Performance Evaluation of Advanced Routing Algorithms for Unstructured Peer-to-Peer Networks

Michele Amoretti
Distributed Systems Group
Information Technology
Departement
University of Parma (Italy)
amoretti@ce.unipr.it

Francesco Zanichelli
Distributed Systems Group
Information Technology
Departement
University of Parma (Italy)
zanichelli@ce.unipr.it

Gianni Conte
Distributed Systems Group
Information Technology
Departement
University of Parma (Italy)
conte@ce.unipr.it

ABSTRACT

Peer-to-peer systems have recently emerged to address the problem of enabling the virtualization of distributed resources such as processing, network bandwidth and storage capacity, to create a single system image, granting users and applications seamless access to vast IT capabilities. Participants in peer-to-peer networks are not only potential consumers but also potential resource providers, and operate autonomously with no central authority. Efficient resource sharing and discovery mechanisms are both essential for the functioning of the system as a whole and for the benefit of all participants.

This paper illustrates our contributions to the characterization of unstructured peer-to-peer architectures, in which the overlay network topology and the adopted routing strategy are not deterministically correlated. Starting from classic analytical results in the field of random graphs, we introduce several novel topological models which put the emphasis on capturing the network growth, and that in our view are very significant for peer-to-peer systems. Moreover, we introduce a novel routing algorithm called HALO, which has been compared to the SRDI strategy adopted by JXTA. Simulation results of HALO and JXTA performance are provided for different overlay network topologies.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems; C.4 [**Performance of Systems**]

1. INTRODUCTION

Peer-to-peer has emerged as a promising new paradigm for distributed computing, aiming at efficient workload distribution and high resource availability. The main idea behind the peer-to-peer paradigm is that each peer, *i.e.* each participant, can act both as a client and as a server *in the context of some application.* Resource sharing and collaboration among peer are the distinguishing properties fo peer-to-peer applications.

A peer-to-peer architectural model is characterized by an overlay network topology and a routing strategy. If these aspects are correlated by a deterministic logical model, we say that the architecture is *structured.* This is the case, for example, of a peer-to-peer system in which nodes and message identifiers are taken from the same space, the overlay network topology is a tree, and propagation is based on choosing the neighbor whose identifier is most similar to the message identifier. On the other hand, if there is no deterministic logical model, we say that the architecture is *unstructured.* Moreover, topologies can be classified in: centralized, partially centralized (hybrid), and decentralized (pure), according to the taxonomy proposed, referring to physical networks, by Paul Baran in the early sixties [9].

In this paper we focus on the characterization of unstructured peer-to-peer architectures, in which the overlay network topology and the adopted routing strategy are not deterministically correlated. Nevertheless, if the overlay network topology can be estimated in advance, the routing strategy should be adapted to it. The key to scalable searches in unstructured networks is to find the best direction and to cover the path as quickly as possible, and with as little overhead as possible. To achieve these goals, the routing algorithm designer should pay attention to adaptive termination, message duplication and coverage granularity. An almost complete classification of peer-to-peer architectural models has been provided by Lloret [17]. In this work we consider the Selective Query Model (SQM), which leads to unstructured supernode networks, *i.e.* overlay architectures in which peers with higher bandwidth and process capacity act as supernodes, assuming the responsibility of propagating messages, while peers with less capacities (leaf nodes) connect to supernodes and send them publication/query messages, but do not contribute to the overall routing process. The model is illustrated in figure 1.

Based on the SQM model, FastTrack [2] is a protocol used by applications like KaZaA Media Desktop (shortly, KaZaA) [3], and Skype [4]. An open-source SQM-based peer-to-peer system is JXTA [20], a project originally conceived by Sun Microsystems, and designed with the participation of a growing number of experts from academic institutions and industry.

Valuetools'06 October 11-13, 2006, Pisa, Italy.
Copyright Number: 1-59593-504-5

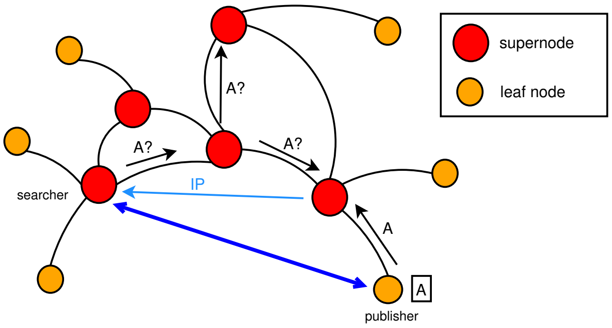

Figure 1: The SQM peer-to-peer architectural model.

The main goal of JXTA is to define a generic peer-to-peer network overlay usable to implement a wide variety of applications and services. The JXTA platform provides core building blocks (IDs, advertisements, peergroups, pipes) and a default set of core policies, which can be replaced if necessary.

The paper is organized as follows. Section 2 illustrates several novel network topologies, which put the emphasis on capturing the network growth, and that in our view are very significant for peer-to-peer systems. Two routing strategies are described in section 3: SRDI, which is the default solution provided by JXTA, and HALO, which is the strategy we defined aiming at good efficiency and low search cost. We simulated JXTA and HALO searches in both classic and novel network topologies, and we obtained many interesting performance results which are illustrated in section 4. Finally, an outline of open issues concludes the paper.

2. GROWTH-BASED NETWORK TOPOLOGY MODELS

The formalism we adopt all over this section comes from the theory of random graphs [11], one of the youngest branches of graph theory. Its great strength is that it uses probabilistic methods to demonstrate the existence of the desired graphs without constructing them. In this context, we depict each network as an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is the set of nodes of the network and $\mathcal{L}$ is the set of links. The number of nodes is $N = |\mathcal{N}|$. The number of links is $L = |\mathcal{L}|$, but it is not important as the *node degree*, which is the number of links starting from a node. The node degree of a network is described in terms of *Probability Mass Function (PMF)*

$$P(k) = P\{\text{node degree} = k\}$$

The first and most investigated random network model has been introduced by Erdös and Rényi (*ER model*). Networks based on the ER model have N vertices, each one connected to an average of $\alpha = \langle k \rangle$ nodes. The presence or absence of a link between two vertices is independent of the presence or absence of any other link, thus each link can be considered to be present with independent probability p. For large N, the degree distribution of the network converges to the Poisson distribution

$$P(k) = \frac{\alpha^k e^{-\alpha}}{k!} \qquad \text{with } \alpha = \langle k \rangle = \sigma^2 \tag{1}$$

An example of Poisson distribution of the node degree is illustrated in figure 2.

