Limited Scale-Free Overlay Topologies for Unstructured Peer-to-Peer Networks

Hasan Guclu, *Member*, *IEEE*, and Murat Yuksel, *Member*, *IEEE*

Abstract—In unstructured peer-to-peer (P2P) networks, the overlay topology (or connectivity graph) among peers is a crucial component in addition to the peer/data organization and search. Topological characteristics have profound impact on the efficiency of a search on such unstructured P2P networks, as well as other networks. A key limitation of scale-free (power-law) topologies is the high load (i.e., high degree) on a very few number of hub nodes. In a typical unstructured P2P network, peers are not willing to maintain high degrees/loads as they may not want to store a large number of entries for construction of the overlay topology. Therefore, to achieve fairness and practicality among all peers, hard cutoffs on the number of entries are imposed by the individual peers, which limits scale-freeness of the overall topology, hence limited scale-free networks. Thus, it is expected that the efficiency of the flooding search reduces as the size of the hard cutoff does. We investigate the construction of scale-free topologies with hard cutoffs (i.e., there are not any major hubs) and the effect of these hard cutoffs on the search efficiency. Interestingly, we observe that the efficiency of normalized flooding and random walk search algorithms increases as the hard cutoff decreases.

Index Terms—Unstructured peer-to-peer networks, scale-free networks, power-law networks, search efficiency, cutoff.

1 Introduction

In decentralized P2P networks, the overlay topology (or connectivity graph) among peers is a crucial component in addition to the peer/data organization and search. Topological characteristics have profound impact on the efficiency of a search on P2P networks, as well as other networks. It has been well known that a search on small-world topologies can be as efficient as $O(\ln N)$ [1], and this phenomenon has recently been studied on P2P networks [2], [3].

The best search efficiency in realistic networks can be achieved when the topology is scale free (power law), which offers search efficiencies like $O(\ln \ln N)$. However, the generation and maintenance of such scale-free topologies are hard to realize in a distributed and potentially uncooperative environments as in P2P networks. A key limitation of scale-free topologies is the high load (i.e., high degree) on a very few number of hub nodes. In a typical unstructured P2P network, peers are not willing to maintain high degrees/loads as they may not want to store large number of entries for construction of the overlay topology. Therefore, to achieve fairness and practicality among all peers, *hard* cutoffs on the number of entries are imposed by the individual peers, which makes the overall network a limited one. These hard cutoffs might limit the *scale-freeness* of the overall topology, by which we mean having a network with a power-law degree distribution from which an exponent can be obtained properly. Thus, it is expected that the search efficiency reduces as the size of the hard cutoff does. In this paper, we use the terms "scale-free network with a hard cutoff" and "limited scale-free network" interchangeably since we mean that the lower the hard cutoff, the more limited the network.

The primary focus of this paper is to 1) investigate the construction of scale-free topologies with hard cutoffs (i.e., there are not any major hubs) in a distributed manner without requiring global topology information at the time when nodes join and 2) to investigate the effect of these hard cutoffs on the search efficiency.

The rest of the paper is organized as follows: First, in the rest of this section, we provide motivation for this work, outline key dimensions to be considered, and briefly indicate major contributions and findings of the work. Then, we survey previous work on P2P networks in Section 2. In Section 3, we survey the previous work on scale-free topology generation and cover two specific models that we use: Preferential Attachment (PA) and Configuration Model (CM). We introduce our practical topology generation methodologies, Hop-and-Attempt PA (HAPA) and Discover-and-Attempt PA (DAPA), in Section 4. In Section 5, we present our simulations of three different search algorithms (i.e., Flooding (FL), Normalized Flooding (NF), and Random Walk (RW)) on topologies generated by the models PA, CM, HAPA, and DAPA. We conclude by summarizing the work and outlining future directions in Section 6.

1.1 Motivation and Key Considerations

The search efficiency of small-world and scale-free topologies is well known. Although scale-free topologies are superior in search efficiency, their hub-based structure makes them vulnerable to threats and impractical due to unfair assignment of network load on a very small subset of all nodes. As peers in a P2P network are typically not fully cooperative, protocols cannot rely on methods working

- *H. Guclu is with the School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623. E-mail: guclu@rit.edu.*
- *M. Yuksel is with the Computer Science and Engineering Department, 171 University of Nevada, Reno, 1664 N. Virginia St., Reno, NV 89557-0042. E-mail: yuksem@cse.unr.edu.*

Manuscript received 2 Aug. 2007; revised 18 June 2008; accepted 30 July 2008; published online 4 Aug. 2008.
Recommended for acceptance by C. Shahabi.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2007-08-0264.
Digital Object Identifier no. 10.1109/TPDS.2008.150.

with the full cooperation of peers. For example, peers may not want to store a large number of entries for the construction of the overlay topology, i.e., connectivity graph. Even though the characteristics of the overlay topology are crucial in determining the efficiency of the network, peers typically do not want to take the burden of storing an excessive amount of control information for others in the network. The effect of this on the overlay topology maintenance is that peers impose *hard cutoffs* on the amount of control information to be stored. Since P2P overlay topology generation and maintenance are very important for realizing a scalable unstructured P2P network, the focus of this paper is *to investigate the effect of the hard cutoffs on the overall search efficiency*.

A key issue is the construction of scale-free overlay topologies without global information. There are several techniques to generate a scale-free topology [4], [5], which rely on *global* information about the current network when a new node joins. Such global methods are not practical in P2P networks, and *local* heuristics in generating such scale-free overlay topologies with hard cutoffs are the key issue, which we investigate in this paper. In other words, each peer has to figure out the optimal way of joining the P2P overlay by only using the locally available (i.e., immediate/close neighbors) information and also causing a minimal inefficiency to the search mechanisms to be run on the network.

1.2 Contributions and Major Results

This paper touches an uncovered set of research problems relating to trade-offs between the maximum number of links a peer can (or is willing to) store and the efficiency of a search on an overlay topology composed of such peers. We defined the maximum number of links to be stored by peers as the *hard cutoff* for the degree of a peer in the network as compared to *natural cutoff*, which occurs due to finite-size effects. Our contributions include the following:

- *Scale-free topology generation methods.* We studied two well-known scale-free topology generation mechanisms (i.e., PA and CM) that use global information about the overlay topology within the context of unstructured P2P networks. We introduced two novel mechanisms (i.e., HAPA and DAPA) that use local topology information solely or partially.
- *Search efficiency on scale-free topologies with hard cutoffs.* Through extensive simulations, we studied the efficiency of FL, NF, and RW on the topologies generated by the four mechanisms PA, CM, HAPA, and DAPA.
- *Guidelines for designing peer join algorithms for unstructured P2P networks.* Our study yielded several guidelines for peers to join to a Gnutella-like unstructured P2P network, so that the search performance of the overall overlay topology remains high.

Our study of hard cutoffs resulted in several interesting findings, some of which are listed as follows:

- *Hard cutoffs may not always affect the search performance adversely.* We found that hard cutoffs may actually improve the search performance, even though the value of the power-law exponent in the degree distribution of the topology might mean otherwise. This is against the expected wisdom that the power-law exponent is directly related to search performance.
- *Search performance depends on the combination of the specific search algorithm and the topology.* We showed that search performance depends on the particular search algorithm being used and on the topological characteristics, including the exponent of the degree distribution, connectedness (the minimum degree is a measure for it in scale-free networks), hard cutoff, and locality. Our simulation experiments clearly showed that practical search algorithms like NF or repeated RWs can perform better on scale-free topologies with smaller hard cutoffs as long as peers join carefully, e.g., as in HAPA and DAPA mechanisms.
- *There exists an interplay between the connectedness and the degree distribution exponent for a fixed cutoff.* More specifically, if connectedness is too low in the topology, then one can improve search performance by applying smaller hard cutoffs. As a particular guideline for optimizing joining techniques of peers, we showed that as long as every peer is required to maintain a minimum of two to three links to the rest of the network rather than just one link, it is possible to diminish the negative effects of hard cutoffs on search performance.

2 Related Work

Our work is related to peer-to-peer (P2P) network protocol designs and the topological analysis of complex networks. Previous work on P2P network protocols can be classified into *centralized* and *decentralized* ones. As centralized P2P protocols (e.g., Napster [6]) proved to be unscalable, the majority of P2P research has focused on decentralized schemes. The decentralized P2P schemes can be further classified into subcategories: *structured*, *unstructured*, and *hybrid*.

In the structured P2P networks, the data/file content of peers is organized based on a keying mechanism that can work in a distributed manner, e.g., Chord [7] and Kademlia [8]. The keying mechanism typically maps the peers (or their content) to a logical search space, which is then leveraged for performing efficient searches. Another positive side of the structured schemes is the guarantees of finding rare items in a timely manner. However, the cost comes from the complexity of maintaining the consistency of mapping the peers to the logical search space, which typically causes a considerable amount of control traffic (e.g., join/leave messaging) for highly dynamic P2P environments. Due to their capability of locating rare items, structured approaches have been very well suited to a wide range of applications, e.g., [9], [10], and [11].

In contrast to the structured schemes, unstructured P2P networks do not include a strict organization of peers or their content. Since there is no particular keying or organization of the content, the search techniques are typically based on flooding. Thus, the searches may take a very long time for rare items, although popular items can be found very fast due to possible leveraging of locality of reference [12] and caching/replication [13].

