

VegaNet: A Peer-to-Peer Overlay Network for Mobile Social Applications

Joonhyun Bae, Seunghun Lee, and Sangwook Kim

Electrical Engineering and Computer Science,
Kyungpook National University, Daegu, Korea
{jhbae, zrcn2sh, swkim}@cs.knu.ac.kr

Abstract—In the context of social relationships, cellular phones and their contacts construct a social network by nature. In this paper, we propose *VegaNet*, a peer-to-peer overlay network enhancing the performance and reliability of DHT routing using social links. The nodes in VegaNet are identified by the users' social identity, and it is structured by underlying DHT overlay exploiting social identities and social relationships. We present algorithms for handling churn and routing over the VegaNet in this paper, and the result of our experiments shows the effect of social links. The performance of lookup operation approximates to $O(\log N)$ with $O(\log N)$ social links at each node in our proposed model.

Keywords—DHT, Peer-to-Peer Overlay, Social Network.

I. INTRODUCTION

The regularities of structured overlay network provide the peer-to-peer systems with efficiency and scalability. However, the chronic headaches of peer-to-peer systems are primarily related with their basis on the anonymity [1]. The anonymity of peers bears the threats of misrouting and makes the whole network to be vulnerable [2]. Besides, many free-riders can join the network because no one knows who is joining the network. For these reason, as far as the reliability is concerned, the structured peer-to-peer systems are also fragile.

The motivation of this paper is to solve these problems by introducing the identified and structured peer-to-peer overlay network model. Hence, in this paper, we propose *VegaNet*, an identified peer-to-peer overlay network which is consists of social identities and social relationships. Mobile subscribers have their unique identities, MIN (Mobile Identity Number). When mobile users share their social objects with their unique identifiers, they are uniformly distributed into the same identifier space as the user identifiers. Social identities and relationships are nodes and links in the social network respectively [4-12, 18].

With the favor of previous DHT-based algorithms [13-16], all the user and data identities are evenly distributed into a common identifier space by the cryptographic hash function like SHA-1. The underlying assumption of our approach is that a peer believes another peer according to the social distance between them. Hence, we introduce two routing strategies in this paper: Social-First-Routing (SFR) and Local-First-Routing (LFR). To understand the effects of social routing in DHT-based structured overlay network, we present the results of our

theoretical analysis and simulation studies on the power-law distributed network [17].

This paper is organized as follows. In Section 2, we review related work. In Section 3, we describe the VegaNet model, as an identified peer-to-peer overlay network utilizing the existing social network. In Section 4, we present the algorithms for handling churn and sharing content. In chapter 5, we evaluate our proposed model by simulation studies. Finally, Section 6 concludes this paper.

II. RELATED WORK

A social network may be thought as a network where nodes are humans and links are the social relationships. Watts et al. [3] have founded that social networks have the property of being searchable. Ordinary people are capable of directing messages though their network of acquaintances to reach a specific but distant target person in only a few steps. Their first contention about social network is an individual in social networks are endowed not only with network ties, but also identities.

Constructing a peer-to-peer system over a social network is a challenging problem. There are many previous researches related with this challenge [4-12,18]. Fast et al. [4] and Pouwelse et al. [5] proposed a structured peer-to-peer overlay network model based on the social network. Marti et al.[6] also proposed SPROUT, a DHT routing algorithm that significantly increases the probability of successful routing by using social links. Matuszewski et al. [7] also proposed a social DHT architecture that mitigates the problems of P2P networks in the mobile environment. Bryan et al. [8] proposed SOSIMPLE, a fully decentralized, peer-to-peer, standard-based approach to communications. It combines the SIP/SIMPLE standards with the self-organizing properties of a DHT-based P2P mechanism. Carchiolo et al. [9] proposes a novel model, called *PROSA*, for growing and evolution of a peer-to-peer network inspired by social networks dynamics. They think that emerging structural and topological properties of natural networks, as said before, could be successfully exploited to improve communication among peers and to efficiently retrieve shared resources.