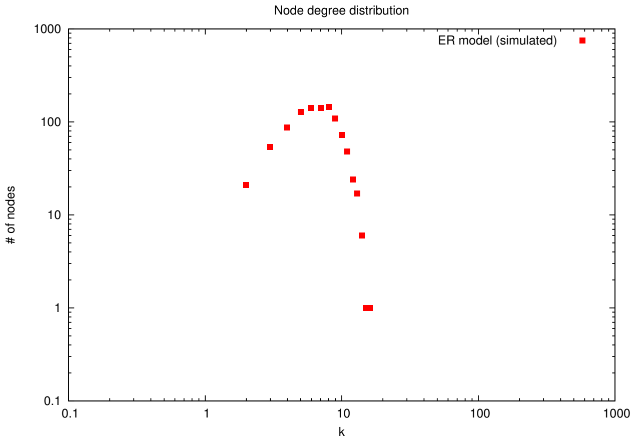

Figure 2: Node degree distribution for a simulated Poisson random network of $N = 1000$ nodes, with average node degree $\alpha = 7$ (in log scales).

Poisson networks are fairly homogeneous, *i.e.* each node has approximatively the same number of links. In contrast, studies about the World Wide Web, the Internet, and other large networks indicate that these systems belong to a class of inhomogeneous networks, for which the node degree PMF decays as *power law* [19, 18] distribution:

$$P(k) = ck^{-\tau} \tag{2}$$

with $\tau > 1$ (to be normalizable), and

$$c = [\sum_{k=1}^{\infty} k^{-\tau}]^{-1} = [\zeta(\tau)]^{-1} \tag{3}$$

where $\zeta(\cdot)$ is the Riemann zeta function. These networks are called *scale-free*, because their separation degree growth is sublinear with respect to N. In particular, if $2 < \tau < 3$, the diameter is $\langle l \rangle \sim \ln \ln N$ [13]. Even if the number of nodes strongly increases, the mean distance between two nodes remains the same.

Barabási and Albert proposed a simple model (called *BA model*) [7, 8, 6] to construct scale-free networks with $\tau \simeq 3$. The BA model is based on two ingredients: growth (*i.e.* N should not be fixed in advance), and preferential attachment (*i.e.* the probability with which a new node connects to the existing nodes is not uniform as in Poisson random networks). For example, in the actor's network, a new actor is casted more likely in a supporting role, with more established, well-known actors. Figure 3 illustrates the node degree distribution of network based on the BA model.

The BA model incorporates the growing character of the network, starting with a small number N_0 of connected nodes, by adding a new node at every step, with $m \leq N_0$

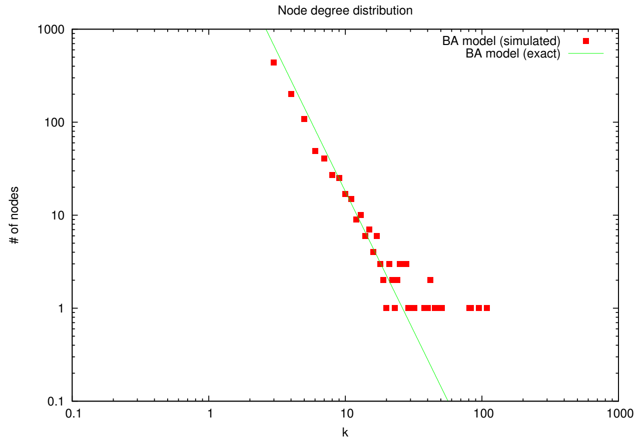

Figure 3: BA model based network of $N = 1000$ nodes, with $m = 3$ and $N_0 = 5$.

links to different old nodes of the system. To incorporate preferential attachment, the probability $\Pi(k, N)$ that the $(N + 1)$-th node will be connected to any node with degree k is assumed to be dependent on the node degree k of that node, so that

$$\Pi(k, N) = \frac{k}{\sum_{j=1}^{N} k_j} \simeq \frac{k}{2mN}$$

Typically, most connected nodes are stable and recognized, thus it is realistic to assume that new nodes join them with high probability. By the way, the network is once again described in terms of probability of connecting to a node, without giving the strategy followed by each peer.

The most elegant way of deriving the node degree distribution of a network constructed with the BA model, is the *master-equation approach* proposed by Dorogovtsev and Mendes [14]. In general, this approach defines the rate of change of the probability of occurrence of a state of the system:

$$\frac{dP_i}{dt} = \sum_j [W_{ij}P_j - W_{ji}P_i] \tag{4}$$

where P_i is the probability of finding the system in state i at time t, and W_{ji} and W_{ij} are the transition rates for changes from state j to i and from state i to j, respectively. In the case of a network, the k-th state is “having k neighbors”, and in particular for the BA model the transition rate is:

$$m\Pi(k, N) = \frac{k}{2N} \tag{5}$$

Thus the master equation is

$$p_{k,N+1} - p_{k,N} = \frac{k-1}{2N} p_{k-1,N} - \frac{k}{2N} p_{k,N} \tag{6}$$

Looking for stationary solutions $p_{k,N+1} = p_{k,N} = P(k)$, expression 6 gives the recursive equation

$$P(k) = \begin{cases} 1 - \frac{1}{2} m P(m) & \text{if } k = m \\ \frac{1}{2}(k-1)P(k-1) - \frac{1}{2}(k)P(k) & \forall k > m \end{cases}$$

which can be rearranged, giving

$$P(k) = \begin{cases} \frac{2}{m+2} & \text{if } k = m \\ \frac{k-1}{k+2} P(k-1) & \forall k > m \end{cases} \tag{7}$$

and finally

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \simeq 2m^2 k^{-3} \quad \forall k \geq m \tag{8}$$

which is a power law with exponent $\tau = 3$.

There is a fundamental difference between the modeling approach we took for Poisson networks, and the one required to reproduce the power law degree distribution. While the goal of the former models is to construct a graph with correct topological features, the modeling of scale-free networks puts the emphasis on capturing the dynamics of the network growth. According to this principle, in the following we define seven models which are related to peer-to-peer systems, in which the construction of the overlay network can be driven to obtain particular regularities. The disadvantage of these novel models is the difficulty of describing them with analytical techniques such the one we used for the BA model, in particular for what concerns the probability $\Pi(k, N)$ that the $(N + 1)$-th node will be connected to a node with degree k.