The main focus of the research on unstructured P2P networks has been the trade-off between the state complexity of peers (i.e., the number of records needed to be stored at each peer) and flooding-based search efficiency. The minimal state each peer has to maintain is the *list of neighbor peers*, which construct the overlay topology. Optionally, peers can maintain *forwarding tables* (also referred as routing tables in the literature) for data items in addition to the list of neighbor peers. Thus, we can classify unstructured P2P networks into two based on the type(s) of state peers maintain: 1) *per-data* unstructured P2P networks (i.e., peers maintain both the list of neighbor peers and the per-data forwarding table) and 2) *non-per-data* unstructured P2P networks (i.e., peers maintain only the list of neighbor peers).

Non-per-data schemes are mainly Gnutella-like schemes [14], where a search is performed by means of flooding query packets. Search performance over such P2P networks has been studied in various contexts, which includes pure RWs [15], probabilistic flooding techniques [16], and systematic filtering techniques [17]. Recent research also recognized the fact that the overlay topology needs to be organized with information about the underlying Internet topology to achieve better routing performance [18], although we focus on improving search performance with better topology organization.

Per-data schemes (e.g., Freenet [19]) can achieve better search performance than non-per-data schemes, though they impose additional storage requirements to peers. By making the peers maintain a number of $\langle \text{key}, \text{pointer} \rangle$ entries, peers direct the search queries to more appropriate neighbors, where "key" is an identifier for the data item being searched, and the "pointer" is the next-best neighbor to reach that data item. This capability allows peers to leverage associativity characteristics of search queries [20]. Studies ranged from grouping peers of similar interests (i.e., peer associativity) [3], [20] to exploiting locality in search queries (i.e., query associativity) [12].

Our work is applicable to both per-data and non-per-data unstructured P2P networks, since we focus on the interactions between search efficiency and topological characteristics. Topology adaptation for better protocol performance was studied in various contexts, such as breaking or establishment of overlay links based on the perceived load on peers [21], age of the peers [22], capacity of the link [23], or lookup latency [24].

3 Scale-Free Network Topologies

Recent research shows that many natural and artificial systems such as the Internet [25], World Wide Web [26], scientific collaboration network [27], and e-mail network [28] have power-law degree (connectivity) distributions. These systems are commonly known as power-law or scale-free networks since their degree distributions are free of scale (i.e., not a function of the number of network nodes N) and follow power-law distributions over many orders of magnitude. This phenomenon has been represented by the probability of having nodes with k degrees as $P(k) \sim k^{-\gamma}$, where γ is usually between two and three [4]. Scale-free networks have many interesting properties such as high tolerance to random errors and attacks (yet low tolerance to attacks targeted to hubs) [29], high synchronizability [30], and resistance to congestion [31].

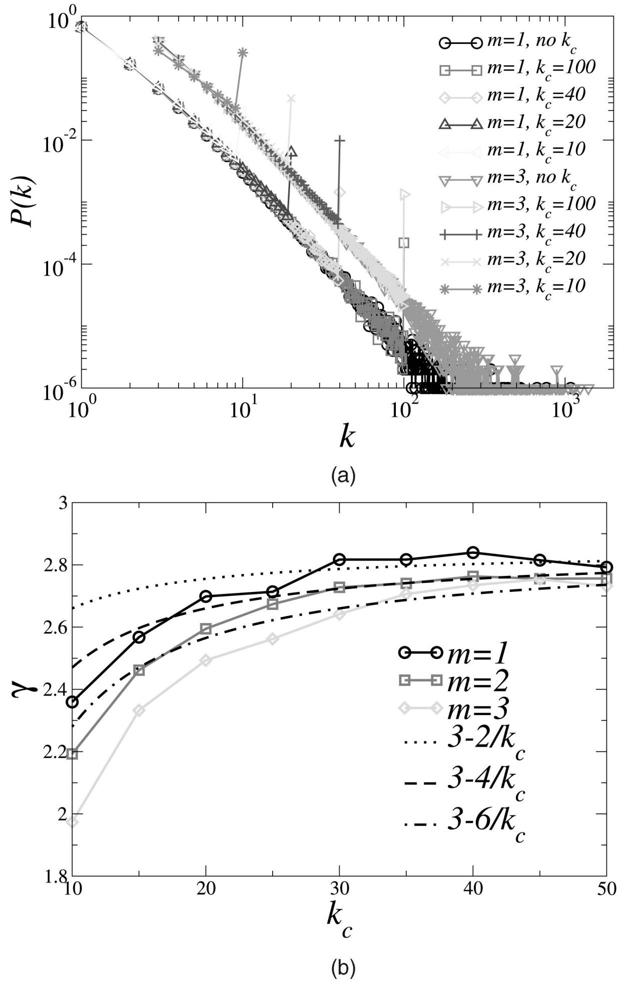


Fig. 1. Degree distributions of the PA model. (a) $P(k)$ with hard cutoffs. (b) $P(k)$ exponent versus cutoff.

The origin of the scale-free behavior can be traced back to two mechanisms that are present in many systems and have a strong impact on the final topology [4]. First, networks are developed by the addition of new nodes that are connected to those already present in the system. This mechanism signifies continuous expansion in real networks. Second, there is a higher probability that a new node is linked to a node that already has a large number of connections. These two features led to the formulation of a growing network model first proposed by Barabási and Albert that generates a scale-free network for which $P(k)$ follows a power law with $\gamma = 3$. This model is known as PA (or a rich-get-richer mechanism), and the resulting network is called the *Barabási-Albert* network [4], [5].

In this study, we use a simple version of the PA model [4]. The model evolves by one node at a time, and this new node is connected to m (the number of stubs) different existing nodes with a probability proportional to their degrees, i.e., $P_i = k_i / \sum_j k_j$, where k_i is the degree of the node i. The average degree per node in the resulting network is $2m$, and the minimum degree is m. Fig. 1a shows the degree distributions of scale-free networks generated by the PA model with different m values. The links are regarded as bidirectional links; however, the results can easily be generalized to directed networks as well [5]. The special case of the PA model is when the

TABLE 1
Scale-Free Network Diameter Behavior

Diameter d	Exponent γ	# of stubs m
$\ln \ln N$	(2,3)	≥ 1
$\ln N / \ln \ln N$	3	≥ 2
$\ln N$	3	1
$\ln N$	> 3	≥ 1

number of stubs is one (i.e., $m = 1$) in which a scale-free tree without clustering (loops) is generated.

Scale-free networks are very robust against random failures and attacks since the probability to hit the hub nodes (few nodes with a very large degree) is very small, and attacking the low-degree satellite nodes does not harm the network. On the other hand, deliberate attacks targeted to hubs through which most of the traffic go can easily shatter the network and severely damage the overall communication in the network. For the same reason, the Internet is called "robust yet fragile" [32] or "Achilles' heel" [29].

Scale-free networks also have *small-world* properties. In small-world networks, the diameter or the mean hop distance between the nodes scales with the system size (or the number of network nodes) N logarithmically, i.e., $d \sim \ln N$. The scale-free networks with $2 < \gamma < 3$ have a much smaller diameter and can be named *ultrasmall* networks [33], behaving as $d \sim \ln \ln N$. When $\gamma = 3$ and $m \geq 2$, d behaves as $d \sim \ln N / \ln \ln N$. However, when $m = 1$ and $\gamma = 3$, the Barabási-Albert model turns into a tree, and $d \sim \ln N$ is obtained. Also, when $\gamma > 3$, the diameter behaves logarithmically as $d \sim \ln N$. These relationships are summarized in Table 1. Since the speed/efficiency of search algorithms strongly depend on the average shortest path, scale-free networks have much better performance in a search than other random networks.

3.1 The Cutoff

One of the important characteristics of scale-free networks is the natural cutoff on the degree (or the maximum degree) due to finite-size effects. Natural cutoff can be defined as [34] the value of the degree above which one expects to find at most one vertex, i.e.,

$$N \int_{k_{nc}}^{\infty} P(k) dk \sim 1. \tag{1}$$

By using the degree distribution for the scale-free network and the exact form of probability distribution (i.e., $P(k) = (\gamma - 1) m^{\gamma - 1} / k^{\gamma}$), one obtains

$$k_{nc}(N) \sim m N^{1/(\gamma - 1)}, \tag{2}$$

which is known as the *natural* cutoff of the network. The scaling of the natural cutoff can also be calculated by using the extreme-value theory [35]. For the scale-free networks generated by the PA model ($\gamma = 3$), the natural cutoff becomes $k_{nc}(N) \sim m\sqrt{N}$.

3.2 Preferential Attachment with Hard Cutoffs

The natural cutoff may not be always attainable for most of the scale-free networks due to technical reasons. One main reason is that the network might have limitations on the number of links the nodes can have. This is especially important for P2P networks in which nodes cannot possibly connect many other nodes. This requires putting an artificial or *hard* cutoff k_c to the number of links one node might have.

In order to implement the hard cutoff in PA, we simply did not allow nodes to have links more than a fixed hard cutoff value during the attachment process. This modified method generates a scale-free network in which there are many nodes with degree fixed to a hard cutoff instead of a few very high degree hubs and the degree distribution still decays in a power-law fashion. The degree distribution of the PA model with cutoff is slightly different than that of PA without a cutoff in terms of exponent and an accumulation of nodes with degree equal to the hard cutoff. The PA model, in its original form, has a degree distribution exponent $\gamma = 3$ for very large networks. However, when a hard cutoff is imposed, it is observed that the absolute value of the degree distribution exponent decreases as in Fig. 1b.