In the application areas, Tomiyasu et al. [10,11] also designed and implemented a query propagation mechanism to realize a social network composed by cellular phones using e-mail function. Kellerer et al. [12] provide several problems to

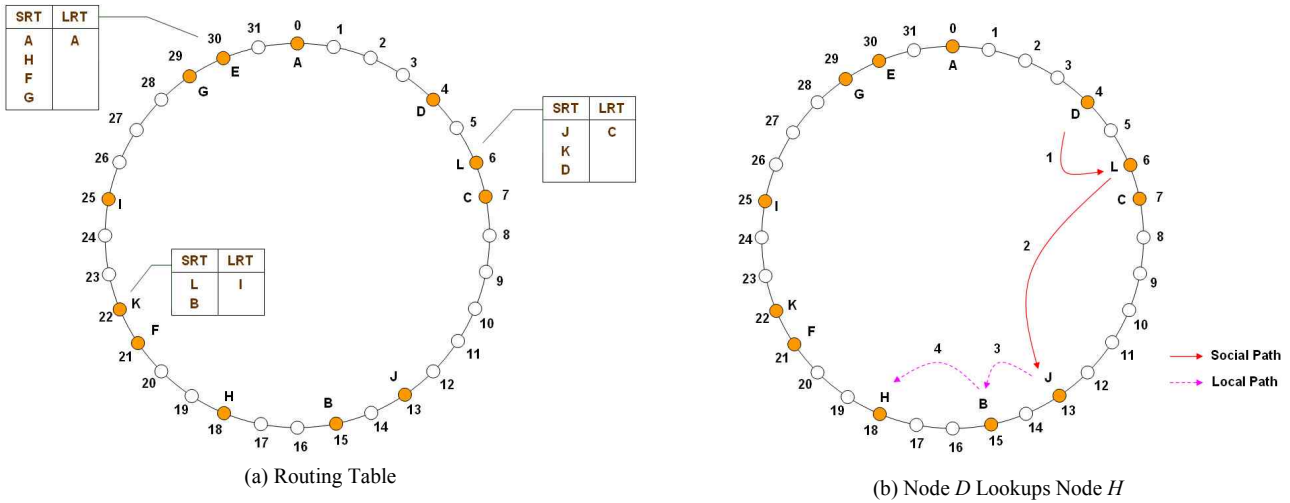


Figure 1. An example structure of VegaNet

realize P2P concepts in mobile networks. They summarize that it is especially the heterogeneity of the mobile environment that constitutes the challenges. They proposed a two-layer architecture consisting of different building blocks for an operator-grade P2P service platform. Bae et al. [18] proposed a mobile peer-to-peer query for a decentralized search.

III. SYSTEM MODEL

In this section, we describe the VegaNet, an identified peer-to-peer overlay network utilizing the existing social network.

A. Networking Entities

We can identify the following three major types of entities of social network that enables sharing resources among people: social identity, social relationship, and social object.

- **Social Identity:** In a social network, social relationships are represented by nodes and ties. A node is an individual within the social network, and a tie is the relationship between two nodes. In our approach, a mobile phone represents the social identity of its owner. Hence, it is apparent that the IMSI (International Mobile Subscriber Identity) is a good candidate for the identifier in our overlay network. However, IMSI provisioned by cellular communication protocols is voice-oriented. Instead of it, we can utilize Instant Messenger IDs or SIP URIs because of data communication requirement of content sharing.
- **Social Relationship:** In the context of social relationships, contacts in mobile phones imply that a directed connection from the owner of contacts to that of indicating mobile phones. With the same reason as above, in our approach, a contact itself does not represent the social links between two users. To enable query routing and content exchange, two linked nodes can communicate mutually with social identities. Instant Messaging protocols or SIP over 3G networks can suitable for this requirement. For example, if an IM user has a roster list of his friends, they can be the

social links because he has a notion of his friends and can communicate mutually.

- **Social Object:** Usually, unlike PC-based environment, contents in mobile devices are self-created and private, e.g., photos, videos, voices, and so on. Only the content with which the owner explicitly wants to share should be public to the community. We regard this kind of content as social object. In our approach, a social object also has a unique identifier. It can be represented in the same manner of Uniform Resource Identifier using the social identifier instead of IP-address or domain name.