2.1 Locally Preferential Networks

The *Locally Preferential (LP)* model assumes that each new node joins the network having a precise view of only n_0 already existing nodes (we call this set *initial peerview*), and choosing the $m \leq n_0$ best connected ones. The network grows around $N_0 \ll N$ initial nodes which form a complete graph. The LP growth process is illustrated in figure 4. Considering n_0 and m as fixed constants, it is still possible to use the master-equation approach, but it is difficult to find the analitical solution. The probability $\Pi(k, N)$ that the $(N + 1)$-th node will be connected to any node with degree k is

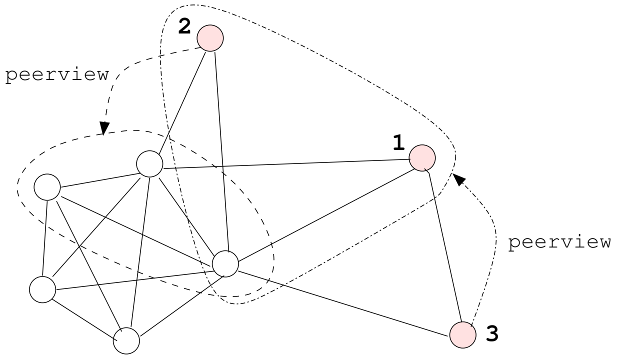

Figure 4: LP growth process with $m = 2$, $n_0 = 3$ and $N_0 = 5$.

$$\begin{aligned} \Pi(k, N) &= 1 - \mathrm{P}\{\text{no peer of the chosen } m \text{ has } k \text{ links}\} \\ &= 1 - [\frac{\binom{pN}{m}\binom{qN}{n_0-m}}{\binom{N}{n_0}} + \frac{\binom{p'N}{m}\binom{q'N}{n_0-m}}{\binom{N}{n_0}}] \end{aligned}$$

where

$$\begin{aligned}
p &= \mathrm{P}\{\text{having} > k \text{ links}\} = \sum_{j=k+1}^{\infty} P(j) \\
q &= \mathrm{P}\{\text{having} \leq k \text{ links}\} = 1 - p = \sum_{j=1}^{k} P(j) \\
p' &= \mathrm{P}\{\text{having} < k \text{ links}\} = q - P(k) = 1 - p - P(k) \\
q' &= \mathrm{P}\{\text{having} \geq k \text{ links}\} = p + P(k) = 1 - q + P(k)
\end{aligned}$$

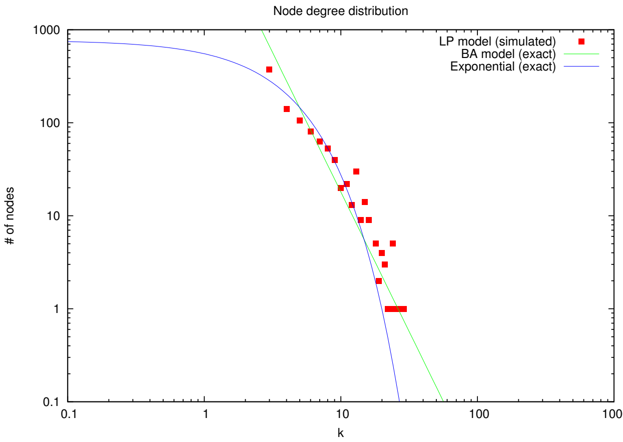

Figure 5: Node degree distribution of a network generated with the LP model, with $N = 1000$, $m = 3$, $n_0 = 5$ and $N_0 = 5$.

Figure 5 illustrates the node degree distribution of a network generated with our simulator (see section 4) using the LP model. This result and those following (figures 8, 9 and 10) are compared to the exponential degree distribution

$$P(k) = e(1 - e^{-\frac{1}{m}}) \exp\left(-\frac{k}{m}\right) \quad \forall k > m \qquad (9)$$

which characterizes a network growing without preferential attachment, and also to the degree distribution generated by the BA model (eq. 8). The result of figure 5 shows that the node degree distribution of a LP network is in between the exponential and the scale-free. Unfortunately, the LP model does not fully respect the construction of a real peer-to-peer network, because the initial peerview is different for each joining peer, and the probability that two peers have the same initial peerview decreases as N increases.

2.2 Locally Preferential Random Networks

An extension of the LP model is the *Locally Preferential Randomized (LPR)* model, in which n_0 and m are not the same for each peer. In details, the peerview size of each joining node is random in $[1, N_0]$. Moreover, the number m of nodes to which each joining node connects is random in $[1, n_0]$. The LPR growth process is illustrated in figure 6. The resulting network has N_0 classes of peers, one for each possible value of the initial peerview size. The n-th class ($n \in [1, N_0]$) has n subclasses, one for each possible value of $m \in [1, n_0]$. Thus n_0 and m are random processes, *i.e.* their values are randomly different for each new node joining the network. For large values of N, we can assume these processes to be stationary.

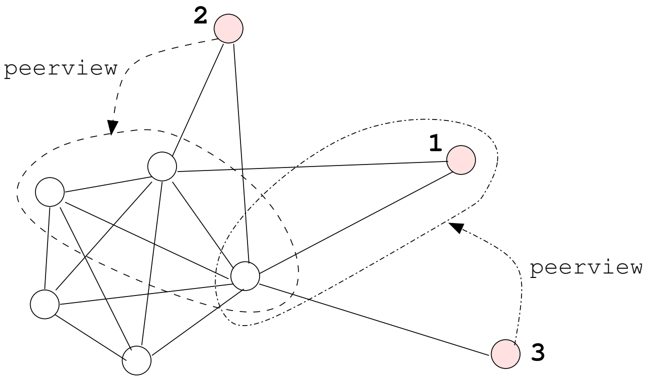

Figure 6: LPR growth process with $N_0 = 5$ (m and n_0 are random).

In particular we consider n_0 and m to be uniform random variables, thus:

$$\langle n_0 \rangle = \frac{N_0}{2}$$

$$\begin{aligned}
\langle m \rangle &= \sum_{m=1}^{N_0} m P(m) = \sum_{m=1}^{N_0} m \sum_{n=1}^{N_0} P(m|n_0 = n) P(n_0 = n) \\
&= \sum_{m=1}^{N_0} m \sum_{n=1}^{N_0} \frac{1}{n} \frac{1}{N_0} = \frac{N_0(N_0+1)}{2N_0} \sum_{n=1}^{N_0} \frac{1}{n} \\
&\simeq \frac{N_0+1}{2} \ln N_0
\end{aligned}$$

Moreover, the probability of being in a class is $\frac{1}{N_0}$ for each class, and the mean size of a class is $\frac{N}{N_0}$. The probability of being in a subclass of the n-th class is $\frac{1}{n}$, and the mean size of a subclass of the n-th class is $\frac{N}{N_0 n}$.

