Peer-to-Peer Clustering for Semantic Overlay Network Generation

Michalis Vazirgiannis[1], Kjetil Nørvåg[2], and Christos Doulkeridis[1]

[1] Department of Informatics
Athens University of Economics and Business (AUEB)
Athens, Greece
{mvazirg,cdoulk}@aueb.gr
[2] Department of Computer Science
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway
Kjetil.Norvag@idi.ntnu.no

Abstract. The peer-to-peer (P2P) paradigm presents an attractive solution for applications that require scalability, fault-tolerance and autonomy. P2P systems in their basic unstructured form suffer high costs when it comes to efficiently locating content, mainly due to the lack of global knowledge. It is therefore crucial to organize content in an unsupervised way by creating groups of peers with similar content, in order to support efficient search mechanisms. In this paper, we discuss the need for content organization in unstructured P2P networks and present the requirements that must be fulfilled by any approach. We propose P2P clustering as a potential solution to Semantic Overlay Network (SON) generation for organizing P2P networks, and we present our unsupervised approach for decentralized SON creation towards this end.

1 Introduction

The peer-to-peer (P2P) paradigm presents an attractive solution for several applications that require scalability, fault-tolerance and autonomy. Numerous P2P applications, with file-sharing being the most prominent, have already proved their merit and are extensively used. Other, more ambitious approaches have been recently proposed in the literature, for example P2P web search [22].

P2P systems are classified in *unstructured* and *structured* systems. Unstructured P2P systems do not impose any constraints to the participating peers, other than establishing a limited number of neighbors for each peer. The basic search mechanisms are flooding [14] and its variants, like directed or normalized flooding [13]. This pure P2P architecture has several advantages, like resilience to failures and peer autonomy, but presents some drawbacks as well, such as high search costs with no guarantees of locating content. In order to solve some of these problems, structured P2P systems have been proposed [25, 27, 29, 34]. These systems are based on distributed hash tables (DHTs) that can support efficient key-based lookups, with predictable logarithmic cost. However, structured P2P systems impose restrictions on data or index placement, and in general they are less resilient to failures. Since our main interest lies in scalable,

self-organizing and fault-tolerant systems, we focus in this paper on unstructured P2P architectures.

To improve the efficiency and quality of search in unstructured P2P systems, Semantic Overlay Networks (SONs) [8] have been proposed. The basic idea behind SONs is to group together peers that contain similar contents, so that at search time, queries can be forwarded to only those peers containing content that satisfies the constraints of the query context, thus reducing the communication cost of the query and increasing result quality. One of the problems of SONs is the actual construction of these overlays in a P2P manner, assuming the lack of knowledge of both global content and network topology. In a P2P architecture each peer is initially aware only of its neighbors and their content. Thus finding other peers with similar contents to form a SON, a procedure that we call peer clustering, becomes a tedious problem.

The main topic of this paper is the SON creation, and this paper also motivates the use of SONs to facilitate search in unstructured P2P networks and it captures the requirements for SON generation. Further, the research results of our method for *distributed* and *decentralized* SON construction, called DESENT, are presented, providing an efficient mechanism for search in unstructured P2P networks.

The rest of this paper is organized as follows: in Section 2, an extensive overview of the related research is presented, namely semantic overlay networks and P2P clustering. In Section 3, we state the requirements for SON generation in large-scale P2P networks. Our approach to distributed and decentralized SON generation is described in Section 4, and finally, in Section 5, we summarize the conclusions of our work and identify future research directions related to SON generation.

2 Related Work

Performance and scalability problems in unstructured P2P networks, like Gnutella [14] and Freenet [5], are well-known [20, 26] and approaches that try to rectify the search performance have been previously proposed [3, 7, 13]. Another study [11], has pointed the problem of free riding in P2P networks and in particular for Gnutella the authors reached the conclusions that: a) nearly 70% of Gnutella users shared no files and b) nearly 50% of all responses are returned by the top 1% of sharing hosts. All these results bring out the problems of search using unstructured P2P networks in their basic form and motivate the development of more efficient methods.

In Gia [3] the combination of several techniques are proposed to effectively improve searches: topology adaptation, hot-spot avoidance, one-hop replication and biased random walks. Gkantsidis *et al.* [13] study hybrid search schemes for unstructured P2P networks, including normalized flooding and random walks with shallow flooding. In [33], broadcast policies are proposed for improving search and three families of techniques are proposed: a) iterative deepening, b) directed breadth-first search, and c) local indices. Another approach based on directed searches that improves on blind flooding is presented Crespo and Garcia-Molina [7]. Each peer maintains local *routing indices* that help choosing the most promising directions for neighbor selection. A similar approach utilizing taxonomy-based routing indices is proposed in [23].

