 0$ iff peer $i$ knows its $l$th neighbor does not have item $j$; $F^{(i)}_{ij} = \text{Null}$ iff peer $i$ does not know whether its $l$th neighbor has item $j$ or not. Furthermore, we define peer $i$’s resource possession knowledge set $K_i \subseteq \{1, \ldots, m\}$ about its $l$th neighbor ($1 \leq l \leq f_i$), to be

$$K_i = \{ j | F^{(i)}_{lj} = 1 \text{ or } F^{(i)}_{lj} = 0 \}.$$

With the neighbor-item matrix $F^{(i)}$ and the possession knowledge set $K_i$, the strength of the relationship that peer $i$ has towards its $l$th neighbor is defined as

$$r_{il} = \frac{k_{il}}{m},$$

where $k_{il}$ denotes the number of items that peer $i$ clearly knows its $l$th neighbor has or has not, and $k_{il} = |K_i|$.

The strength of social relationship $r_{il}$ indicates the ratio of items that peer $i$ clearly knows its $l$th neighbor has or has not. We assume that the strength of relationship is asymmetric, which is reasonable in both real-life relationships and on-line relationships.

The following example illustrates how the strength of the social relationship about peer $i$ to its neighbors is calculated. Considering a case where peer $i$ has five neighbors, and peer $i$’s $l$th neighbor is denoted as $nb(i,l)$. There are four items in the network ($m = 4$). Table I shows the knowledge of the resources possessed by peer $i$’s neighbors. The knowledge set about $nb(i,1)$, $K_{i1} = \phi$. Thus, $r_{i1} = 0$. For the second neighbor $nb(i,2)$, $K_{i2} = \{1\}$. Thus, $r_{i2} = \frac{|K_{i2}|}{m} = 0.25$. $r_{il}$ can be computed using the similar method when $l = 3, 4, 5$.

<table>
<thead>
<tr>
<th>Neighbor</th>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>nb(1)</td>
<td>Null</td>
<td>Null</td>
<td>Null</td>
<td>Null</td>
<td>0</td>
</tr>
<tr>
<td>nb(2)</td>
<td>0</td>
<td>1</td>
<td>Null</td>
<td>Null</td>
<td>0.25</td>
</tr>
<tr>
<td>nb(3)</td>
<td>0</td>
<td>1</td>
<td>Null</td>
<td>Null</td>
<td>0.25</td>
</tr>
<tr>
<td>nb(4)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>Null</td>
<td>0.75</td>
</tr>
<tr>
<td>nb(5)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

C. Assumptions

Motivated by the classification of the real-life social relationships [10], we assume that there are two types of neighbors...
in our on-line relationship model: the strong tie and the weak tie. **Strong Tie Set** of peer \(i\), denoted as \(STS(i)\), is a set of neighbors that have strong tie relationships with peer \(i\), while **Weak Tie Set** of peer \(i\), \(WTS(i)\), is a set of neighbors that have weak tie relationships with peer \(i\). Let \(f_{i,s}\) and \(f_{i,w}\) be the sizes of \(STS(i)\) and \(WTS(i)\), respectively, with \(f_{i,s} + f_{i,w} = f_i\). This classification is featured in some social network applications. For example, the term *top friend* in [13] is used to distinguish different levels of acquaintances.

Note that users are likely to form links with “friends” on the platform provided by social network applications. However, with the expansion of friends, many users’ friend lists tend to include people they do not actually know or seldom communicate with. This brings forth weak tie relationships which are not strong enough to provide the resource information analyzed in this paper. Thus, in the following discussion, we assume that, for each peer \(i\) in the network, the knowledge states corresponding to the weak tie neighbors, in peer \(i\)’s neighbor-item matrix \(F^{(i)}\), are Null. This suggests that peer \(i\) is not clear about the resources storage of neighbors in \(WTS(i)\). Meanwhile, it is unlikely that peer \(i\) clearly knows all of its neighbors, thus we assume that the \(WTS(i)\) is not an empty set.

**IV. Analytical Results**

In this section, we first summarize the search efficiency of the random walk without social information. Then, we derive the analytical results in a new scenario where the social information is available.

**A. Pure Random walk without social information**

In the analysis below, we follow the results from [19], [20]: random walk achieves statistical properties similar to independent sampling for every reasonable network and network model. So the process of visiting each node can be regarded as an independent sampling from a space of nodes with uniform distribution. The query is forwarded to a neighbor randomly in such a way that it depends only on the overlay structure. Therefore, the current query is independent of the previous query and the previous selection of the querying peer. Each query is called a “random walker”. In this situation, the probability that the search will succeed at one step, denoted as \(q\), equals to the resource probability \(p\).

Based on the analysis about the success rate in [2], the search efficiency for a random walk search algorithm that deploys \(k\) independent random walkers can be represented as

\[
E(k \cdot T) = \frac{1 - (1 - q)^{k \cdot T}}{k \cdot T} = \frac{1 - (1 - p)^{k \cdot T}}{k \cdot T}
\]

where \(T\) is **TTL** for each random walker.