For this kind of system it is not possible to use the master-equation approach, because transition rates W_{ij} are time-dependent random variables, *i.e.* random processes. In general, if $W_{ij}(\infty) \neq 0$ and the Markov process defined by the master equation has a finite number of states and it is irreducible (*i.e.* all the configurations can be reached from a given one by a string of transitions with nonzero probability), all solutions of the master equation converge [12]. But since $W_{ij}(\infty)$ for this kind of network is undefined, it makes no sense to write the master equation.

The LPR model is based on more realistic dynamics than those of the LP model, assuming variable initial peerview size and number of peers chosen for connection.

2.3 Seeded Networks

We observed that both LP and LPR models do not fully respect the construction of a real peer-to-peer network, because the initial peerview is different for each joining peer, and the probability that two peers have the same initial peerview decreases as N increases. These problems can be neglected if we construct the network as a cluster of clusters, *i.e.* if there are S non-overlapping groups of peers, each of them starting with N_0 completely connected peers. Each group has $n(i)$ peers ($i = 1, .., S$), with $\sum_i n(i) = N$. Groups are not isolated clusters because each one has $s(i)$ *seed peers* among the initial N_0 members, forming a complete graph with all other seed peers. The resulting node degree distribution has a Poisson queue, which is the contribution of seed nodes. This is what we espect to have in peer-to-peer seeded networks, such as JXTA.

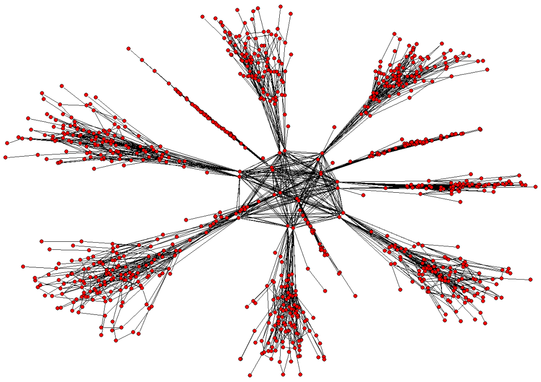

Figure 7: Graphical representation of a seeded network with $N = 1000$ nodes, and $S = 10$ groups.

2.4 Seeded ER Networks

Based on this strategy is the *Seeded ER (SER)* model, in which each node in a group is connected to an average of α nodes. The presence or absence of a link between two nodes is independent of the presence or absence of any other link, thus each link can be considered to be present with independent probability p. If nodes are independent, the degree distribution of the network is binomial:

$$P(k) = \binom{n(i) - 1}{k} p^k (1 - p)^{n(i)-1-k} \qquad (10)$$

which for large $n(i)$ converges to the Poisson distribution (eq. 1), as illustrated in figure 8.

2.5 Seeded BA Networks

The seeding strategy applied to the BA model leads to the *Seeded BA (SBA)* model, in which each new node joining a group connects to $m \leq N_0$ different old nodes of the group, with preferential attachment. Each group is characterized by a power law degree distribution with exponent $\tau = 3$ (remember eq. 8). Figure 9 illustrates that the node degree distribution of a network generated using the SBA model is very similar to the scale-free distribution generated by the BA model, with the exception of the Poisson queue which is the common characteristic of all seeded networks.

2.6 Seeded Locally Preferential Networks

The seeding strategy applied to the LP model leads to the *Seeded Locally Preferential (SLP)* model, in which each peer joining a group has a limited peerview, with fixed size $n_0 \in [1, N_0]$. Moreover, the joining peer choses the m best connected nodes in its peerview, with $m \leq n_0$ fixed.

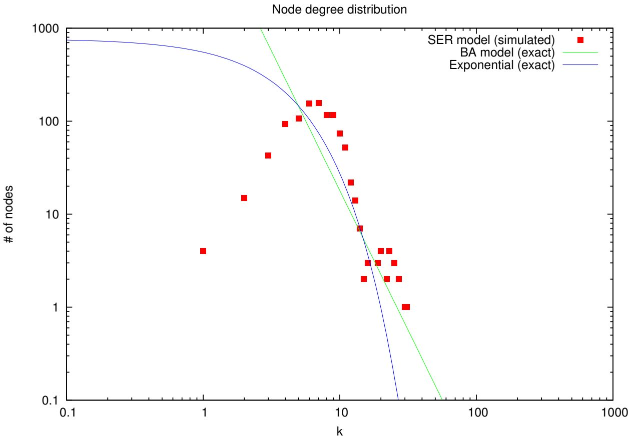

Figure 8: Network generated with the SER model, with $N = 1000$, $s(i) = S_0 = 2$ and $\alpha = 7$. Each group is assumed to have no more than $N_S = 150$ members.

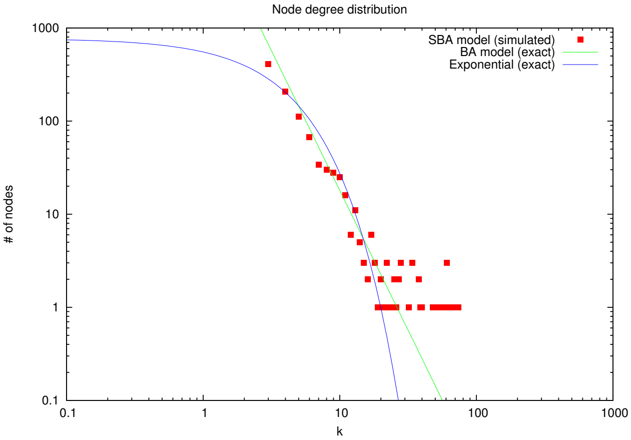

Figure 9: Network generated with the SBA model, with $N = 1000$, $s(i) = S_0 = 2$, $m = 3$ and $N_0 = 5$. Each group is assumed to have no more than $N_S = 150$ members.

Figure 10 illustrates a SLP distribution, which is similar to the LP distribution in figure 5, with the exclusion of the Poisson queue which is typical of seeded networks.

2.7 Seeded Locally Preferential Random Networks

The seeding strategy applied to the LPR model leads to the *Seeded Locally Preferential Randomized (SLPR)* model, in which each peer joining a group has a limited peerview, with uniformly random size $n_0 \in [1, N_0]$. Moreover, the number m of nodes to which each joining node connects is uniformly random in $[1, n_0]$. Finally, the m chosen nodes are the best connected ones among those in the peerview.