One can use the master-equation [36] approach to analyze the effects of the hard cutoff on the topological characteristics. We grow the network by introducing new nodes one by one for simplicity. Each new node links to m earlier nodes in the network. The probability that the new node attaches to a previous node of degree k is defined to be A_k/A, where A_k is the rate of attachment to a previous node, and this rate depends only on the degree of the target node, while $A = \sum_{k=m}^{k_c - 1} A_k N_k$ is the total rate for all events, and N_k is the number of nodes of degree k in the network. Thus, A_k/A equals the probability for the newly introduced node to attach to a node of degree k. The new feature that we study is the effect of a hard cutoff in the degree of each node. Once the degree of a node reaches k_c, it is defined to become inert so that no further attachment to this node can occur. Thus, only nodes with degrees $k = m, m+1, \ldots, k_c - 1$ are active. This restriction is the source of the cutoff in the definition of the total attachment rate. We now study the degree distribution, $N_k(N)$, as a function of the cutoff k_c and the total number of nodes in the network N.

The master equations for the degree distribution can be written by using the fact that N_k is proportional to N, and thus, $N_k \to N n_k$ and $A \to \nu N$ as

$$n_k = \begin{cases} -\frac{m n_m}{\nu} + 1, & k = m, \\ \frac{(k-1) n_{k-1} - k n_k}{\nu}, & k = m+1, \ldots, k_c - 1, \\ \frac{(k_c - 1) n_{k_c - 1}}{\nu}, & k = k_c. \end{cases} \tag{3}$$

By the nature of these equations, it is evident that n_{k_c} is of a different order than n_k with $k < k_c$. Starting with the solution $n_m = \nu/(m + \nu)$, we can find n_k by subsequent substitutions. This recursive approach gives us a chance to write n_k values as products [36], and by converting these products into Euler gamma functions, we show that n_{k_c} scales as $k^{-\nu}$, while for $k < k_c$, n_k scales as $k^{-(\nu+1)}$. We can obtain the coefficient ν in $A = \nu N$ self-consistently from $A = \sum_{k=m}^{k_c - 1} A_k n_k \equiv \nu N$, or equivalently, $\nu = \sum_{k=m}^{k_c - 1} A_k n_k$. By rewriting the sum above as a difference between two sums with limits from the minimum degree to ∞ and from cutoff to ∞ and by taking asymptotic limits [37] of large N and k_c, we get

$$\nu \to 2 - \frac{2m}{k_c}. \tag{4}$$

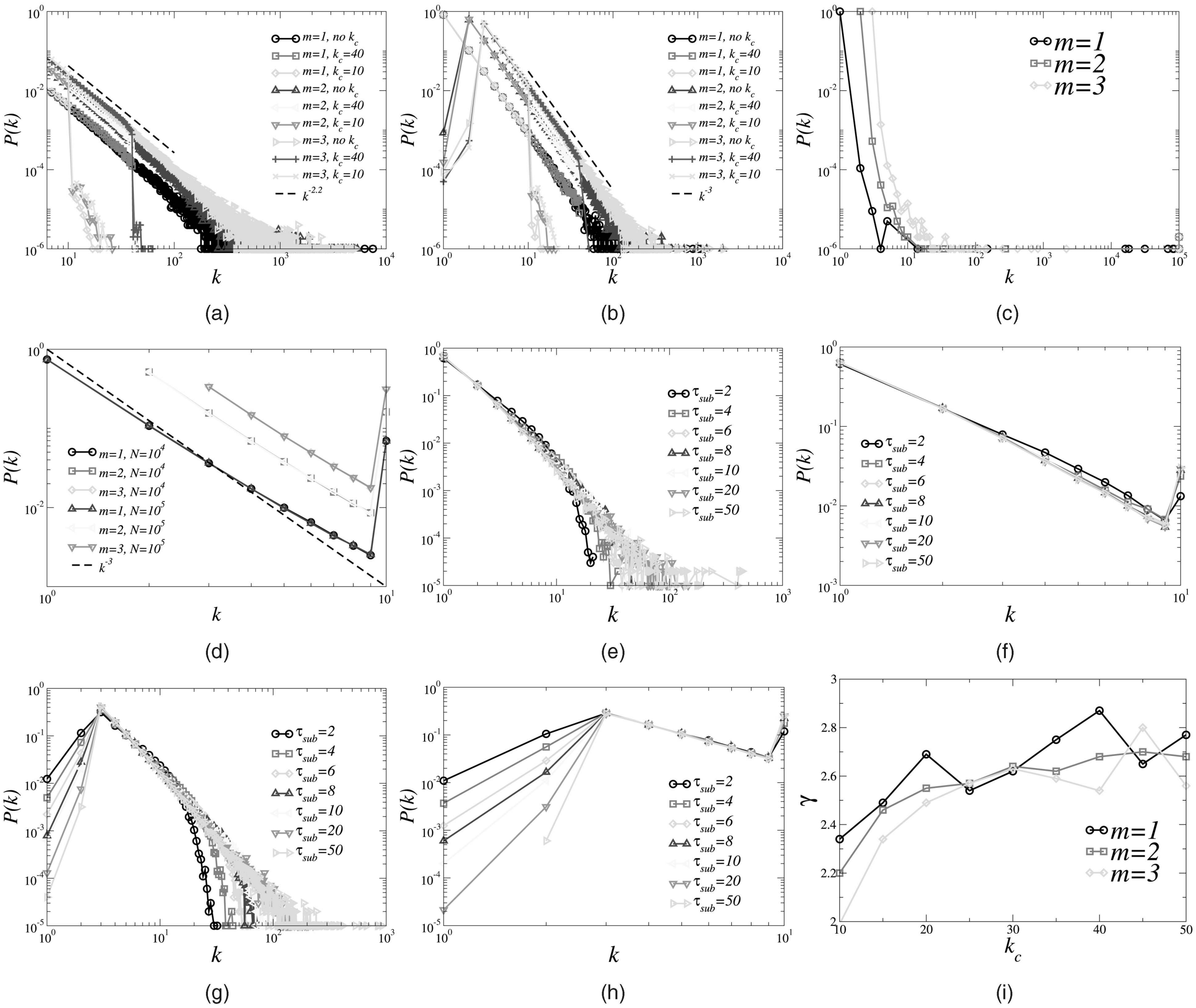


Fig. 2. Degree distributions for CM, HAPA, and DAPA. (a) CM $(\gamma = 2.2)$. (b) CM $(\gamma = 3)$. (c) HAPA (no cutoff). (d) HAPA $(k_c = 10)$. (e) DAPA ($m = 1$, no cutoff). (f) DAPA $(m = 1, k_c = 10)$. (g) DAPA ($m = 3$, no cutoff). (h) DAPA $(m = 3, k_c = 10)$. (i) DAPA (γ versus k_c).

This result shows that $n_k \sim k^{-(3-2m/k_c)}$ for $k < k_c$ and $n_{k_c} \sim k_c^{-(2-2m/k_c)}$, confirming the change we observe in the degree distribution exponent, as in Fig. 1b, for which we rescaled the above equation with respect to the maximum exponent one can get for this system size, which is around 2.85.

3.3 Configuration Model (CM)

Given that the PA model yields lower degree distribution exponents as the hard cutoff reduces, we were motivated to work on the generation of power-law networks with different exponents. In this manner, the spikes at the hard cutoff value in Fig. 1a can be prevented, and a smooth power-law distribution of degrees can be obtained. For this reason, modified PA models such as nonlinear PA [36], dynamic edge-rewiring, [5], and fitness models [38] have been proposed. Here, we use the CM with a predefined degree distribution to generate a static scale-free network [39].

CM was introduced as an algorithm to generate uncorrelated random networks with a given degree distribution. In CM, the vertices of the graph are assigned a fixed sequence of degrees $\{k_i\}_{i=1}^{N}$, $m \leq k_i \leq k_c$, where typically $k_c = N$, chosen at random from the desired degree distribution $P(k)$ and with the additional constraint that the $\sum_i k_i$ must be even. Then, pairs of nodes are chosen randomly and connected by undirected edges. This model generates a network with the expected degree distribution and no degree correlations; however, it allows self-loops and multiple connections when it is implemented as described above. It was proved in [35] that the number of multiple connections when the maximum degree is fixed to the system size, i.e., $k_c = N$, scales with the system size N as $N^{3-\gamma} \ln N$. Since we work with hard cutoff values typically less than the natural cutoff, the number of multiple links is much less than the original CM for which $k_c = N$ [40]. After this procedure, we simply delete the multiple connections and self-loops from the network, which gives a very marginal error in the degree distribution exponent. Deleting this discrepancies also causes some very negligible number of nodes in the network to have degrees less than the fixed minimum degree (m) value (even zero), as seen in Figs. 2a and 2b. One other characteristic of the CM is that the network is not a connected network when $m = 1$, i.e., it has disconnected clusters (or components). For $m > 1$, the network is most likely connected, having one giant component including all the nodes.

The main disadvantage of PA and CM methods is that they require global knowledge about the network, i.e., the degrees of all peers or the maximum/total degree, which might be usually difficult to store and share by the nodes for unstructured P2P networks. This motivated us to modify the PA model so that it makes use of local information as much as possible. In the next section, we explore the local heuristics in creating scale-free networks and introduce two new attachment models.

4 Local Heuristics for Scale-Free Overlay Topology Construction

In the PA model, as outlined in the previous section, the new node has to make random attempts to connect to the existing nodes with a probability depending on the degree of the existing node. To implement this in a P2P (or any distributed) network, the new node has to have information about the global topology (e.g., the current number of degrees each node has for the PA model), which might be very hard to maintain in reality. Such global topology information is needed in the CM method as well.