The concept of Semantic Overlay Networks (SONs) is introduced in the P2P literature in [8]. The authors recognize the following challenges when building SONs: a) classification of queries and peers, b) level of granularity for each classification, c) the condition(s) that should be satisfied for a peer to join a cluster, and d) which clusters to use for answering a query. However they do not provide any other algorithm for searching than flooding. In order to be useful in a large system, unsupervised and decentralized creation of SONs is necessary, as well as efficient routing of queries to the appropriate SON(s).

Lately several approaches have been proposed for using SONs to improve search in P2P systems, partly addressing some of the aforementioned issues. While several papers refer to clustering and semantic cluster creation, they usually apply classification to generate groups of documents, and subsequently peers. In fact this is an important feature that discerns completely unsupervised methods from methods that rely on some background knowledge. Liu *et al.* [16] create groups of peers that are topologically near each other, which they call clusters, and within each cluster specific peers are assigned a set of predefined categories. Cohen *et al.* [6] propose *associative overlays*, which are formed by peers that have provided answers to previous queries. Also they use *possession rule overlays*, formed by having peers maintaining a list of other peers, with which they index the same item. Parreira *et al.* [22] propose SONs for P2P web search. Their method is based on rearranging the connections between peers to link *friend* peers to each other. A similar approach is followed in [4], where the notion of *acquaintances* is proposed. In [32], a P2P architecture where nodes are logically organized into a fixed number of clusters is presented. The main focus of the paper is fairness with respect to the load of individual nodes. The allocation of documents to clusters is done by classification, so it is not unsupervised. In [18], clustering policies are proposed to generate semantic clusters in super-peer networks. Particular emphasis is put on managing heterogeneous data schemes. Clustering peers based on schemas is also studied in [21], while in [1], GridVine is presented, which is about SONs based on schema mappings. An approach for distributed document clustering based on k-means is presented in [12]. In [24], an approach for connectivity-based clustering that creates topological clusters, which can be used as starting points for flooding, is presented. Tempich *et al.* [31] present an approach where peers join overlay networks based on observations about queries that were successfully answered by other peers. This information is later used to direct searches only to peers that are likely to answer the query.

Hierarchical SONs have also been proposed in the literature, mainly because of their efficiency. In [15], the authors present HSPIR, an approach to index documents in the network hierarchically, in order to support efficient distributed information retrieval. HSPIR uses a structured P2P network (CAN [25]) to organize the nodes, while support for semantics is guaranteed by the use of Latent Semantic Indexing (LSI). A different focus is given in [28], where hierarchical summary indices for content search are created, following a super-peer approach. Taxonomy-based overlays are studied in [17], where existing classification of peers into taxonomy concepts is exploited to improve query routing.

Several other approaches for SON creation over structured P2P systems have also been proposed [2, 9, 15, 19, 30]. Since the focus of this paper is in unstructured P2P

systems, we confine to merely mention these approaches, but we will not describe their functionality in more detail.

3 Requirements for SON Generation

Although several P2P research papers adopt the use of semantic overlay networks, they also adopt a set of assumptions that more or less relax the basic constraints imposed by the P2P paradigm. In this section, we go a few steps back, as we identify with the benefit of hindsight from existing approaches the basic requirements for SON generation in a dynamic P2P environment: unsupervised approaches for P2P clustering, scalability, self-organization, autonomy and decentralization. We do not consider this list to be complete, we rather see it as a basic set of requirements that should be enforced, as they increase the value and benefit of any novel SON generation algorithm.

3.1 Unsupervised Approaches for P2P Clustering

P2P networks in their initial, visionary form are systems characterized by lack of global knowledge. Instead, only local knowledge of content and topology can be safely assumed. However, several approaches make assumptions about the existence of background knowledge, in order to facilitate SON generation. A challenge is to organize the P2P network, assuming minimal pre-existing knowledge.

In principle, clustering algorithms are particularly suitable for SON generation, because they constitute an unsupervised approach. Apart from the input parameters that some clustering algorithms need to execute, P2P clustering based on peers' contents assumes no further pre-existing knowledge. Nevertheless, several existing SON generation approaches rely on classification to group similar peers. The difference is that some background knowledge is assumed, usually in the form of a pre-defined taxonomy or as an already existing labeling scheme. While this assumption sounds reasonable for certain applications, it cannot by any means be generalized and presented as suitable for any P2P system. All in all, we identify a need for unsupervised approaches and we believe that future research should follow this direction.