B. VegaNet Model

VegaNet consists of three kinds of sets: a set of identities P , a set of links L , and a set of objects D .

$$VegaNet = (P, L, D) \quad (1)$$

An identifier space I is a set of all the keys which can be produced by an m -bit consistent hash function $hash$.

$$I = \{k \mid 0 \leq k \leq 2^m - 1, k \text{ is an integer}\} \quad (2)$$

A set of peers' keys P and data's keys D are the collections of keys which are converted with the peers' and data's identifiers.

$$P = \{k \mid k \in I, k = hash(userid),\} \quad (3)$$

$$D = \{k \mid k \in I, k = hash(dataid)\} \quad (4)$$

C. Social and Local Links

The links between two nodes in VegaNet are established by either social or local relationships. A *social link* can be made by the user with the notion of real world's social relationship like a buddy in a messenger or contacts in phonebook information. Hence, if there is social link between two nodes, we assume that two peers mutually trust each other and willingly communicate and share messages and information.

A *local link* can be established by the closeness in the identifier space of two nodes. It is maintained to design an

```

ALGORITHM: Lookup
node.lookupSuccessor(k)
  if (k in (node.key, next.key))
    return node;
  follow ← node.closestInSocialLinks(k)
  if (follow = null)
    follow ← node.closestInLocalLinks(k)
  return follow.lookupSuccessor(k);

node.closestInSocialLinks(k)
  for i=|SRT|-1 downto 0
    if (k < SRT(node.key)[i].key)
      return SRT(node.key)[i]
  return null;

node.closestInLocalLinks(k)
  for i=|LRT|-1 downto 0
    if (k < LRT(node.key)[i].key)
      return LRT(node.key)[i]
  return null;

```

Figure 2. Lookup operations with LRT and SRT.

```

ALGORITHM: Churn
node.join(bootNode)
  node.pred ← null
  node.succ ← bootNode.lookupSuccessor(node)

node.stabilize()
  n ← node.pred
  if n.key in (node.key, succ.key)
    succ ← n
  succ.notifyPred(node)
  node.updateLRT()
  node.updateSRT()

node.notifyPred(n)
  if (pred is null or n in (pred.key, node.key))
    pred = node;

node.leave()
  succ.notifyPred(pred)
  pred.stabilize()

```

Figure 3. Join and Leave operations.

efficient routing protocol with the advantage of structured overlay network. When a message is routed to a destination peer, the approximation should be guaranteed by the notion of locality about the network topology, namely, a peer should have the better knowledge about the peers which is closer in the identifier space.

A set of vectors L , which represents a link between two peers, is the union of a set of social links L_S and that of local links L_L . i.e., $L = L_S \cup L_L$.

$$L_S = \{(i, j) \mid i \in P, j \in P, sd(i, j) = 1\} \quad (5)$$

$$L_L = \{(i, j) \mid i \in P, j \in P, sd(i, j) < 1\} \quad (6)$$

IV. ALGORITHMS

In this section, we provide algorithms for looking up a node, handling churn, and sharing objects.

A. Lookup

The basic operation of structured peer-to-peer overlay networks is to find a node which is closest to a given arbitrary

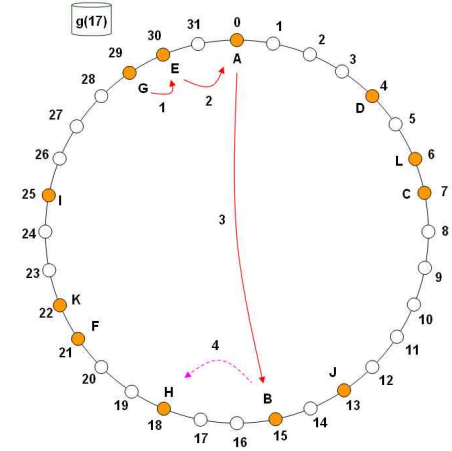


Figure 3. Sharing an object g in VegaNet.

key k . It searches a closest node to k in SRT (Social Routing Table) first, and then, scans LRT (Local Routing Table) to find a closest node to k , until the successor of k is encountered. All the records in the routing tables are stored in the order sorted by the offset from the node of the owner of the tables.

For example, in Figure 1(a), the SRT of node E is an array of nodes in $SRT(30) = \{0, 18, 21, 29\}$, where $SRT(30)[0] = 0$, $SRT(30)[1] = 18$, $SRT(30)[2] = 21$, and $SRT(30)[3] = 29$. In Figure 1 (b), node D looks up node H . When a node D initiates a query looking up a node H , it finds a node L in its SRT which is closer to node H . Also, L find node J to forward a query. Since node J doesn't has any social links closer to H , J forwards query to B and finally, node B finds node H and lookup operation is terminated. Figure 2 shows the pseudo-codes for lookup operations.