3. ADVANCED ROUTING ALGORITHMS IN SQM-BASED NETWORKS

Peers in SQM-based networks can be divided into leaf nodes and supernodes. Each leaf node is connected to a (single)

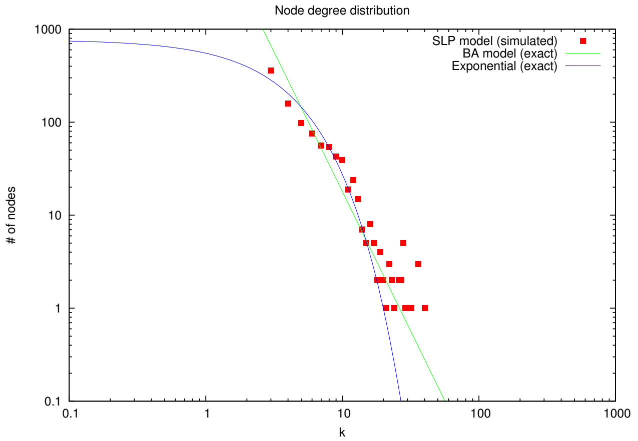

Figure 10: Network generated with the SLP model, with $N = 1000$, $s(i) = S_0 = 2$, $m = 3$, $n_0 = 5$ and $N_0 = 5$. Each group is assumed to have no more than $N_S = 150$ members.

supernode. On the other side, supernodes maintain many leaf connections, as well as a small number of connections to other supernodes. This hierarchy leads to scalable systems, in which supernodes shield leaf nodes from virtually all ping and query traffic.

In the following we illustrate two routing strategies for SQM-based peer-to-peer networks, namely SRDI, adopted by JXTA [20], as well as our HALO algorithm.

3.1 JXTA Routing

Message routing for resource sharing and discovery in JXTA networks is based on two components: the *Shared Resource Distributed Index (SRDI)*, and the *loosely-consistent DHT walker*. The SRDI module implemented in each JXTA peer is used on one hand to extract entries from resource advertisements and push them to the network, and on the other hand for lookups. The walker is used for routing when no index information is locally available. The process is illustrated in figure 11, for both the ideal case of complete supernode network, and the more common case of uncomplete supernode network.

JXTA supernodes are called rendezvous super-peers, while leaf nodes are called edge peers. Resources, services, peers and peergroups are described by XML documents, the so-called advertisements. When an advertisement is published by an edge peer (E1), its entries (which are attribute-value pairs) are sent to the connected supernode (R1). Supernodes store the entries in dynamic indexes including also, for each entry, the ID of the peer which originated them and an expiration time. Moreover, each index entry is replicated from supernode R1 to another supernode in R1's RPV using a DHT function. In details, the 160 bit SHA1 hash address space is evenly divided amongst the ordered RPV (sorted by peer ID), so an index entry is routed by hashing its value and mapping its location in the RPV (suppose R11). If the RPV is > 3, each index entry is also replicated to the RPV neighbors of R11 ($+1$ and -1 in the RPV ordered circular list).

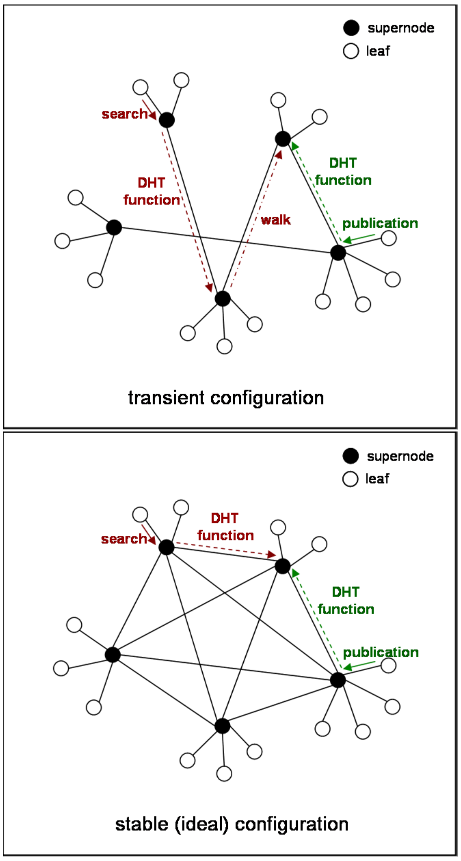

Figure 11: JXTA routing with the SRDI strategy.

Search is based on the same DHT function, but also on a limited range walker to resolve inconsistency of the DHT within the dynamic rendezvous network. Queries are messages which contain advertisement entries (attribute-value pairs). If a query is sent by an edge peer (E2), it reaches the connected supernode (R2). Once R2 receives the query, it first attempts to match it locally. If a match is not found, then R2 forwards the query to another supernode in R2's RPV (suppose R22) using the DHT function. The more the rendezvous network is near to completeness, *i.e.* the RPV is consistent across all supernodes, the more the DHT-based routing algorithm is efficient. To compensate for any RPV skew, a *limited range walker* is used. For example, suppose R22 fails to match the query, and its RPV is

R20 R21 *R22* R23

Assume that R20 and R21 have the same RPV of R22, while R23 has the following RPV:

R23 R24 R25

Thus an R22 originated limited range walker query is walked to R21 with a TTL of 2, and to R23 with a TTL of 1, where the TTL is adjusted to 2 on R23 and walked to R25. When there is a query hit, the response is forwarded to the query originator (in this case, edge E2).

The more the supernode network is near to completeness, *i.e.* the peerview is consistent across all supernodes, the more the routing algorithm is efficient. The strategy is summarized in algorithm 1.

```
if (publication) then
    save locally
end if
if (search) then
    match locally
end if
if (leaf peer) then
    send message to supernode
end if
if (supernode) then
    if ((message from leaf peer) || (local message)) then
        find target supernode neighbor $t$ using DHT function
        send message to target $t$ supernode neighbor
        if ((publication) && (peerview > 3)) then
            send message to $t+1$ and $t-1$ supernode neighbors
        end if
    end if
    if (message from supernode) then
        if (search) then
            walk the peerview
        end if
    end if
end if
```

Algorithm 1: JXTA message routing.

3.2 HALO Routing

The novel approach we propose, called *HALO*, is based on high-degree node search, *i.e.* messages are routed choosing at each step the highest-degree neighbor, and using the DHT function for corrections and for the final hop. The idea for this strategy comes from the observation that the pseudo-DHT strategy adopted by JXTA gives its best when the supernode network is highly connected. HALO routes messages towards best connected nodes, and uses the same DHT function used by JXTA if neighbors are less connected than current peer, and for the final step. If many neighbors have the same highest node degree, the target is chosen by proximity of its ID with the message ID. Algorithm 2 summarizes the HALO strategy. Note that search and publishing strategies are exactly symmetrical. The maximum number of hops for a message is TTL.