The expression of \(E(k \cdot T)\) shows that the search efficiency is mainly affected by the popularity of the requested resource, \(p\), for a given value of \(k \cdot T\).  

**B. Search efficiency of Random walk with social information**

Before analyzing the search efficiency when social information is involved, we first study the impact of the presence of the social knowledge.

By the assumption, each peer has a set of strong tie neighbors which may provide potential resource information. Suppose the size of \(STS(i)\) equals to \(f\), thus peer \(i\) has a sequence of social knowledge states of size \(f\) for each item. The pattern of social information, \(C_{ij} = < F_{ij}^{(1)}, F_{ij}^{(2)}, \ldots >\), is a sequence of knowledge states maintained by peer \(i\), with each element corresponding to a knowledge state associated with item \(j\) for a peer in \(STS(i)\).

\(C_{ij}\) does not include the knowledge states about neighbors from \(WTS(i)\), and it may take on different representations due to the combination of knowledge states in \(C_{ij}\), each of which is expected to have three possible values: \(0, 1, \text{Null}\).

As Fig.1 shows, \(STS(i) = \{\text{nb}(i,1), \text{nb}(i,2)\}\), and \(C_{ij} = < F_{ij}^{(1)}, F_{ij}^{(2)} >\). In this case, the social information pattern may be represented by one of the nine possible combinations: \(< 1, 0 >, < 0, 1 >, < 1, \text{Null} >, < \text{Null}, 1 >, < 1, 1 >, < 0, \text{Null} >, < \text{Null}, 0 >, < \text{Null}, \text{Null} >\).

To identify the impact of the social information with different patterns on the single step success probability, the patterns are classified into two cases. In the first case, \(C_{ij}\) contains state “1”. In the second case, \(C_{ij}\) does not contain state “1”. Since the social information in the two cases influences the success probability at each step in different ways, we will analyze the impact on success probability in each case, respectively.

**Case A:** The knowledge sequence \(C_{ij}\) includes at least one “1” state, which means that peer \(i\) knows at least one of its neighbors possesses the requested item \(j\). Refer to the example above, \(< 1, 0 >, < 0, 1 >, < 1, \text{Null} >, < \text{Null}, 1 >, < 1, 1 >, < 0, \text{Null} >, < \text{Null}, 0 >, < \text{Null}, \text{Null} >\) are classified into Case A. Peer \(i\) will forward the query to one of its neighbors, \(l\), where \(F_{ij}^l = 1\), at the next step. Thus, the probability that the query will succeed at this step is \(q_l = 1\).

**Case B:** The knowledge sequence \(C_{ij}\) does not include any “1” state. \(C_{ij}\) is a sequence consisting of two possible kinds of states: \(0, \text{Null}\). This suggests that peer \(i\) has no idea about who possesses item \(j\), though it has some information about who does not possess \(j\). Peer \(i\) randomly forwards the query.
to one of the neighbors, excluding the peers which do not possess the item based on peer $i$'s knowledge. The possible candidate may come from $WTS(i)$ or the peers in $STS(i)$ that satisfy $F_{ij} = Null$. $\theta$ is used to denote the number of state “0” in the sequence $C_{ij}$, ranging from 0 to the size of $STS(i)$. Case B is further grouped into $b_0, ..., b_{f_i}$, based on the value of $\theta$. Subcase $b_0$ corresponds to the group of social information patterns whose number of state “0” equals to $\theta$. The classification of the subcases in this example is listed as follows.

$$b_0 : \{< Null, Null >\},$$

$$b_1 : \{< 0, Null >, < Null, 0 >\},$$

$$b_2 : \{< 0, 0 >\}.$$  

In subcase $b_0$, there is neither state “1” nor state “0” in $C_{ij}$. Thus, peer $i$ will randomly select one from its four neighbors. Then, it forwards the query to the selected neighbor. In subcase $b_1$, there is one state “0” in $C_{ij}$. Thus, the scope of selection is restricted to three neighbors, excluding the one peer $i$ knows in which, either $nb(i,1)$ or $nb(i,2)$, item $j$ is not stored. In subcase $b_2$, peer $i$ will select one peer from either $nb(i,3)$ or $nb(i,4)$. Peer $i$ is aware of the fact that $\theta$ neighbors do not have the requested resource so they can be removed from the candidature of the random selection. Thus, the number of peers that may possess item $j$ is $n - \theta$. The probability that a peer which is randomly selected by peer $i$ possesses item $j$ is $\frac{s_j}{n - \theta}$. Therefore, the probability that the query will succeed at this step for subcase $b_0$ is $p_{b_0} = \frac{s_j}{n - \theta}$.