When a node n wants to lookup a node which is the successor of a key k , algorithm **Lookup** uses a greedy method to approximate clockwise to the destination. If n does not have a direct connection to the destination node, there is a node n' in SRT or LRT, which is closer to the destination than node n . When n forwards a message to n' , the distance to the destination decreases. If the distance is smaller than $O(\log_2 N)$, as the size of LRT is $O(\log_2 N)$, the destination can be reached by one more hopping, i.e. $O(1)$. If the distance to the destination is bigger than the size of LRT, a closest node in SRT or LRT can be used to advance. Suppose that a forward to n' approximates half the distance from n to the destination. Then, let h be the number of hops necessary to reach within the range of LRT . The size of subset of nodes $N/2^h = 1$, namely, $h = \log_2 N$. As the size of N 's subset is diminished, the probability of approximation is reduced accordingly. Only a local links can be reliable to reach the following node in a small range of identifier space. In spite of this conditions, if there are $O(\log_2 N)$ social links per nodes which can divide the range of identifier space into the half, a node can reach to the successor of an arbitrary key with $O(\log_2 N)$ hops.

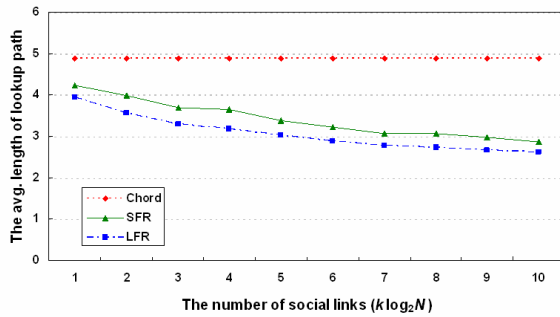
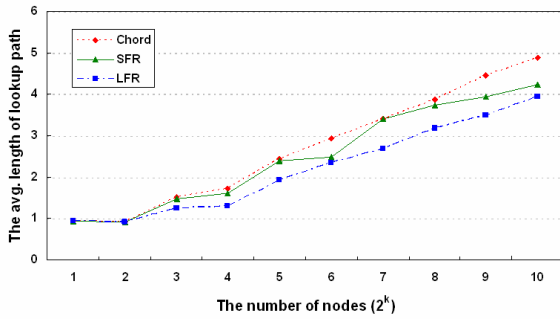


Figure 5. The cost of lookup operation

B. Join and Leave

A node joining the network should have the information about the location of the network to be placed. To join the proposed overlay network, a node is invited by a participating node. On the contrary, a node can join the network by the help of bootstrap node which is already participating in the network.

Figure 3 shows the algorithms for join and leave operations. A first node to create an overlay network does not need to have a bootstrap node. A node which is joining the existing network should know the predecessor and successor of its own key in the identifier space. With the help of boot node, a new node can have the reference point to previous and next node in the identifier space. The previous and next nodes also should be notified by the joining node to change the local links. Leaving the network is as simple as a deletion from a doubly linked list. Finding a successor of a key is described in the previous sections. Hence, the join operation returns the successor of a key, to give a node with the local links with the previous and next adjacent nodes.

C. Sharing Contents

The DHT abstraction provides the same functionality as a traditional hash table, by storing the mapping between a key and a value. This interface implements a simple *put* and *get* functionality, where the value is always stored at the live overlay node(s) to which the key is mapped by the lookup operation. Values can be objects of any type. *Put* operation inserts a pair of a key and a value into a database. It returns no value, which is not intuitive but is less restrictive. Keys and values are represented in strings. *Get* operation fetches pairs by a key and returns a list of pairs whose keys are the same as the

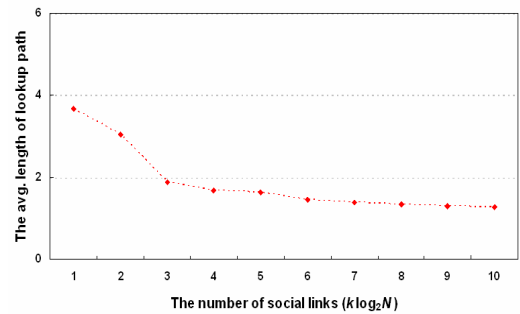


Figure 6. The rating of path lengths

specified key. The list is returned asynchronously, which is less restrictive. Note that some implementations might return only one pair.