Thus, in order for a topology construction mechanism to be practical in P2P networks, it must allow joining of new nodes by just using locally available information. Of course, the cost of using only local information is expected to be loss of scale-freeness (or any other desired characteristics) of the whole overlay topology, which will result in the loss of search efficiency in return. In this section, we present two practical methods using local heuristics requiring partial or no global information about the topology: HAPA and DAPA.

4.1 Hop-and-Attempt Preferential Attachment (HAPA)

In this method, the new node randomly selects an existing node and attempts to connect. Then, it randomly selects a node that is a neighbor of the previously selected node and attempts to connect. Thus, the new node hops between the neighboring nodes by using the existing links in the network and attempts to connect until it fills all its stubs, i.e., the number of links it has reached m.

This hopping process gives a better chance to the new node to find the high-degree hubs in the network than the PA does since the hubs in scale-free networks are only a couple of hops away from the low-degree nodes, and it is less likely to find hubs by random node selection. Therefore, some nodes in the network (probably, they are the initial nodes, and their number is $m+1$ in this algorithm) become dominant and attract almost all the nodes to themselves and thus deserve the name *superhubs.* The superhubs have degrees on the order of the network size. It is easily seen that this procedure makes the topology of the system a starlike topology if the network is not limited by a cutoff. Naturally, without a hard cutoff, the degree distribution is not a power law, and the average shortest path/diameter is very small with respect to scale-free networks generated by PA (see Fig. 2c). As shown in Fig. 2d, when a hard cutoff is introduced, the degree distribution gets closer to a power law having an exponent $\gamma = 3$ but with possibly logarithmic factors, making a degree exponent calculation very hard.

4.2 Discover-and-Attempt Preferential Attachment (DAPA)

The DAPA model imitates the method for finding peers in Gnutella-like unstructured P2P networks. First, we assume that we have a network, called a *substrate* network, with a predefined and preconstructed topology at hand. Then, we construct an overlay network on this substrate by using the PA method among the set of nodes visible/reachable to a node (the horizon of the node) in a number of hops, which we call *local time to live* (TTL) and represent with τ_{sub}. The substrate network corresponds to the underlying Internet connectivity among the nodes, and the nodes become peers in an overlay network by reaching each other via the substrate network. This process imitates the way Gnutella-like overlay networks are established.

We use a two-dimensional geometric (euclidean) random network (GRN) [41] with a giant component as a substrate network. To construct a two-dimensional GRN, first, nodes are randomly distributed on a unit square. Then, pairs of nodes are linked if they are closer to each other than some specific distance. GRNs have Poissonian degree distributions with only one parameter, i.e., the average degree. The average degree of a GRN depends on the specific distance and system size, e.g., for the specific distance of 0.012 and the system size 10^4, the network has an average degree of 4.5. This specific distance is the critical distance beyond which the network has a giant component. Throughout the paper, we use a two-dimensional GRN as a substrate network with an average degree of 10 and size $2N$. We use a GRN as a substrate network because it is topologically closer to real-life nodes in the Internet than a regular or highly random network. We could use also a two-dimensional mesh, but our conclusions about the search efficiency would not change.

In DAPA, initially, a few nodes are randomly selected from the substrate network and added to the previously empty overlay network; then, these nodes are connected to each other in the overlay network. At each step, one random node is chosen in the substrate network, and let it send a query to its neighborhood reachable in τ_{sub} hops to get a list of peers in its horizon. Then, by using the rules of PA, the new node connects to m peers with a probability proportional to their degrees divided by the total degrees of the peers in its horizon. If the number of peers in the horizon is less than m, then the new node connects to all the peers it can find. The nodes that can find at least one peer in their horizon is added to the overlay network and becomes a peer. A peer that belongs to the overlay network cannot be selected again to look for new peers. This process is continued until the number of peers in the overlay network reaches the desired number of nodes.

The degree distribution of the network generated by the DAPA model exhibits some interesting characteristics. For small values of τ_{sub}, the nodes are shortsighted, i.e., they cannot see enough peers in their short horizon, causing the degree distribution to be exponential. For high-enough τ_{sub} values, the degree distribution changes into a power law. Thus, one can go from an exponential to a scale-free network by playing with the measure of locality (τ_{sub}), as can be seen in Figs. 2e, 2f, 2g, 2h, and 2i. As the hard cutoff gets smaller, the difference between the degree distributions

TABLE 2
Comparison of Different Network Generation Models

Model	Usage of Global Information
PA	Yes
CM	Yes
HAPA	Partial
DAPA	No (τ_{sub} dependent)

with different τ_{sub} becomes invisible. For higher values of m (i.e., $m > 1$), it is possible to find peers with degree less than m, as in Figs. 2g and 2h, since some nodes cannot find enough peers in their horizon to fill all their stubs. The degree distribution exponent has a similar behavior to PA as we change the hard cutoff value, i.e., as the cutoff decreases, the exponent increases (see Fig. 2i). The data in Fig. 2i is very noisy, and the data points contain quite large error bars because they are obtained from very scattered degree distribution tails.

A comparison of different network generation models in terms of locality can be seen in Table 2. When a peer is to join the current overlay topology, PA and CM need global information about the current topology, while HAPA and DAPA methods use local information partially or mostly, respectively. Therefore, HAPA and DAPA methods are more practical in the context of unstructured P2P networks.

4.3 Effect of Hard Cutoffs on Topologies

Depending on the way one applies hard cutoffs to an initially scale-free topology results in different topological characteristics, such as the degree distribution, diameter (or expected search efficiency). Table 1 summarizes the interrelationship between these three dimensions that PA models were studied in literature.

When we applied hard cutoffs to the regular PA topologies, their degree distribution looked like a power-law distribution except for a spike in the frequency of nodes having a degree equal to the hard cutoff (see Fig. 1a). Unlike the original PA topologies without any cutoff, these topologies exhibit different power-law exponents when the spike on the hard cutoffs is not taken into account. We estimated the power-law exponents for these PA topologies with hard cutoffs and plotted Fig. 1b, which shows the power-law exponent of the degree distribution versus the hard cutoff. As expected, Fig. 1c shows that the degree distribution exponent γ degrades to lower values when harder cutoffs are applied, suggesting that search efficiency (in connection with the diameter size) will also degrade for smaller cutoffs.

CM does not allow changes in the degree distribution exponent because the degree sequence is drawn from a predefined distribution generated by using a specific degree exponent (see Figs. 2a and 2b). The only change in degree distribution exponent is due to the deletion of self-loops and multiple connections, and this can be considered negligible. It is also observed that applying harder (smaller) cutoffs to the degrees decreases the probability to have self-loops and multiple connections.

In the HAPA model, it is not even possible to say that we still have power-law degree distributions. Without a hard cutoff, the degree distribution decreases very fast as the degree increases, and there are a few nodes with degree on the order of the system size, i.e., starlike topology (see Fig. 2c). Applying a cutoff destroys the starlike topology and changes the degree distribution into one similar to a power law with possibly cutoff-dependent logarithmic corrections, as can be seen in Fig. 2d.

The DAPA model is qualitatively very similar to PA for high-enough τ_{sub} values (see Figs. 2e, 2f, 2g, 2h, and 2i). The small τ_{sub} makes the network an exponential one. By tuning this parameter, one can change the degree distribution from exponential to power law. As in the PA model, applying a harder cutoff decreases the degree distribution exponent, as can be seen in Fig. 2i.

5 Simulations

In P2P networks that do not have a central server, including Gnutella and Freenet, files are found by forwarding queries to neighbors until the target is found. In the previous sections, in addition to studying well-known techniques like PA and CM for scale-free topology construction, we introduced new algorithms (i.e., HAPA and DAPA) with the same purpose within the context of unstructured P2P networks. Here, we study a number of message-passing algorithms that can be efficiently used to search items in P2P networks utilizing the power-law (the presence of hubs) degree distribution in sample networks generated by our topology construction algorithms. These algorithms are completely decentralized and do not use any kind of global knowledge about the network. We consider three different search algorithms: FL, NF, and RW. The goals of our simulation experiments include the following:

- *Effect of hard cutoffs on search efficiency.* Applying hard cutoffs on power-law topologies reduces the degree distribution exponent, which should affect the search efficiency (i.e., the *number of hits per unit time*) on such topologies. We are interested in observing this effect for the three search algorithms on the topologies constructed by our algorithms.
- *Topology construction with global versus local information.* Though we showed in the previous section that using local information when a peer is joining yields a less scale-free topology, the effect of this on search efficiency still needs to be shed light on. Our simulations aim to investigate this too.
- *Messaging complexity.* One side effect of changing topology characteristics is that it will affect the messaging complexity (i.e., the *number of messages per search request*) of the search algorithms. We would like to observe this effect as well.

5.1 Search Algorithms

FL. FL is the most common search algorithm in unstructured P2P networks. In the search by FL, the source node s sends a message to all its nearest neighbors. If the neighbors do not have the requested item, they send on to their nearest neighbors, excluding the source node (see Fig. 3a). This process is repeated a certain number of times, which is usually called *TTL*, and we represent it with τ in this paper. After a message is forwarded an amount of time equal to τ,

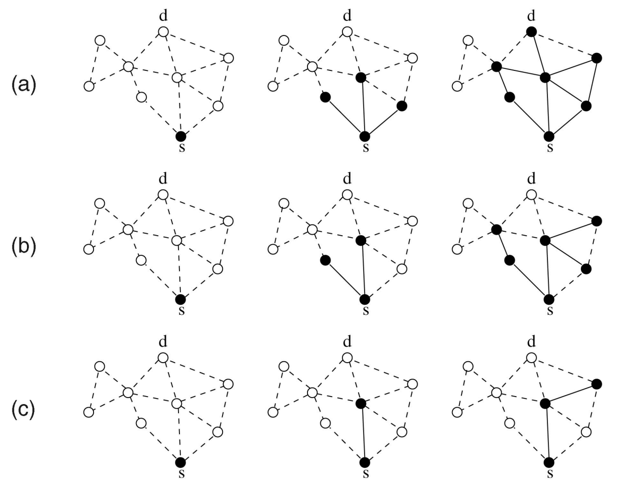


Fig. 3. Search strategies. (a) FL or broadcast search. (b) NF. (c) RW search.

it is discarded. Independent floods by the nodes make the FL algorithm parallel. On the other hand, in this algorithm, a large number of messages is created since the destination node cannot stop the search. This corresponds to a complete sweep of all the nodes within a τ-hop distance from the source. The delivery time in the search by FL is measured by the number of intermediate links traversed and is equal to the shortest path length. Since the average shortest path for small-world networks, including scale-free ones, is proportional to the logarithm of the system size N or even slower, the average delivery time (T_N) is logarithmic as well, i.e., $T_N = \log(N)$.