3.2 Scalability

One of the claims of P2P computing is the unlimited scalability that can be achieved by exploiting the aggregate capabilities of all participating peers. Semantic overlays are proposed as a mechanism that improves the efficiency of search, so any such approach should pay particular attention to scalability. Potential bottlenecks in terms of communication costs (consumed bandwidth, latency, etc.) should be thoroughly studied. Synchronization is also costly and should be avoided. Load-balancing is equally important, especially in the cases where the individual peer load may have an aggregate effect, i.e., increase with the size of the network. Any P2P system based on SONs should be able to scale well with the number of peers. In the absence of sufficiently large testbeds, researchers use simulations to test the scalability of proposed systems.

Overlay networks, such as Gnutella, have problems related to scalability [26]. In particular, the time required to locate content in a large network can be extremely long, with high associated costs. Structured P2P systems solve this issue, by being able to find the answer to a query with logarithmic cost, however their feasibility is still questionable, especially in the case of high churn rates. The dynamism of a P2P system, where peers may arbitrarily fail or join the network, poses another threat against ensuring scalable solutions.

Current P2P research focusing on SON generation should put scalability as number one requirement, as this need will become more evident in the future, where the urge for such viable, completely distributed systems is expected to increase.

3.3 Self-organization

Informally, the spontaneous activity towards organization of a system is described by the concept of self-organization. The basic mechanism for self-organization is dynamic topology adaptation, as a means to reorganize each peer's neighbors. In this way overlay networks are created on top of the initial P2P overlay network. We stress here the essence of self-organization: there is no need for enforcing external observation and maintenance mechanisms. Self-organization is one of the most challenging requirements in P2P systems and, at the same time, one of the most difficult to achieve.

3.4 Autonomy

Peer autonomy is an important concept in P2P networks, which is indirectly related to other issues like fault-tolerance. Peer autonomy means that each peer can be as independent as possible of the limitations imposed by the P2P protocol, concerning both its behavior and as well as its content. In particular, independence with respect to content means that each peer does not have to replicate its local data or provide explicit indices to its local data to other peers. Moreover, a peer should not be imposed to host indices to data that belong to other peers. In this sense, unstructured P2P systems respect peer autonomy, in contrast with structured P2P systems. As a consequence, unstructured P2P systems are more resilient to failures, because in general, a peer failure makes only its local content unavailable, while in a DHT-based network, recovery mechanisms must be enforced for recovery and continuous correct operation. However, peer autonomy comes with a cost: it is usually difficult to provide efficient searches.

3.5 Decentralization

While distribution is inherently related to the P2P concept, the same does not always hold for decentralization. The most evident example of centralized P2P system was Napster, one of the first P2P systems to enable file sharing between participating computers. However, while the actual file exchange was performed in a P2P manner, the index to files was held in a centralized location. This is not an acceptable approach for dynamic P2P systems, since it presents a single point of failure.

Learning from the shortcomings of such approaches, we can extend and apply the lessons learned to any new P2P system design. This also holds for SON generation, especially in large-scale networks. If operations are centralized, this endangers the completeness of SON generation, with obvious consequences to the correctness of the final overlays. Also, while for small networks a centralized solution may seem appropriate, due to the assembly of global knowledge, where better decisions can be made, however it usually presents problems when applied to large-scale networks. The main reason is communication bottlenecks that often result in non-applicable or infeasible approaches, or in other words algorithms that do not scale.

4 The DESENT Approach to SON Generation

In this section, we summarize the results of our approach for unsupervised SON generation. Our approach is called DESENT, which stands for DEcentralized SEmantic overlay NeTwork generation. DESENT's design attempts to take into account most of the requirements expressed in the previous section. The approach is based on creating local zones of peers, forming semantic clusters based on data stored on these peers, and then merging zones and clusters recursively until global zones and clusters are obtained. We assume that peers store documents, though other data representations can also be supported.

4.1 Peer Clustering

The peer clustering process is divided into 5 phases: 1) local clustering, 2) zone initiator selection, 3) zone creation, 4) intra-zone clustering, and 5) inter-zone clustering.