Figure 4 shows that a social object g is shared with *put* operation. Since the hash value of g is 17, the successor of g is node $H = succ(17)$. After node G lookups H , it replicates the identifier and associated link and indexing tags to node H . It will be used for searching and getting g by the other users.

V. EVALUATION

In this section, we evaluate the effect of social links on the performance of routing in our proposed model.

A. Experimental Setup

To test the routing performance in *VegaNet*, we decided to compare our algorithm with original Chord [13] algorithm. Hence, we use 160-bit *SHA-1* algorithm as a consistent hash function.

We developed two different routing strategies on top of original Chord algorithm: 1) *SFR* and 2) *LFR*. *SFR* (*Social First Routing*) strategy chooses the next node from SRT (Social Routing Table) first. A node from LRT (Local Routing Table) is chosen only if there is no closest preceding node in SRT. In the other strategy, *LFR* (*Local First Routing*) selects the closest next node from SRT and LRT. Hence, we expect that LFR strategy is faster than SFR, because of the approximation of social routing is guaranteed by the selection of closer node to the destination.

We simulated these strategies in a network with $N=2^k$ nodes. We varied k from 1 to 10 and conducted an experiment in each network. The number of data used in lookup count is 100×2^k data generated by a hash value of a string "1" to "D($=100 \times 2^k$).". The number of social links per node is $\log_2 N$, and they are also generated by a random number with preferential attachment [17].

B. Evaluation Result

Figure 5(a) shows the result of lookup cost of each operation. It presents the average path length of each strategy. As expected, LFR shows better performance than original Chord in all the experiment results. However, SFR would be sometimes worse than Chord. It is because the selection in SRT is not guaranteed for the approximation to the target node

as in the *finger table* of Chord [13] system. Though, we see that the asymptotic average cost of SFR is also $O(\log N)$.

Unlike the original Chord algorithm, in *VegaNet*, as the number of social links increases, the performance of social lookup goes better. To understand this property, we perform an experiment of increasing the number of social links for each node from $O(\log N)$ to $10 \times O(\log N)$ in a network with 2^{10} number of nodes. Figure 5(b) shows the result of this experiment when we conduct the same lookups as in the previous experiment.

The advantage of our social routing algorithms over other DHTs is that there are reliable connections in routing paths. Because the social routing tables are constructed by the real world human relationships, we consider the selection of next node in SRT as a highly reliable routing. To verify this hypothesis, we conduct an experiment to estimate the path's reliability using a trust model as the following:

For given two nodes i and j , the trust factor $T(i, j)$ is the inverse of the social distance $sd(i, j)$.

$$T(i, j) = \frac{1}{sd(i, j)} \quad (7)$$

With this simple and deterministic trust function, for the purpose of social routing, *VegaNet* maintains the reputation of nodes in the network. Our strategy of estimating a node's reputation is affected by two factors: 1) the number of trusted neighbors and 2) the reputation of the neighbors. *VegaNet* present the equation to estimate a node's reputation, $R(i)$, for a given node i , with the number of neighbors l , as follows:

For a node i , the reputation of this node $R(i)$ is the sum of the number of social links and the rooted value of neighbor's reputation.

$$R(i) = \sum_{k=1}^l T(i, k) + \sqrt{\sum_{k=1}^l R(k)} = l + \sqrt{\sum_{k=1}^l R(k)} \quad (8)$$

We issue 10×2^{10} lookup in a network with 2^{10} number of nodes, and estimate the average rating of path according to the number of hops. Figure 6 shows the result of the average path rate for each path length from 1 to 10. As expected, the more the social paths are, the higher the trust of average lookup path.

VI. CONCLUSION

In this paper, we proposed a novel peer-to-peer overlay network model, *VegaNet*, inspired by social relationships in mobile environment. With this model, we investigated the performance of peer-to-peer social routing algorithms, with two different strategies, LFR and SFR.