HALO is different from the algorithm proposed by Adamic *et al.* [5], which after the initial climb towards high degree nodes walks down the degree sequence, and does not define the strategy for choosing between equal degree nodes.

In the following section, we show many simulation results which demonstrate HALO's good performance in scale-free networks and in other typical peer-to-peer topologies, such as SBA and SLP. With such kind of overlay networks, HALO is more efficient than JXTA, which is suited for Poisson or quasi-complete networks.

4. SIMULATION RESULTS

Previously illustrated routing strategies have been evaluated with the help of a new component we developed for the universally known *ns-2* network simulator. This component provides a C++ API implementing basic functionalities such as key management and peer linkage, which are common to all peer-to-peer networks, and specific routing algorithms such as Gnutella, Random Walk, JXTA and HALO; a TCL script library to setup the simulations, by defining the physical network topology, the overlay network topology, and the protocol to simulate; several scripts for log analisys and graphical representation of results.

```
if (publication) then
    save locally
end if
if (search) then
    match locally
end if
if (leaf peer) then
    send message to supernode
end if
if (supernode) then
    if (Hops < $TTL - 1$) then
        search for supernode neighbor with higher degree
        if (found $n \geq 1$ supernode neighbors with higher degree) then
            choose highest degree supernode neighbor with ID $\simeq$ message ID
            send to chosen supernode neighbor
        else
            find target supernode neighbor $t$ using DHT function
            send message to target $t$ supernode neighbor
        end if
    end if
    if (Hops == $TTL - 1$) then
        find target supernode neighbor $t$ using DHT function
        send message to target $t$ supernode neighbor
    end if
end if
```

Algorithm 2: HALO message routing.

Simulations have been performed on a Pentium IV machine with 2.8Ghz cpu, 512KB cache and 1GB ram. The typical problems of large simulations in ns-2, the main constraints being that of memory and cpu-time, are (1) start-up time or the time taken before the actual simulation can start, (2) run-time or how long it takes for the simulation to complete and (3) memory requirement for the simulation. In particular, the most common problem that people face while running large simulations is running out of memory. To evaluate our peer-to-peer simulator, we adopted all known methods and practices to reduce memory usage in ns-2. We also assumed UDP connections between peers, to minimize packet headers' length and thus reduce computation load.

In the following we illustrate the results of several simulations aimed at comparing JXTA and HALO in unstructured supernode networks. Each simulation should be considered as a "snapshot" representing the life of the network for a very short time interval. Constant parameters are:

- $N = 4000$ nodes (including R supernodes and E leaf nodes);
- the physical network topology, generated with BRITE [1]: formed by 100 subnetworks, each one including 40 nodes connected with $10 Mbps$ links, according to the

BA model (with $m = 3$); subnetworks are clustered by nodes which are connected with $1Gbps$ links, according to Waxman model [21] (with $\alpha = 0.15$, $\beta = 0.2$ and $m = 2$); packet losses are considered only for $1Gbps$ links;

- $r = 41245$ resources evenly distributed among all N nodes;
- $q = 1000$ queries for existing resources, evenly distributed among all N nodes; query starting time is exponentially distributed, with mean value $0.05sec$.

To study a wide range of configurations, we used different values for the following parameters:

- the number of supernodes (R);
- the supernode network topology (10 models, from ER to SLPR, each one with its own parameters);
- the TTL value, *i.e.* the maximum number of hops per message.

The efficiency (*i.e.* % of query hits) of JXTA and HALO is not affected by the increasing of R, for every kind of topology, as illustrated in figure 12.

The efficiency of JXTA and HALO, for increasing TTL values and five different kinds of topologies, is compared in figure 13. We can observe that, in every case, there is a small threshold after which JXTA's efficiency does not further increase. This is a good feature if the topology is ER, for which JXTA quickly reaches a good value of efficiency, but it is a bad feature in every other case. On the other side, HALO's efficiency increases as TTL increases, reaching a generally acceptable value when $TTL = \ln R$.

Figure 14 shows the efficiency of JXTA, in ER topologies, for increasing α values, with $R = 1000$ supernodes among $N = 4000$ total nodes in the network. We can observe that JXTA efficiency is acceptable only for $\alpha > 5$, and that packet losses on backbones strongly affects the performance of the system.

Figure 15 shows the efficiency of HALO, in BA topologies, for increasing m values, with $R = 1000$ supernodes among $N = 4000$ total nodes in the network. We can observe that HALO efficiency is unaffected by m variations, and slightly affected by packet losses on $1Gbps$ links, compared to JXTA.

To fully characterize routing algorithms, hops number and time distributions of query hits must be considered. We previously observed that JXTA and HALO show better efficiency respectively with ER and BA topologies. We fixed $R = 1500$ (thus $E = N - R = 2500$) and $TTL = 6$, and we considered an ER topology with $\alpha = 7$, and a BA topology with $m = 3$ and $N_0 = 5$.

The resulting hops count and time distributions have Poisson shape, with different mean value. In particular, JXTA routing is very costly, requiring 16 hops (in the average) for

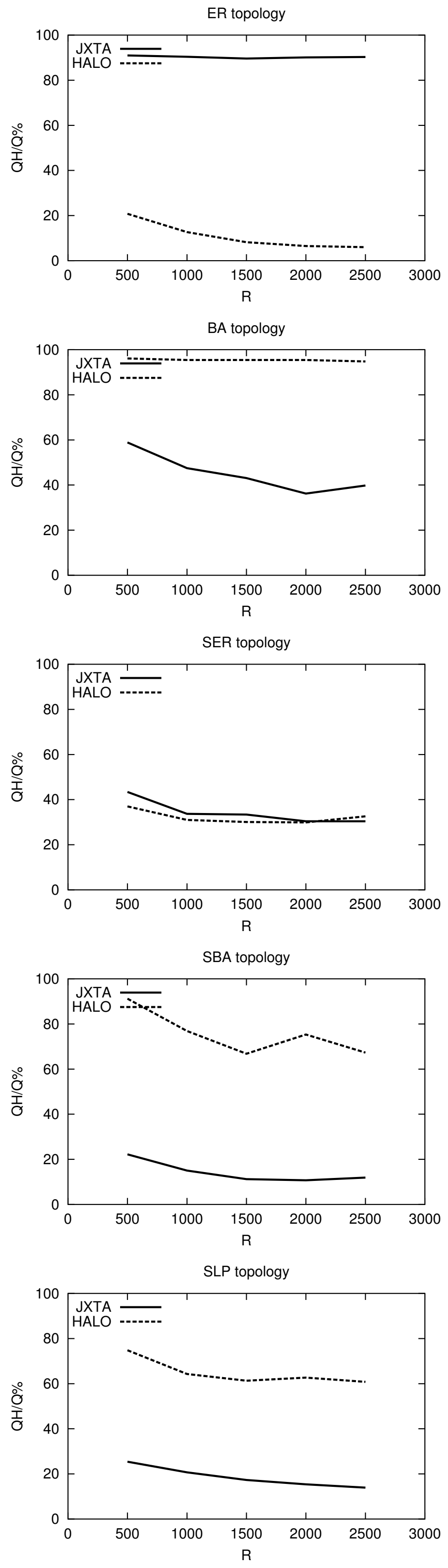

Figure 12: JXTA vs HALO efficiency for different R values. We assumed $R = 1000$ supernodes, among $N = 4000$ total nodes in the network.