Phase 1: Local Clustering. In the process of determining peers that contain related documents, *feature vectors* are used instead of the actual documents because of the large amounts of data involved. A feature vector F_i is a vector of tuples, each tuple containing a feature (word) f_i and a weight w_i. The feature vectors are created using a feature extraction process. By performing clustering of the document collection at each peer, a set of document clusters is created, each cluster represented by a feature vector.

Phase 2: Initiator Selection. Assuming Z_i is the set of all peers in zone i, the zone consists of $|Z_i|$ peers, and one of these peers is given the role of *initiator*, which subsequently initiates and controls the clustering process within the zone. The process of choosing initiators is completely distributed. Because of load-balancing, the aim is to have as uniform zone sizes as possible, of approximately S_Z peers per zone. Assuming the IP of a peer P_i is IP_{P_i} and the time is T (rounded to nearest t_a[3]), a peer will discover that it is an initiator if $(IP_{P_i} + T)\ MOD\ S_Z = 0$. The aim of the function is to select initiators that are uniformly spread out in the network and an appropriate number of initiators relative to the total number of peers in the network. By including time in the function we ensure that we obtain different initiators each time the clustering algorithm is run. This tackles the problem of being stuck with faulty initiators, as well as reduces the problem of permanent cheaters.

[3] Assuming that each peer has a clock that is accurate within a certain amount of time t_a, note that DESENT itself can be used to improve the accuracy.

Phase 3: Zone Creation. After a peer P_i has discovered that it is an initiator, it uses a probe-based technique to create its zone. This zone creation algorithm has a low cost, and in the case of excessive zone sizes, the initiator can decide to partition its zone, thus sharing its load with other peers. When this algorithm terminates, 1) each initiator has assembled a set of peers Z_i and their capabilities, in terms of resources they possess, 2) each peer knows the initiator responsible for its zone and 3) each initiator knows the identities of its neighboring initiators. An interesting characteristic of this algorithm is that it ensures that all peers in the network will be contacted, as long as they are connected to the network. This is essential, otherwise there may exist peers whose content will never be retrieved. We refer to [10] for more details on initiator selection and zone creation.

Phase 4: Intra-zone Clustering. After the zones and their initiators have been determined, global clustering starts by collecting feature vectors from the peers and creating clusters based on these feature vectors. The initiator of each zone i collects the feature vectors from the peers in Z_i and performs a clustering algorithm, resulting in a set of N_C^0 basic clusters. The initiator selects a representative peer R_i for each cluster, based on resource information that is provided during Phase 3, like peer bandwidth, connectivity, etc. The result kept at the initiator is a set of cluster descriptions (CDs), one for each cluster C_i. A CD consists of the cluster identifier C_i, a feature vector F_i, the set of peers $\{P\}$ belonging to the cluster, and the representative R of the cluster, i.e., $\text{CD}_i = (C_i, F_i, \{P\}, R)$. Each of the representative peers are informed by the initiator about the assignment and receive a copy of the CDs (of *all* clusters in the zone). The representatives then inform peers on their cluster membership by sending them messages of the type (C_i, F_i, R).

Phase 5: Inter-zone Clustering. At this point, each initiator has identified the clusters in its zone. These clusters can be employed to reduce the cost and increase the quality of answers to queries involving the peers in one zone. However, in many cases peers in other zones will be able to provide more relevant responses to queries. Thus, we need to create an overlay that can help in routing queries to clusters in remote zones. In order to achieve this, we recursively apply merging of zones to larger and larger super-zones, and at the same time merge clusters that are sufficiently similar into super-clusters: first a set of neighboring zones are combined to a super-zone, then neighboring super-zones are combined to a larger super-zone, etc. Note that level-i initiators are a subset of the level-$(i-1)$ initiators. Due to lack of space, we refer to [10] for the actual details of inter-zone clustering.

We emphasize that even though parts of this process resemble a centralized approach, this is not the case: initiators are chosen at random and perform their tasks completely independent of each other. Also, the role of the final peer in the super-initiator hierarchy is only to determine that the global process is finished. Failure of this peer will be discovered and another peer can perform the task. As can be noted, initiators actually have similarities with super-peers, but one important difference is that their role is not constant.

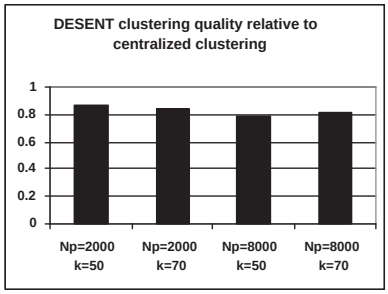

Fig. 1. Simulation results: Cluster quality compared to centralized clustering for different network sizes and values of k.