The contribution of this paper is to introduce the design of structured overlay network routing algorithms to be efficient and scalable in the mobile environment. Although mobile networks and devices are evolving, still the communication costs and device capacities are relatively limited. In our approach, the cost of the lookup operation to find a node with a key is $O(\log_2 N)$ in N -node overlay network.

In the future work, we will advance to an identified peer-to-peer network that includes the functions of locating and retrieving a social object on the overlay network model which is proposed in this paper. Also, to make it more fascinating, we will add an incentive mechanism with trust and reputation model defined in this paper.

REFERENCES

- [1] E. Adar and B. A. Huberman, "Free Riding on Gnutella," *First Monday*, vol. 5, no. 10, Oct. 2000.
- [2] M. Castro, P. Drushel, P. Ganesh, A. Rowstron, and D. Wallach, "Secure Routing for Structured Peer-to-Peer Overlay Networks," *Proc. of OSDI'02*, Boston, Massachusetts, USA, Dec. 2002.
- [3] D. J. Watts, P. S. Dobbs, and M. E. J. Newman, "Identity and Search in Social Networks," *Science*, vol. 296, pp. 1302-1305, 2002.
- [4] A. Fast, D. Jensen, and B. N. Levine, "Creating Social Networks to Improve Peer-to-Peer Networking," *Proc. of ACM SIGKDD'05*, Chicago, Illinois, USA, 2005.
- [5] V. Carchiolo, M. Malgeri, G. Mangioni, and V. Nicosia, "Emerging Structures of P2P Networks Induced by Social Relationships," *Computer Communications*, Vol. 31, Issue 3, pp. 620-628, Feb. 2008.
- [6] J. A. Pouwelse, P. Garbacki, J. Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. van Steen, and J. J. Sips, "TRIBLER: A Social-based Peer-to-Peer System," *Concurrency Computat.: Pract. Exper.* Vol. 20, pp. 127-138, 2008.
- [7] H. Tomiyasu, T. Maekawa, T. Hara, and S. Nishio: "Social Network Applications using Cellular Phones with E-mail Function," *Proc. of Int'l Special Workshop on Databases for Next Generation Researchers (SWOD 2005)*, pp. 136-139, 2005.
- [8] H. Tomiyasu, T. Maekawa, T. Hara, and S. Nishio: "Profile-based Query Routing in a Mobile Social Network," *Proc. of the 7th Int'l Conference on Mobile Data Management (MDM 2006)*, p. 105, 2006.
- [9] D. A. Bryan, B. B. Lowekamp, and C. Jennings, "SOSIMPLE: A Serverless, Standards-based, P2P SIP Communication System," *Proc. of AAA-IDEA'05*, pp. 42-49, 2005.
- [10] W. Kellerer, Z. Despotovic, M. Michel, Q. Hofstätter, and S. Zöls "Toward a Mobile Peer-to-Peer Service Platform," *Proc. of SAINTW'07*, p. 2, Jan. 2007.
- [11] S. Marti, P. Ganesan, and H. Garcia-Molina, "SPROUT: P2P Routing with Social Networks," *Proc. of P2P&DB'04*, Mar. 2004.
- [12] M. Matuszewski, and M. A. Garcia-Martin, "Social Distributed Hash Table," *Proc. of WCNC'07*, pp. 2812-2818, Mar. 2007.
- [13] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications," *IEEE/ACM Transactions on Networking*, vol. 11, no. 1, 2003.
- [14] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A Scalable Content Addressable Network," *Proc. of ACM SIGCOMM'01*, San Diego, USA, Aug. 2001.
- [15] A. Rowstron and P. Drushel, "Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-Peer Systems," *Proc. of IFIP/ACM International Conference on Distributed Systems Platforms (Middleware'01)*, Heidelberg, Germany, Nov. 2001.
- [16] B. Zhao, J. Kubiatowicz, and A. Joseph, "Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing," *Technical Report UCB/CSD-01-1141*, Computer Science Division, Univ. of California, Berkeley, Apr. 2001.
- [17] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in Power-Law Networks," *Physical Review E.*, vol. 64, no. 4, 046135, 2001.
- [18] J. Bae, S. Lee, E. Jung, and S. Kim, "A Mobile Peer-to-Peer Query in a Social Network," *Proc. Of ALPIT'08*, pp. 450-453, 2008.