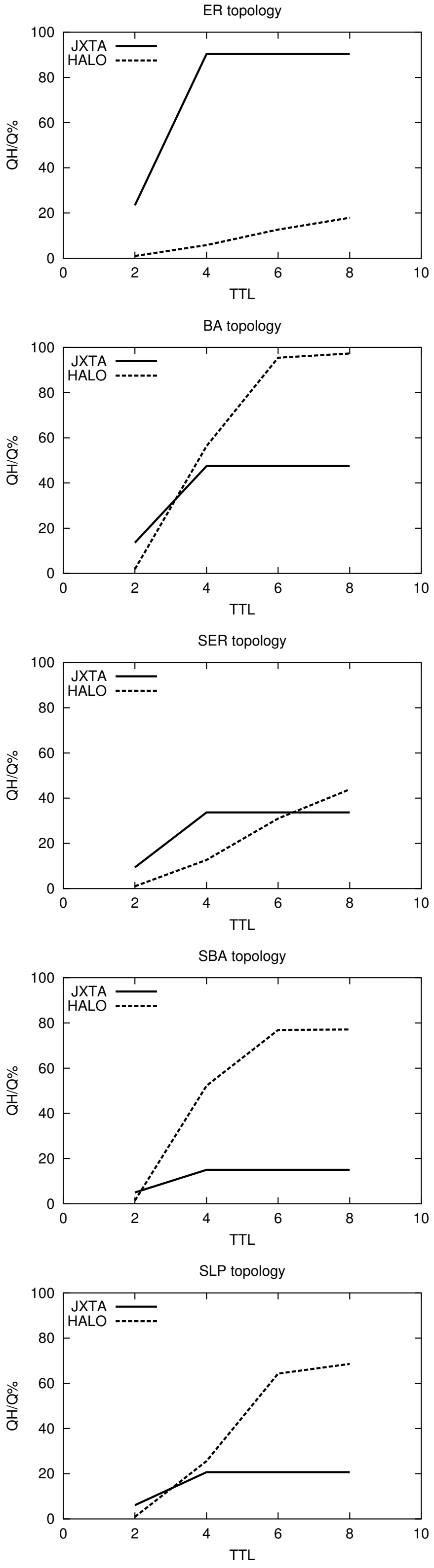

Figure 13: JXTA vs HALO efficiency for different TTL values. We assumed $R = 1000$ supernodes, among $N = 4000$ total nodes in the network.

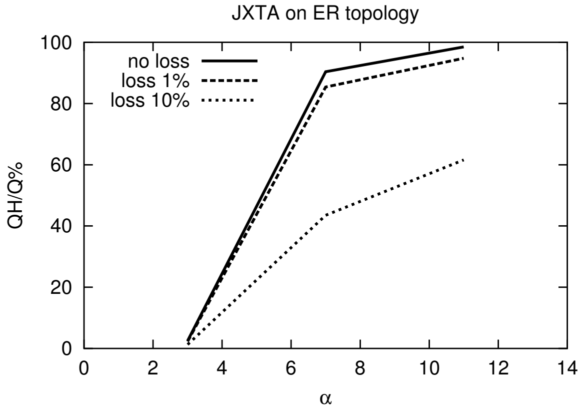

Figure 14: JXTA efficiency in ER topologies for different α values, considering also packet losses on $1Gbps$ links.

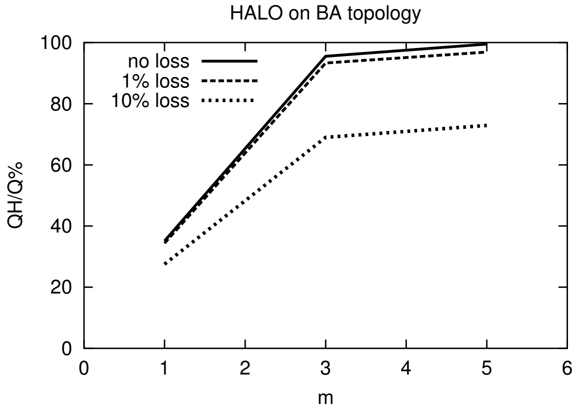

Figure 15: HALO efficiency in BA topologies for different m values, considering also packet losses on $1Gbps$ links.

message. On the other side, HALO queries perform only 3 hops in the average. These results are not surprising, because we know JXTA uses a random walker which frequently re-initializes its TTL, while HALO strictly guarantees a maximum of TTL hops for each message.

Simulating application data transfers in ns-2 is often complicated by the gap between the ns2 implementation of TCP and the real-world socket interfaces provided by most operating systems. The FullTcpAgent provided by ns-2 does not transfer real payloads (user data), although the receiver knows the size of the payload when it receives a TCP packet. This prevents application layer protocols (such as peer-to-peer routing protocols) from being implemented. FullTcpAgent is also not compatible with the behavior of a real TCP socket in that it does not spawn new connection dynamically as TCP connection requests arrive, and that it assumes unlimited TCP send buffer such that the applications are never blocked for sending. We are currently trying to overcome these problems, which would allow to perform the same simulations we presented in this section, but using TCP instead of UDP.