4.2 Experimental Results

We have developed a simulation environment in Java, which covers all intermediate phases of the overlay network generation. At initialization of the P2P network, a topology of N_P interconnected peers is created. We used the GT-ITM topology generator[4] to create random graphs of peers, and our own SQUARE topology, which is similar to GT-ITM, only the network is more dense.

A collection of N_D documents is distributed to peers, so that each peer retains N_D/N_P distinct documents. Every peer runs a clustering algorithm on its local documents resulting in a set of initial clusters. In our experiments we chose the Reuters-21578 text categorization test collection[5], and we used 8000 pre-classified documents that belong to 60 distinct categories. We tried two different experimental setups with 2000 and 8000 peers. We then performed feature extraction (tokenization, stemming, stop-word removal and finally keeping the top-k features based on their TF/IDF[6] value and kept a feature vector of top-k features for each document as a compact document description). Thus, each document is represented by a top-k feature vector.

We used hierarchical agglomerative clustering to create clusters of documents at each initiator. Clustering is based on computing document similarities and merging feature vectors, by taking the union of the clusters' features and keeping the top-k features with higher TF/IDF values. We used the cosine similarity with parameter the similarity threshold T_s for merging. Obviously, other clustering algorithms, as well as other similarity measures can be used.

We compare the quality of clustering results (unsupervised learning procedure) to the document classification (supervised learning) as carried out by humans. This is probably not a completely fair comparison, since we take as granted the ground truth of the human classification into categories. Here, even the centralized clustering performs rather poorly. Thus we compare the clustering quality of our approach to the centralized clustering results. We used in our experiments the F-measure as a cluster quality

[4] http://www.cc.gatech.edu/projects/gtitm/

[5] http://www.daviddlewis.com/resources/testcollections/reuters21578/

[6] Notice that the inverse document frequency (IDF) is not available, since no peer has global knowledge of the document corpus, so we use the TF/IDF values produced on each peer locally, taking only the local documents into account.

measure. F-measure ranges between 0 and 1, with higher values corresponding to better clustering. The average values of DESENT F-measure relative to centralized clustering are illustrated in Fig. 1, and show that DESENT achieves high clustering quality. Also note that the results exhibit a relatively stable behavior as the network size increases. This indicates that DESENT scales well with the number of participating peers. This conveys that the proposed system achieves high quality in forming SONs despite of the lack of global knowledge and the high distribution of the content.

5 Conclusions and Further Work

In this paper we have focused on semantic overlay generation in unstructured P2P networks. We have expressed a set of basic requirements that any SON generation algorithm should try to enforce, namely: unsupervised approaches for P2P clustering, scalability, self-organization, autonomy and decentralization. Furthermore, we summarized the research results of our approach regarding unsupervised semantic overlay generation in large-scale P2P networks. We recognize the following future research directions: a) emergence in the SON generation process, b) design of algorithms applicable to Internet-scale environments, c) metrics to ensure the quality of the formed overlays and d) maintenance of semantic overlays in the presence of high churn.

References

1. K. Aberer, P. Cudre-Mauroux, M. Hauswirth, and T. V. Pelt. Gridvine: Building internet-scale semantic overlay networks. In *Proceedings of ISWC'04*, 2004.
2. K. Aberer et al. P-Grid: A Self-organizing Structured P2P System. *SIGMOD Record*, 32(3):29–33, 2003.
3. Y. Chawathe, S. Ratnasamy, L. Breslau, N. Lanham, and S. Shenker. Making Gnutella-like P2P systems scalable. In *Proceedings of SIGCOMM'03*, 2003.
4. V. Cholvi, P. Felber, and E. Biersack. Efficient search in unstructured peer-to-peer networks. In *Proceedings of SPAA'04*, 2004.
5. I. Clarke, O. Sandberg, B. Wiley, and T. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability*, 2000.
6. E. Cohen, H. Kaplan, and A. Fiat. Associative search in peer to peer networks: Harnessing latent semantics. In *Proceedings of INFOCOM'03*, 2003.
7. A. Crespo and H. Garcia-Molina. Routing Indices for Peer-to-Peer Systems. In *Proceedings of ICDCS'02*, 2002.
8. A. Crespo and H. Garcia-Molina. Semantic Overlay Networks for P2P Systems. Technical report, Stanford University, 2002.
9. F. Cuenca-Acuna et al. PlanetP: using gossiping to build content addressable peer-to-peer information sharing communities. In *Proceedings of HPDC'03*, 2003.
10. C. Doulkeridis, K. Nørvåg, and M. Vazirgiannis. DESENT: Decentralized and Distributed Semantic Overlay Generation in P2P Networks. Technical report, AUEB, 2005 http://www.db-net.aueb.gr/index.php/publications/technical_reports/.
11. E.Adar and B.Huberman. Free riding on Gnutella. *First Monday*, 5(10), 2000.
12. M. Eisenhardt, W. Mueller, and A. Henrich. Classifying Documents by Distributed P2P Clustering. *GI Jahrestagung*, pages 286–291, 2003.