The main disadvantage with FL is that it requires a large amount of messaging traffic because most of the nodes are visited and forced to exchange messages, which makes the search by FL unscalable. Another disadvantage is that FL has poor granularity, i.e., each additional step in the search significantly increases the number of nodes visited [13]. Yet the search efficiency of FL (i.e., the number of hits per search) provides a way of determining how other realistic and scalable search algorithms can perform in comparison to the best possible, i.e., the search efficiency of FL.

NF. In the search by FL, when large-degree nodes (hubs) are reached, the number of neighbors for the next step in FL increases dramatically, leading to poor granularity. This also causes a lot of shared edges, reducing the performance in terms of the number of messages per distinct number of discovered nodes. To overcome this problem, the search by the NF algorithm was introduced in [16]. In NF, the minimum degree m in the network is an important factor. The NF search algorithm proceeds as follows: When a node of degree m receives a message, the node forwards the message to all of its neighbors, excluding the node that forwarded the message in the previous step. When a node with a larger degree receives the message, it forwards the message only to randomly chosen m neighbors, except for the one that forwarded the message. The NF mechanism is illustrated in Fig. 3b. In this simple network with $m = 2$, the source node sends a message to its randomly chosen two neighbors, and these neighbors forward the message to their randomly chosen two neighbors. In the third step, the message reaches its destination.

The NF search algorithm is based on the minimum degree in the network. The fixed minimum degree is equal to m by definition in PA and HAPA, whereas in CM an DAPA, it is not guaranteed that the minimum degree will be m. In CM, the deletion of self-loops and multiple links might reduce the minimum degree to values less than m. In DAPA, however, the minimum degree might be less than m because of the short range of horizon for some peers that are geographically far from others. But still, since the ratio of nodes with degree less than m is small, we ignored them and ran the NF algorithm based on the predefined minimum degree value m.

RW. RW or multiple RWs have been used as an alternative search algorithm to achieve even better granularity than NF. In RW, the message from the source node is sent to a randomly chosen neighbor. Then, this random neighbor takes the message and sends it to one of its random neighbors, excluding the node from which it got the message. This continues until the destination node is reached or the total number of hops is equal to τ. A schematic of RW can be seen in Fig. 3c. RW can also be seen as a special case of FL where only one neighbor is forwarded the search query, providing the other extreme situation of the trade-off between delivery time and messaging complexity. The RW search is inherently serial (sequential), which causes a large increase in the delivery time [42], [15]. In particular, computer simulations performed on a generalized scale-free network with degree exponent $\gamma = 2.1$, which is equal to the value observed in P2P networks, yield the result [43] $T_N = N^{0.79}$. Although the RW search is worse than the FL search in scale-free networks in terms of the time needed to locate a given node, the average total traffic in the network is equal to T_N and therefore scales sublinearly with N, better than the linear growth of the FL search.

5.2 Results

We simulated the three search algorithms FL, NF, and RW on the topologies generated by PA, CM, HAPA, and DAPA and provide results for all the combinations with various hard cutoffs. Through the PA, CM, HAPA, and DAPA methods, we generated topologies with 10,000 nodes. We used different cutoff values from 10 to 100 (or just a few in this range), in addition to the natural cutoff, i.e., no hard cutoff. When generating DAPA topologies, we used τ_{sub} values of 2, 4, 6, 8, 10, 20, and 50 with the expectation that a larger τ_{sub} should yield better search efficiency. The minimum degree values (or m) in our topologies were 1, 2, and 3. We varied the τ values of search queries in FL up to the point we reach the system size, and for NF/RW, up to 10. To compare search efficiencies of RW and NF fairly in our simulations, we equated the τ of RW searches to the number of messages incurred by the NF searches in the same scenario. Thus, for the search efficiency graphs of RW when τ is equal to a particular value such as four, this means that the number of hits corresponding to that $\tau = 4$ value is obtained by simulating an RW search with τ equal to the number of messages that were caused by an NF search using a τ value of four. A similar normalization was done in [16].

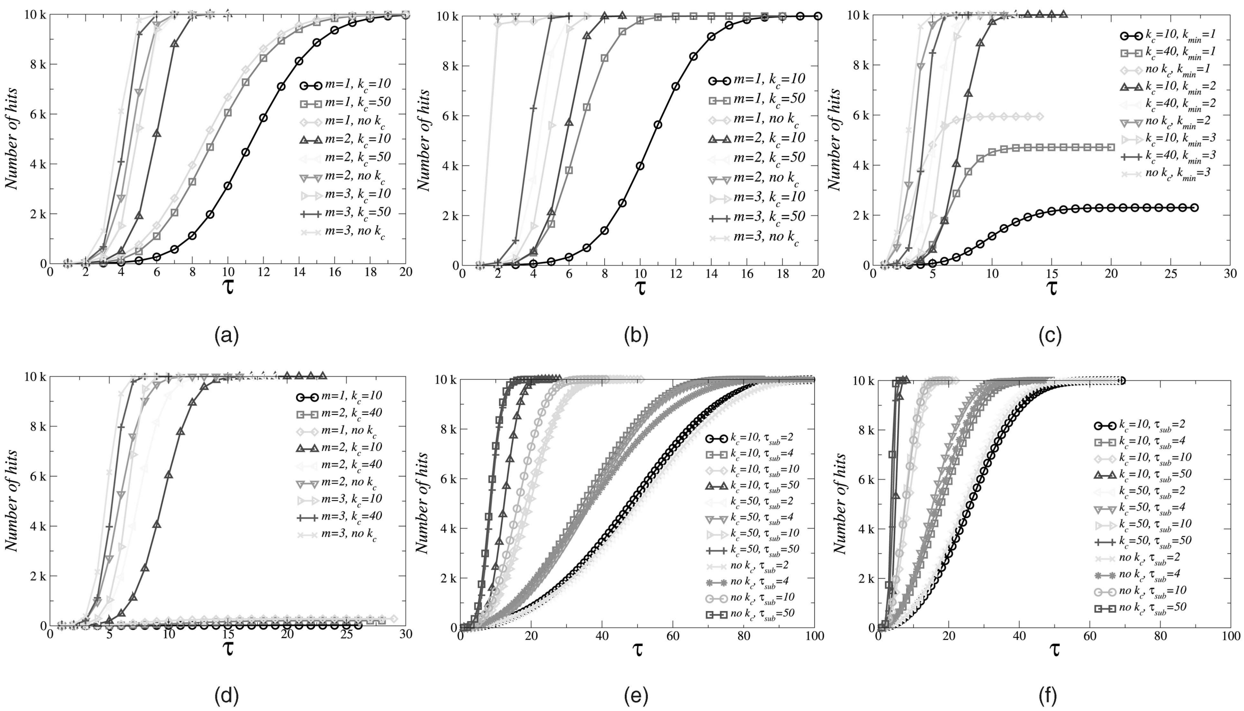


Fig. 4. FL results for PA, CM, and HAPA. (a) PA. (b) HAPA. (c) CM ($\gamma = 2.2$). (d) CM ($\gamma = 3$). (e) DAPA ($m = 1$). (f) DAPA ($m = 3$).

Another important point to note is that the number of hits in RW and NF include the source node as well, which makes the values one more than the usual.

Search efficiency. As τ varies, Fig. 4a shows the number of hits achieved by FL on various topologies with size $N = 10^4$ generated by the PA method. To illustrate the effects of the network size, we show FL results (scaled to one) for three different system sizes in Fig. 5a. As can be seen in this figure, scaled FL results show that as the system size gets larger, the performance decreases relatively for $m = 1$, but for larger values of m, this effect disappears. Similarly, Figs. 4c and 4d shows the search efficiency of FL on the topologies generated by the CM. In both of these figures, as expected, when there is no hard cutoff in the topology, the FL algorithm can achieve higher search efficiency by capturing more of the peers in the network for a specific τ value. Also, the effect of imposing a hard cutoff reduces when the minimum degree in the topology is higher. One interesting feature of CM is that when the minimum number of links is one (i.e., $m = 1$), the number of hits cannot reach the system size even for very large τ values because the network is not a connected one for $m = 1$. Fig. 4b shows a similar search efficiency behavior for FL on HAPA topologies for a fixed-size network, and Fig. 5b, for different network sizes with even more apparent effect of hard cutoff. For small values of cutoff, PA and HAPA give similar performance in FL, whereas for higher values of cutoff, HAPA has better hit results due to the starlike topology. The FL in DAPA is less efficient than that in PA, although for higher values of τ_{sub}, it gets closer to PA, and the efficiency of FL increases, as can be seen in Figs. 4e and 4f. We did not show the FL results for different network sizes for CM and DAPA because the behavior with respect to the network size is virtually identical in all cases.