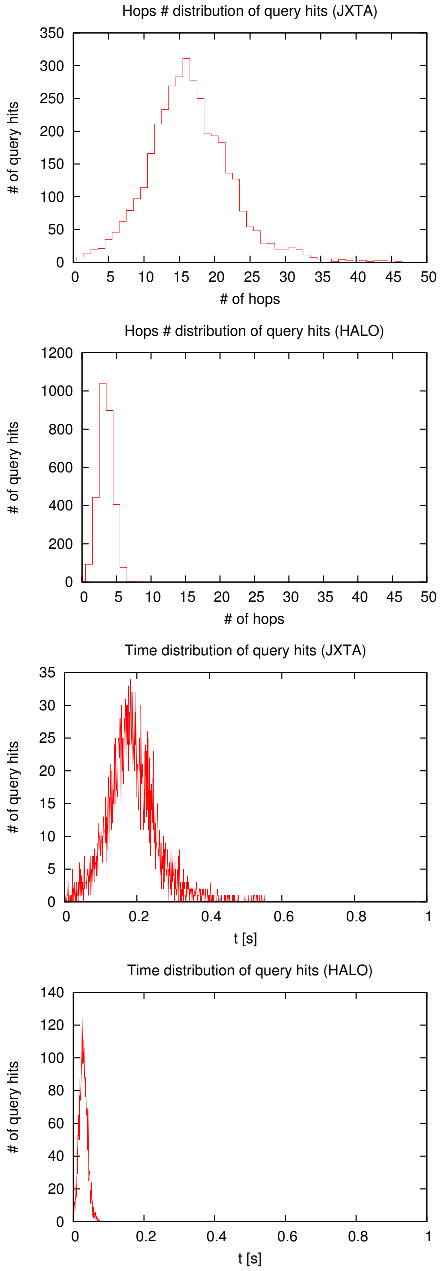

Figure 16: JXTA and HALO hops count and time distribution of query hits.

5. RELATED WORK

First generation unstructured peer-to-peer networks have lost popularity due to their poor scalability. The typical example is Gnutella, a FRM (flooded requests model) protocol which has been analytically studied and simulated by many researchers. In particular we considered [15], which illustrates a ns-2 based package for peer-to-peer newtork simulation, and its utilization to measure Gnutella performance. Before developing our own ns-2 based simulator, we tried to extend that package to implement JXTA and HALO, but our efforts were prevented by several structural limitations of the simulator, and by the substantial indifference of the authors, who exhibited no interest in improving their project.

More recently, simulations of unstructured peer-to-peer networks have been proposed in [10]. The authors share with us the interest for quantifying the performance benefits provided by SQM architectures. In particular, they propose performance metrics for Gnutella 2, which is the supernode-based evolution of Gnutella.

The need for supernode networks takes origin also from the topology mismatch problem between physical and logical network, which affects all peer-to-peer systems. In [16], the authors first analyze the relationship between the property of the overlay and the corresponding message duplications incurred by queries in a given overlay, and prove that computing an optimal overlay with global knowledge is an NP-hard problem. Motivated by the analysis results, they design a distributed overlay optimization algorithm to attack topology mismatch. Its performance is demonstrated by comprehensive simulations in dynamic environments.

6. CONCLUSIONS

The major achievements of this work are related to the characterization of unstructured peer-to-peer architectures. Starting from classical analytical results, we introduced several novel topological models which put the emphasis on capturing the network dynamics, and in our view are very significant for peer-to-peer systems. The formalism we adopted to describe these models comes from the theory of random graphs, one of the youngest branches of graph theory. Its great strength is that it uses probabilistic methods to demonstrate the existence of the desired graphs without constructing them.

Moreover, we proposed HALO, a novel routing algorithm, and we demonstrated its good performance in scale-free networks and in other typical peer-to-peer topologies. With such kind of overlay networks, HALO is more efficient than JXTA, which is suited for Poisson or quasi-complete networks.

Future work will concern a new research topic related to the study of emergent behaviour in peer-to-peer architectures. In this context, growth-based models for network topologies, such as those we presented in this paper, can be considered. The initial challenge will be the simulation and implementation of adaptive routing strategies.

Acknowledgements

This work has been supported by the ”WEB-MINDS” FIRB project of the National Research Ministry. The ns-2 component for peer-to-peer network simulations which has been presented in this work, has been developed in collaboration with R. Lancellotti, M. Andreolini, and P. S. Yu. We would like to thank Veronica Biazzo for the interesting and useful discussions on combinatorics.

7. REFERENCES

[1] Brite homepage. http://www.cs.bu.edu/brite/.

[2] FastTrack. http://en.wikipedia.org/wiki/FastTrack.

[3] KaZaA. http://www.kazaa.com.

[4] Skype. http://www.skype.com.

[5] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman. Search in power-law networks. *Physical Review E*, 64(4):1842–1845, September 2000.

[6] R. Albert and A. Barabási. Statistical mechanics for complex networks. *Reviews of Modern Physics*, 74(1):47–97, January 2002.

[7] A. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999.

[8] A. Barabási, R. Albert, and H. Jeong. Mean-field theory for scale-free random networks. *Physica A*, 272(1-2):173–187, October 1999.

[9] P. Baran. Introduction to Distributed Communications Network. Technical report, RAND Corporation, Aug. 1964.

[10] F. Benevenuto, J. Ismael, and J. Almeida. Quantitative Evaluation of Unstructured Peer-to-Peer Architectures. In *HOTP2P 2004, Volendam, The Netherlands*, October 2004.

[11] B. Bollobás. *Random Graphs*. Academic Press, 1985.

[12] J. Brey and A. Prados. Normal solutions for master equations with time-dependent transition rates: Application to heating processes. *Physical Review E*, 47:1541–1545, 1993.

[13] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *PHYS.REV.LETT*, 90:058701, 2003.

[14] S. Dorogovtsev and J. Mendes. Evolution of networks. *Advances in Physics*, 51:1079, 2002.

[15] Q. He, M. Ammar, G. Riley, H. Raj, and R. Fujimoto. Mapping Peer Behavior to Packet-level Details: A Framework for Packet-level Simulation of Peer-to-Peer Systems. In *MASCOTS 03*, October 2003.

[16] Y. Liu, L. Ni, and A. Esfahanian. Approaching Optimal Peer-to-Peer Overlays. In *MASCOTS 2005, Atlanta, Georgia, USA*, October 2005.

[17] J. Lloret. Interconnecting Unstructured P2P File Sharing Networks. *P2P Journal*, March 2005.

[18] M. Mihail and C. Papadimitriou. On the Eigenvalue Power Law. In *6th International Workshop on Randomization and Approximation Techniques, Cambridge, Massachusetts*, September 2002.

[19] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos. Power-laws and the AS-level Internet Topology. *IEEE/ACM Transactions on Networking*, 11(4):514–524, August 2003.

[20] B. Traversat, A. Arora, M. Abdelaziz, M. Duigou, C. Haywood, J. Hugly, E. Poyoul, and B. Yeager. Project JXTA 2.0 Super-Peer Virtual Network. Technical report, Sun Microsystems, May 2003.

[21] B. Waxman. Routing of multipoint connections. *IEEE Journal on Selected Areas in Communications*, 9(6), 1988.