13. C. Gkantsidis, M. Mihail, and A. Saberi. Hybrid search schemes for unstructured peer-to-peer networks. In *Proceedings of INFOCOM'05*, 2005.
14. Gnutella, http://www.gnutella.com/.
15. F. Liu, F. Ma, M. Li, and L. Huang. Distributed information retrieval based on hierarchical semantic overlay network. In *Proceedings of GCC'04*, 2004.
16. X. Liu, J. Wang, and S. T. Vuong. A category overlay infrastructure for peer-to-peer content search. In *Proceedings of IPDPS'05*, 2005.
17. A. Loeser. Taxonomy-based Routing Overlays in P2P Networks. In *Proceedings of IDEAS'04*, 2004.
18. A. Loeser, F. Naumann, W. Siberski, W. Nejdl, and U. Thaden. Semantic overlay clusters within super-peer networks. In *Proceedings of DBISP2P'03*, 2003.
19. J. Lv and X. Cheng. WonGoo: A pure peer-to-peer full text information retrieval system based on semantic overlay networks. In *Proceedings of NCA'04*, 2004.
20. E. P. Markatos. Tracing a large-scale P2P system: an hour in the life of Gnutella. In *Proceedings of CCGrid'02*, 2002.
21. W. Nejdl, M. Wolpers, W. Siberski, C. Schmitz, M. Schlosser, I. Brunkhorst, and A. Loeser. Super-Peer-based Routing and Clustering Strategies for RDF-based P2P Networks. In *Proceedings of WWW'03*, 2003.
22. J. Parreira, S. Michel, and G. Weikum. p2pDating: real life inspired semantic overlay networks for web search. In *Proceedings of SIGIR Wokshop on heterogeneous and distributed information retrieval*, 2005.
23. L. Pireddu and M. Nascimento. Taxonomy-based routing indices for peer-to-peer networks. In *Proceedings of the SIGIR Workshop on Peer-to-Peer Information Retrieval*, 2004.
24. L. Ramaswamy, B. Gedik, and L. Liu. Connectivity based Node Clustering in Decentralized Peer-to-Peer Networks. In *Proceedings of P2P'03*, 2003.
25. S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker. A Scalable Content-addressable Network. In *Proceedings of SIGCOMM'01*, 2001.
26. J. Ritter. Why Gnutella can't scale. No, really, http://www.darkridge.com/ jpr5/doc/gnutella.html, 2001.
27. A. Rowstron and P. Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In *Middleware'01*, 2001.
28. H. T. Shen, Y. Shu, and B. Yu. Efficient semantic-based content search in P2P network. *IEEE Transactions on Knowledge and Data Engineering*, 16(7):813–826, 2004.
29. I. Stoica et al. Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications. In *Proceedings of SIGCOMM'01*, 2001.
30. C. Tang, Z. Xu, and S. Dwarkadas. Peer-to-Peer Information Retrieval Using Self-Organizing Semantic Overlay Networks. In *Proceedings of SIGCOMM'03*, 2003.
31. C. Tempich, S. Staab, and A. Wranik. REMINDIN': Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors. In *Proceedings of WWW'2004*, 2004.
32. P. Triantafillou, C. Xiruhaki, M. Koubarakis, and N. Ntarmos. Towards High Performance Peer-to-Peer Content and Resource Sharing Systems. In *Proceedings of CIDR'03*, 2003.
33. B. Yang and H. Garcia-Molina. Improving search in peer-to-peer networks. In *Proceedings of ICDCS'02*, 2002.
34. B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz. Tapestry: A resilient global-scale overlay for service deployment. *IEEE Journal on Selected Areas in Communications*, 22(1):41–53, Jan. 2004.