A minimum of three links for all peers eliminates the negative effects of hard cutoffs. An interesting observation is that the negative effect of hard cutoffs on the FL performance on the PA and HAPA topologies can be easily reduced to negligible values by increasing the number of stubs m (or connectedness). The number of stubs as small as three leaves virtually no difference between the search performance of overlay topologies with or without hard cutoffs. This result provides the guideline that *to achieve a better FL performance, a requirement of having at least three links to the rest of the network will be adequate to assure that no one else in the network will need to maintain an unbearably large number of links.* However, the necessity of complete or partial *global* information about the overall network when constructing a PA or HAPA topology is a major discouragement of using them for generating overlay topologies of unstructured P2P networks.

There exists an interplay between connectedness and the degree distribution exponent for a fixed cutoff. As the DAPA method is a purely local method, it is more interesting to observe search performance on the DAPA topologies. Figs. 4e and 4f show the FL performance on DAPA topologies generated with minimum degrees (or numbers of stubs) of one and three, respectively. In each of these figures, search performance is shown for different τ_{sub} values: 2, 4, 6, 8, 10, 20, and 50. Interestingly, when there is weak connectedness (i.e., $m = 1$), Fig. 4e shows that imposing hard cutoffs improves the search performance. This is due to the fact that hard cutoffs increase the connectedness of the topology by moving the links that would normally go to a hub in a topology without a hard cutoff. However, when the number of stubs is larger (Fig. 4f), we observe an interplay between the degree distribution exponent and the connectedness for a fixed cutoff. We observe that the improvement caused by hard cutoffs depends on the value of the hard cutoff, suggesting that reducing the hard cutoff value hurts the search performance

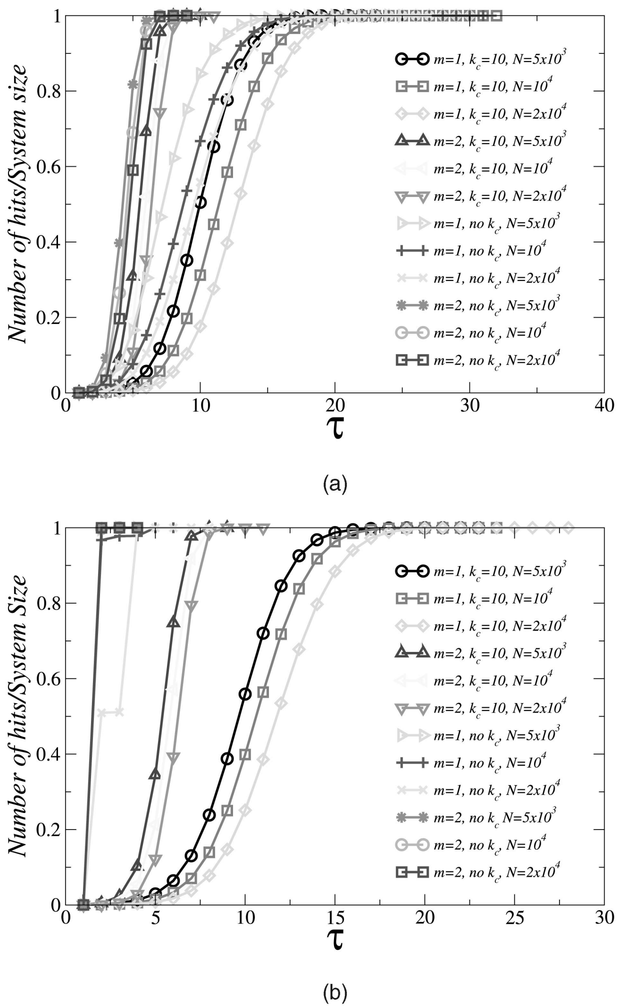


Fig. 5. Scaled FL results for PA and HAPA models for different network sizes. (a) PA. (b) HAPA.

after a while. That is, *potential improvements by having smaller hard cutoffs diminish as the performance starts to become dominated by the degree distribution exponent rather than the connectedness.* Another observation to be made is that the impact of local information plays a major role in the search performance, as can be seen in Figs. 4e and 4f.

Hard cutoffs may improve search efficiency in NF and RW. More interestingly, for NF and RW, improvements due to having hard cutoffs are apparent in all three topology generation methods, including the PA, regardless of the number of stubs m. The only exception to this behavior is the CM, as shown in Figs. 6b and 8b for NF and RW, respectively. This means that practical search algorithms like NF and multiple RWs are affected positively by having hard cutoffs on the overlay topology. For NF, this is evidenced in Figs. 6a, 6d, 6c, and 6f for the PA and HAPA topologies, respectively. Having a little more local connectivity to the network by having a minimum of two to three links in every peer increases the search performance rapidly for the same τ values (i.e., by comparing Figs. 6a and 6d). For RW, a very similar behavior is exhibited in Figs. 8a, 8d, 8c, and 8f, with the only difference being that the effect of hard cutoffs is more apparent due to the fact that NF does better averaging of search possibilities. The observed behavior of RW illustrates how good the effect of hard cutoffs can be on the search performance. It is intuitive that multiple RWs would perform more similar to NF in terms of performance.

More global information is more important for search efficiency when target connectedness is high. Fig. 7 shows the performance of NF on various DAPA topologies with different parameters. Figs. 7a and 7c shows the search performance on a linear scale when $m = 1$, while Figs. 7b and 7d show it on a semilogarithmic scale when $m = 3$. We observe again that as the hard cutoff is getting smaller, the search efficiency improves, regardless of the connectedness m. Also, having a little better connectedness (e.g., $m = 3$) improves the search performance greatly. An interesting observation is that when constructing the overlay topology, having more information (i.e., larger τ_{sub}) about the global topology (and, thus, more scale-freeness in the overall topology) yields more important improvements on the search performance for topologies with more connectedness, i.e., larger m. This means that for the purpose of constructing topologies with better search performance, when the target connectedness value is high, one needs to be more patient and obtain as much information as possible before finalizing its links to the rest of the peers.

DAPA with high τ_{sub} and HAPA models perform almost as good as the CM with $\gamma = 2.2$. An interesting characteristic to observe is how close the performance of DAPA and HAPA is to the low-γ CM for the NF and RW search algorithms. Specifically, topologies generated by the CM do not have spikes at the hard cutoff values (e.g., Fig. 1a) in their degree distributions, in such a way that the links are configured in the perfect manner to assure that no node has links more than the target hard cutoff and the degrees of nodes follow exactly the predefined power law. This can be seen by comparing Figs. 2a and 2b with its counterparts Figs. 1, 2c, 2d, 2e, 2f, 2g, 2h, and 2i. As it can be seen in Figs. 6c and 6f, with connectedness $m = 1$, 2, and 3, HAPA performs slightly worse than CM when using NF. Similarly, the DAPA performance for moderate τ_{sub} values (e.g., six) is very close to that of CM (e.g., Fig. 9). For small or no connectedness $m = 1$, the behavior is the same, in that the DAPA and HAPA performance is close to the CM performance, as shown in Figs. 7a and 7c, 6c, and 6b for DAPA, HAPA, and CM, respectively.

We also simulated our search algorithms on networks with different sizes such as $N = 5,000$ and $N = 20,000$. We observed that for NF and RW at $m = 1$, the results are almost identical for different network sizes. At higher values of m, the same exponential behavior observed but with a different prefactor for each. Typically, higher m values allow NF and RW algorithms to reach more distant nodes on larger networks.

Messaging complexity. We also looked at the complexity of messaging overhead for the search algorithm and topology combinations. We specifically looked at the average number of messages incurred by a search request. As the FL algorithm is an extreme and is not scalable in terms of messaging complexity, we did not study its performance. In all cases, NF performs better than RW consistently, though the difference between the two algorithms diminishes as τ increases for weak connectedness, i.e., $m = 1$. But, for stronger connectedness, i.e., $m > 1$, the difference between NF and RW is more apparent. More importantly, although the effect of hard cutoffs is negative