Let $p(A)$ be the probability of Case A. To simplify our analysis, we assume that the average size of the strong tie neighbor set for each peer equals to $f$ and the average strength of relationship is $r$. The popularity of the requested resource equals to $p$. Our analysis below also assumes that the success probability at each step is independent. Hence, the probability that $F_{ij} = 1$, $P(F_{ij} = 1) = r \cdot p$. Similarly, $P(F_{ij} = 0) = r(1 - p)$, $P(F_{ij} = Null) = 1 - r$. Based on the above analysis, the expression for $p(A)$ is given by

$$p(A) = 1 - (1 - rp)^f.$$  

Let $p(b_0)$ be the probability of the occurrence of subcase $b_0$. Hence, $p(b_0) = C_{f}^\theta \cdot [r(1 - p)]^\theta \cdot (1 - r)^{f - \theta}$.  

Thus, the probability that the search can succeed at one step is

$$q = p(A) \cdot q_a + \sum_{\theta=0}^{f} p(b_\theta) \cdot q_{b_\theta}.$$  

$$= 1 - (1 - r \cdot p)^f + \sum_{\theta=0}^{f} \left(C_{f}^\theta \cdot [r(1 - p)]^\theta \cdot (1 - r)^{f - \theta} \cdot \frac{s_j}{n - \theta}\right).$$

(2)

Social overlay networks demonstrate the characteristics of high user participation and a large number of friends. As a result, the expression of $q$ can be further simplified. The statistical properties in [16], $n = 240$ millions and the average number of peers’ friends equals to 49, implies that $\theta \ll n$. Thus, we have the following approximation

$$\frac{s_j}{n - \theta} = p \cdot \frac{n}{n - \theta} \approx p.$$  

(3)

Recall that the expression of the success rate of a random walk search, which employs $k$ random walkers with $TTL$ equally being $T$, is given in [2], [3] by

$$P_{Succ}(k \cdot T) = 1 - (1 - q)^{k \cdot T}$$  

(4)

where $q$ denotes the probability that the search can succeed in one step.

Substituting (3) into (2), then into (4), we derive the expression for the search efficiency of $k$-walker random walk algorithm

$$E(k \cdot T) = \frac{P_{Succ}(k \cdot T)}{k \cdot T} = \frac{1 - ((1 - r \cdot p)^f - (1 - r \cdot p)^f \cdot p)^{k \cdot T}}{k \cdot T} = \frac{1 - (1 - p)^{k \cdot T} \cdot (1 - r \cdot p)^{f \cdot k \cdot T}}{k \cdot T}.$$  

V. SIMULATION RESULTS

Our simulations are conducted on a graph with $10^4$ nodes, with each curve obtained from $10^4$ simulation runs. Degrees of nodes vary in the range of $[4, 200]$, with the mean value equals to 50. The size of Strong Tie Set for each peer $f$ is set to 4 which is close to the actual value of Gnutella networks as observed in [21]. The number of random walkers $k$ is 5 and the TTL for each random walker is $T = 10$.

In order to implement the social relationship with strength $r$ from peer $A$ towards peer $B$, a total of $r \cdot m$ resource items are randomly selected and peer $B$’s possession states about these items are passed to peer $A$. The corresponding knowledge states of peer $A$ are assigned to “0” or “1”, depending on whether peer $B$ has that item or not.

We simulated the random walk search algorithm with the social information and showed the effects of different resource popularity and different strengths of social relationships $r$. The popularity of resources adopted in the algorithm ranges from 0.0004 to 0.004. Fig.2(a) compares the simulated search success rate against the theoretical results from our derivation, under four different levels of social relationship $r = 0, 0.2, 0.5, 1$. As shown in Fig.2(a), the pair of curves with $r = 0$ correspond to the case discussed in [2], [3], while the other three pairs of curves demonstrate that the simulation results also fit well with the theoretical results even in the case where social information is available. We also observe that resources with higher popularity will have more significant improvements due to the introduction of social information.

On the other hand, Fig.2(b) shows that the success rate increases steadily with the increase of strength $r$. The curves
corresponding to the searches for more popular resources, i.e., greater $p$, rise more rapidly than those of smaller $p$ as $r$ increases.

VI. CONCLUSION

In this paper, we study the random walk search model in P2P networks with social relationship information. Specifically, we model on-line social relationships among peers in P2P networks. A new concept, named the strength of social relationship denoted as $r$, is applied to the traditional random walk search model discussed in [2], [3]. Then, we further exploited the derivation of the search efficiency of the random walk search algorithms. We demonstrated the effectiveness of applying social information through both mathematical analysis and extensive simulations. Both of them show similar performance. On the other hand, it can be shown that the mathematical analysis in this paper is a more general description of random walk search algorithms than that in the previous work [2], [3]. Specifically, their descriptions can be served as a special case in our derivation where $r = 0$.

In this paper, we assumed that all the peers have a fixed size Strong Tie Set and the topology of the network does not change during the whole process. Obviously, such static and synchronous scenarios are rare in real-world systems, but this preliminary study provides insights on the relationship between the search efficiency of the random walk algorithm and the social relationships among peers. We would like to explore the ideas of constructing and updating social relationships among peers in P2P networks in the future.

REFERENCES