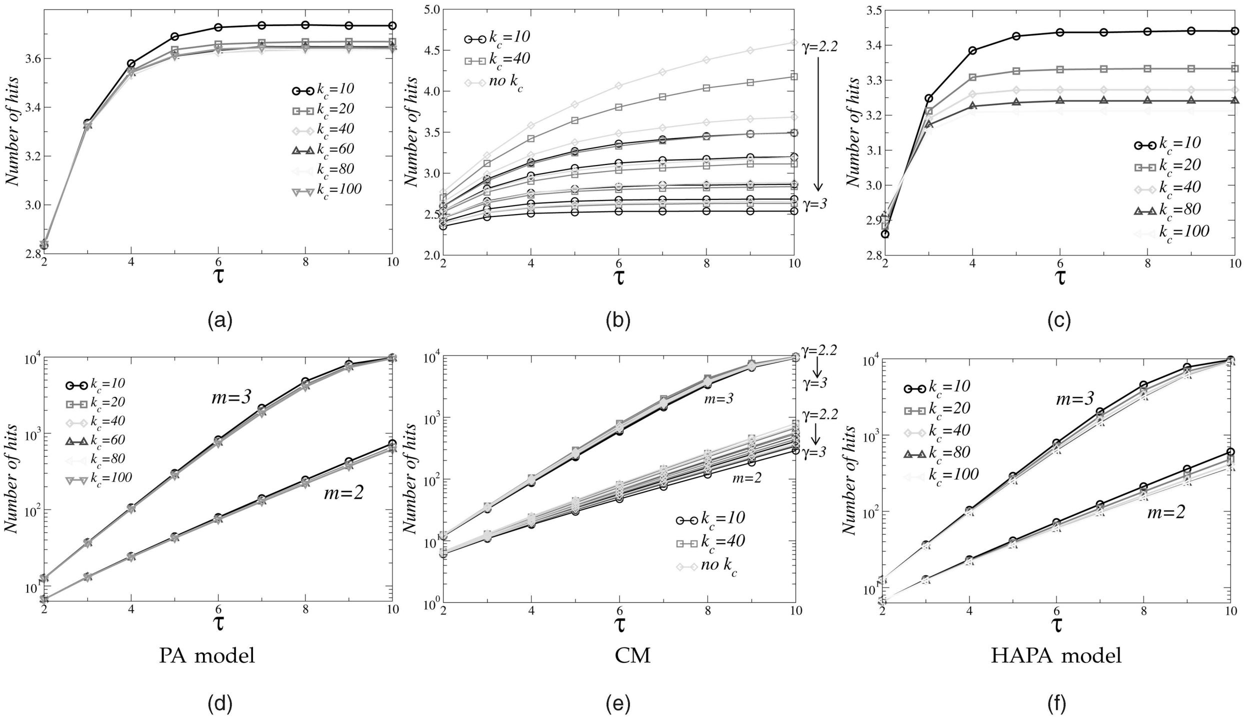


Fig. 6. NF results for PA, CM, and HAPA. (a) $m = 1$. (b) $m = 1$. (c) $m = 1$. (d) $m = 2$ and $m = 3$. (e) $m = 2$ and $m = 3$. (f) $m = 2$ and $m = 3$.

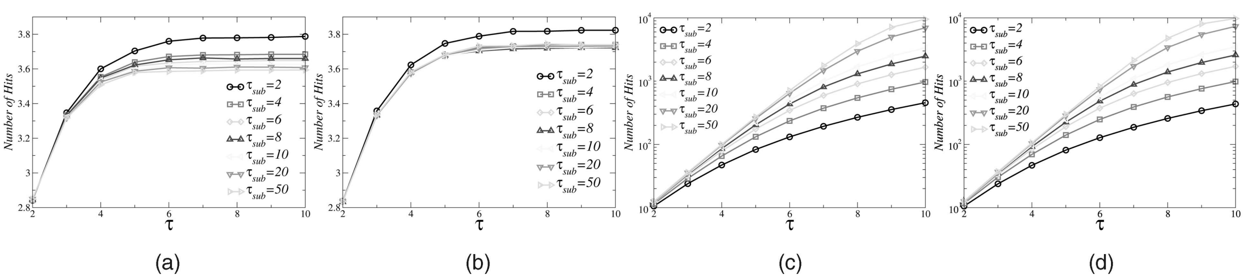


Fig. 7. NF results for the DAPA model. (a) $m = 1$ and no cutoff. (b) $m = 1$ and $k_c = 10$. (c) $m = 3$ and no cutoff. (d) $m = 3$ and $k_c = 10$.

in terms of messaging complexity, we observed that this negative effect is very minimal and negligible, given that improvements on the search performance were observed for smaller hard cutoffs.

Design guidelines and principles. Based on the observations we have made in our simulation experiments, we outline the following design guidelines for constructing an overlay P2P topology:

- *Enforce establishing a minimum of three links to the rest of the network when a peer is joining.* Our results showed that to achieve better flooding performance a requirement of having at least three links to the rest of the network will be adequate to assure that no one else in the network will need to maintain an unbearably large number of links.
- *For weakly connected networks with a treelike topology, apply smaller hard cutoffs.* When the connectedness of peers is weak (i.e., the number of links is one or two), enforcing tighter hard cutoffs is better for search performance. This will ensure that the joining peers will attempt to connect to the peers with perhaps a very small number of links, and thus, this reinforcement will yield a topology with better connectedness and search performance.
- *For strongly connected networks, force the joining peers to be more patient and collect more global information about the existing network topology.* Our results showed that having more information (i.e., larger τ_{sub}) about the global topology (and, thus, more scale-freeness in the overall topology) yields more important improvements on the search performance for topologies with more connectedness, i.e., larger m. This means that for the purpose of constructing topologies with better search performance, when the existing network has peers with high connectedness, the joining peers need to be more patient and obtain as much information as possible before finalizing its links to the rest of the peers.

6 Summary and Discussions

We studied effects of the hard cutoffs peers impose on the number of entries they store on the search efficiency. Specifically, we showed that the exponent of the degree distribution reduces as hard cutoffs imposed by peers become smaller. We introduced new scale-free topology

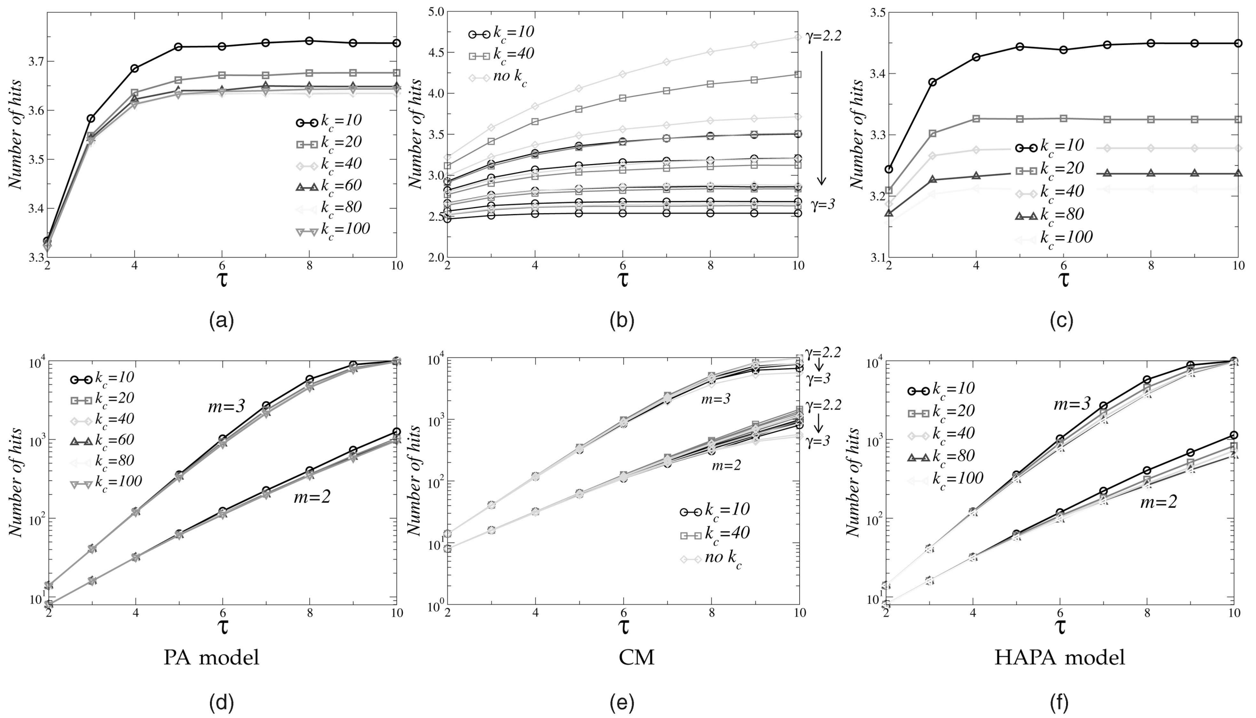


Fig. 8. RW results for PA, CM, and HAPA models. (a) $m = 1$. (b) $m = 1$. (c) $m = 1$. (d) $m = 2$ and $m = 3$. (e) $m = 2$ and $m = 3$. (f) $m = 2$ and $m = 3$.

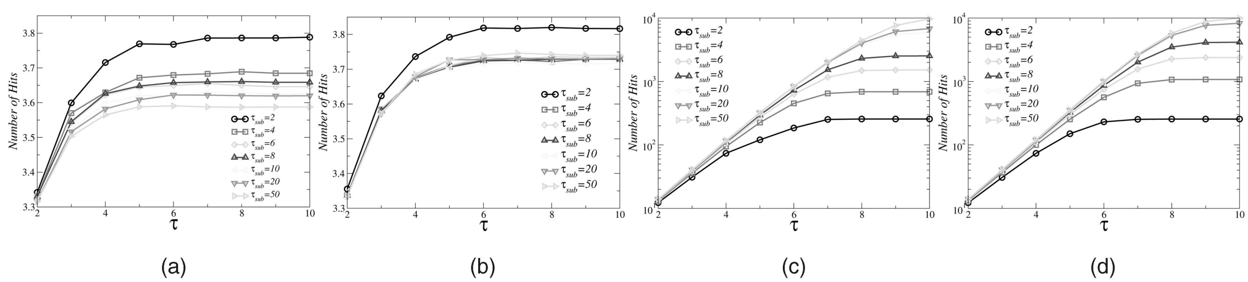


Fig. 9. RW results for the DAPA model. (a) $m = 1$ and no cutoff. (b) $m = 1$ and $k_c = 10$. (c) $m = 3$ and no cutoff. (d) $m = 3$ and $k_c = 10$.

generation mechanisms that use completely or partially local information unlike traditional scale-free topology generation mechanisms using global topology information. We showed that topologies generated by our mechanisms allow better search efficiency in practical search algorithms. Our study also revealed that interplay between the degree distribution exponent with a fixed hard cutoff and connectedness is likely to occur when using our mechanisms.

Future work will include the study of join/leave scenarios for the overlay topologies while attempting to maintain the scale-freeness of the overall topology. The challenge is to achieve minimal messaging overhead for join/leave operations of peers while keeping the scale-freeness in a topology with a hard cutoff. Another issue to investigate is the heterogeneity of peers in terms of the load they are willing to undertake. In our study, to observe the fundamental effects of hard cutoffs, we considered the baseline case where all peers apply a fixed hard cutoff. In practice, to achieve better control and search efficiency, many P2P protocols devise superpeers, which are designated for high loads. This situation can be reflected in the hard cutoff distribution by assuming a different hard cutoff for each peer. Such a study will require inference of realistic hard cutoff distributions.

Acknowledgments

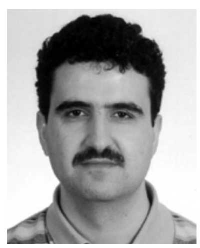
This work was supported by the US Department of Energy (DOE) under Contract DE-AC52-06NA25396 and by the US National Science Foundation under Award 0627039. The authors would like to thank Sid Redner and György Korniss for their feedback.

References

[1] J. Kleinberg, "Navigation in a Small World," *Nature,* p. 845, 2000.
[2] K.Y.K. Hui, J.C.S. Lui, and D.K.Y. Yau, "Small-World Overlay P2P Networks: Constructing and Handling of Dynamic Flash Crowd," *Computer Networks,* vol. 50, pp. 2727-2746, 2006.
[3] A. Iamnitchi, M. Ripeanu, and I. Foster, "Small-World File-Sharing Communities," *Proc. IEEE INFOCOM,* 2004.
[4] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science,* vol. 286, p. 509, 1999.
[5] R. Albert and A.-L. Barabási, "Topology of Evolving Networks: Local Events and Universality," *Physical Rev. Letters,* vol. 85, p. 5234, 2000.
[6] The Napster Home Page, http://www.napster.com, 2008.

[7] I. Stoica, R. Morris, D. Karger, H. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," *Proc. ACM SIGCOMM,* 2001.
[8] P. Maymounkov and D. Mazieres, "Kademlia: A Peer-to-Peer Information System Based on the XOR Metric," *Proc. First Int'l Workshop Peer-to-Peer Systems (IPTPS),* 2002.
[9] A. Rowstron and P. Druschel, "Past: A Large-Scale, Persistent Peer-to-Peer Storage Utility," *Proc. 18th ACM Symp. Operating Systems Principles (SOSP),* 2001.
[10] R. Cox, A. Muthitacharoen, and R. Morris, "Serving DNS Using a Peer-to-Peer Lookup Service," *Proc. First Int'l Workshop Peer-to-Peer Systems (IPTPS),* 2002.
[11] D. Kato, "GISP: Global Information Sharing Protocol—A Distributed Index for Peer-to-Peer Systems," *Proc. Second Int'l Conf. Peer-to-Peer Computing (P2P),* 2002.
[12] K. Sripanidkulchai, B. Maggs, and H. Zhang, "Efficient Content Location Using Interest-Based Locality in Peer-to-Peer Systems," *Proc. IEEE INFOCOM,* 2003.
[13] E. Cohen and S. Shenker, "Replication Strategies in Unstructured Peer-to-Peer Networks," *Proc. ACM SIGCOMM,* 2002.
[14] Gnutella Home Page, http://www.gnutella.wego.com, 2008.
[15] C. Gkantsidis, M. Mihail, and A. Saberi, "Random Walks in Peer-to-Peer Networks," *Proc. IEEE INFOCOM,* 2003.
[16] C. Gkantsidis, M. Mihail, and A. Saberi, "Hybrid Search Schemes for Unstructured Peer-to-Peer Networks," *Proc. IEEE INFOCOM,* 2005.
[17] J. Xu, "On the Fundamental Tradeoffs between Routing Table Size and Network Diameter in Peer-to-Peer Networks," *Proc. IEEE INFOCOM,* 2003.
[18] M. Srivatsa, B. Gedik, and L. Liu, "Large Scaling Unstructured Peer-to-Peer Networks with Heterogeneity-Aware Topology and Routing," *IEEE Trans. Parallel and Distributed Systems,* vol. 17, no. 11, pp. 1277-1293, Nov. 2006.
[19] The Freenet Project, http://freenetproject.org, 2008.
[20] E. Cohen, A. Fiat, and H. Kaplan, "Associative Search in Peer-to-Peer Networks: Harnessing Latent Semantics," *Proc. IEEE INFOCOM,* 2003.
[21] B.F. Cooper and H. Garcia-Molina, "Ad Hoc, Self-Supervising Peer-to-Peer Search Networks," *ACM Trans. Information Systems,* vol. 23, no. 2, pp. 169-200, Apr. 2005.
[22] D. Leonard, V. Rai, and D. Loguinov, "On Lifetime-Based Node Failure and Stochastic Resilience of Decentralized Peer-to-Peer Networks," *Proc. ACM SIGMETRICS '05,* pp. 26-37, 2005.
[23] D. Ilie and A. Popescu, "Statistical Models for Gnutella Signaling Traffic," *Computer Networks,* vol. 51, no. 17, pp. 4816-4835, Apr. 2007.
[24] S. Gribble, S. Ratnasamy, S. Shenker, K. Gummadi, R. Gummadi, and I. Stoica, "The Impact of DHT Routing Geometry on Resilience and Proximity," *Proc. ACM SIGCOMM,* 2003.
[25] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *Computer Comm. Rev.,* vol. 29, p. 251, 1999.
[26] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the World Wide Web," *Nature,* vol. 401, p. 130, 1999.
[27] A.-L. Barabási et al., "Evolution of the Social Network of Scientific Collaborations," *Physica A,* vol. 311, p. 590, 2002.
[28] H. Ebel, M.-I. Mielsch, and S. Bornholdt, "Scale-Free Topology of E-mail Networks," *Physical Rev. E,* vol. 66, p. 035103(R), 2002.
[29] R. Albert, H. Jeong, and A.-L. Barabási, "Error and Attack Tolerance of Complex Networks," *Nature,* vol. 406, p. 378, 2000.
[30] G. Korniss, "Synchronization in Weighted Uncorrelated Complex Networks in a Noisy Environment: Optimization and Connections with Transport Efficiency," *Physical Rev. E,* vol. 75, p. 051121, 2007.
[31] Z. Toroczkai and K.E. Bassler, "Jamming Is Limited in Scale-Free Systems," *Nature,* vol. 428, p. 716, 2004.
[32] J.C. Doyle et al., "The "Robust yet Fragile" Nature of the Internet," *Proc. Nat'l Academy of Sciences of the USA,* vol. 102, p. 14497, 2005.
[33] R. Cohen and S. Havlin, "Scale-Free Networks Are Ultrasmall," *Physical Rev. Letters,* vol. 90, p. 058701, 2003.
[34] S.N. Dorogovtsev and J.F.F. Mendes, "Evolution of Networks," *Advances in Physics,* vol. 51, p. 1079, 2002.
[35] M. Boguná, R. Pastor-Satorras, and A. Vespignani, "Cut-Offs and Finite Size Effects in Scale-Free Networks," *European Physical J. B,* vol. 38, p. 205, 2004.
[36] P.L. Krapivsky and S. Redner, "Organization of Growing Random Networks," *Physical Rev. E,* vol. 63, p. 066123, 2001.
[37] R.L. Graham, D.E. Knuth, and O. Patashnik, *Concrete Mathematics: A Foundation for Computer Science,* second ed. Addison-Wesley, 1994.
[38] R. Albert and A.-L. Barabási, "Bose-Einstein Condensation in Complex Networks," *Physical Rev. Letters,* vol. 86, p. 5632, 2001.
[39] M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence," *Random Structures & Algorithms,* vol. 6, p. 161, 1995.
[40] M. Catanzaro, M. Boguná, and R. Pastor-Satorras, "Generation of Uncorrelated Random Scale-Free Networks," *Physical Rev. E,* vol. 71, p. 027103, 2005.
[41] J. Dall and M. Christensen, "Random Geometric Graphs," *Physical Rev. E,* vol. 66, p. 016121, 2002.
[42] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker, "Search and Replication Strategies in Unstructured Peer-to-Peer Networks," *Proc. ACM Int'l Conf. Supercomputing (ICS),* 2002.
[43] L.A. Adamic, R.M. Lukose, A.R. Puniyani, and B.A. Huberman, "Search in Power-Law Networks," *Physical Rev. E,* vol. 64, p. 046135, 2001.

Hasan Guclu received BS and MS degrees in physics from Middle East Technical University, Ankara, Turkey, in 1998 and 2001, respectively, and the PhD degree in physics from Rensselaer Polytechnic Institute, Troy, New York, in 2005. He is an assistant professor in the School of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York. Formerly, he was a postdoctoral fellow at the Complex Systems Group of Los Alamos National Laboratory, Los Alamos, New Mexico. His research is on complex networks and their applications in peer-to-peer/sensor networks, computational epidemiology, and parallel and distributed computation systems. He is a member of the IEEE, the APS, Sigma Xi, and Sigma Pi Sigma.



Murat Yuksel received the BS degree in computer engineering from Ege University, Izmir, Turkey, and the MS and PhD degrees in computer science from Rensselaer Polytechnic Institute in 1999 and 2002, respectively. He is an assistant professor at the University of Nevada, Reno. His research is on various networking issues such as wireless routing, free-space-optical mobile ad hoc networks (FSO-manet), network modeling and economics, peer-to-peer, protocol design, and performance analysis. He is a member of the IEEE, the ACM, and the Sigma Xi.