PRAMANA — journal of physics | © Indian Academy of Sciences | Vol. 71, No. 2 August 2008 pp. 263–273

Stability analysis of peer-to-peer networks against churn

BIVAS MITRA[1], SUJOY GHOSE[1], NILOY GANGULY[1,*] and FERNANDO PERUANI[2,3]
[1]Department of Computer Science and Engineering, Indian Institute of Technology, Kharagpur 721 302, India
[2]Max Plank Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
[3]ZIH, Technical University, Dresden, Zellescher Weg 12, 01069 Dresden, Germany
[*]Corresponding author
E-mail: {bivasm, sujoy, niloy}@cse.iitkgp.ernet.in; peruani@mpipks-dresden.mpg.de

Abstract. Users of the peer-to-peer system join and leave the network randomly, which makes the overlay network dynamic and unstable in nature. In this paper, we propose an analytical framework to assess the robustness of p2p networks in the face of user churn. We model the peer churn through degree-independent as well as degree-dependent node failure. Lately, superpeer networks are becoming the most widely used topology among the p2p networks. Therefore, we perform the stability analysis of superpeer networks as a case study. We validate the analytically derived results with the help of simulation.

Keywords. Superpeer networks; percolation theory; statistical mechanics.

PACS Nos 89.75.Hc; 89.20.Ff; 02.50.-r; 89.75.-k

1. Introduction

Modern large scale peer-to-peer (p2p) networks present several unique aspects that distinguish them from traditional distributed systems [1]. In client/server architecture, each computer on the network works either as a client or as a server where client requests services from the server and server provides service to the client. However, peer-to-peer architecture is a type of network in which each workstation has equivalent capabilities and responsibilities. These kinds of networks diverge responsibility between participant computers in a network rather than conventional centralized resources. Such networks are useful for many purposes like file sharing, telephony, media streaming (radio, video), discussion forums etc [2]. Peers in p2p networks are connected among themselves by some logical links forming an overlay above the physical network [3]. It has been found that these overlay networks, consisting of a large amount of peers, are analogous to complex real world networks and can be modelled using various types of random graphs [4].

In peer-to-peer networks, there is a high rate of peer churn, that is, nodes continuously leave and join the network. This dynamical behaviour of the peers frequently partitions the network into smaller fragments which results in the breakdown of communication among peers. In this paper we concentrate on understanding the stability of the p2p networks in the face of peer churn. We model peer churn by two kinds of node failures in the random graph, based on the work done in [5].

- The most common type of failure is denoted as degree-independent failure where probability of removal of a node is constant and independent of the degree of that node.
- In p2p networks, peers having higher connectivity (e.g. superpeers) are more stable in the network than the peers having lower connectivity (e.g. connected through dial-up line) because those loosely connected peers enter and leave the network quite frequently. This observation leads us to model a new kind of failure where probability of removal of a node is inversely proportional to the degree of that node. We denote this kind of failure as degree-dependent failure.

We characterize the topology of the network by a probability distribution P and dynamics of the nodes by another probability distribution Q. Using these, we develop an analytical framework to examine the stability of generalized graphs where the vertices undergo some dynamics. However, currently superpeer topologies [6,7] are emerging as the most influencing peer-to-peer networks. In this system, superpeer nodes with high bandwidth connect to each other forming the upper level in the network hierarchy. A large number of ordinary peers are connected with superpeers to get service from them. A comprehensive study of the stability of superpeer networks against peer churn is the primary focus of this paper. We also perform simulations to verify the goodness of our theoretical results.

The stability analysis of different complex networks against various disrupting events have been discussed in [8–10]. Their results have shown that scale-free networks display a high degree of tolerance against random failures but quite sensitive against intentional attack. In [11], Newman *et al* have developed the theory of random graphs with arbitrary degree distribution with the help of generating function formalism. In [12], Callaway *et al* have introduced the concept of percolation process [13] and applied it to examine the resilience of various real world networks like Internet. In [14], researchers have addressed a more realistic scenario in which a network is subjected to simultaneous targeted and random attacks.

The rest of the paper is organized as follows. In §2 we extend the analytical framework [5] to analyse the stability of peer-to-peer networks. Section 3 defines and models various environmental parameters like p2p overlay networks and peer churn. In this section, we also elaborate the simulation environment generated to mimic large superpeer networks and specify the stability metric of the network. Section 4 theoretically analyses the stability of peer-to-peer networks for degree-independent and degree-dependent failures. Section 5 customizes those models for superpeer networks and compares the theoretical results with simulation. Finally §6 concludes the paper.

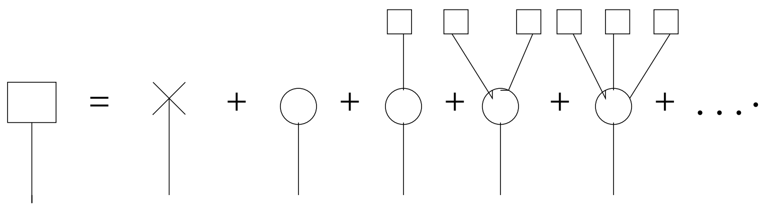

Figure 1. Schematic representation of the sum rule for the connected component of nodes reached by following a randomly chosen edge. The entire sum can be expressed in closed form as eq. (1) and similarly (2).

2. Developing analytical framework using generating function formalism

In this section, we use generating function formalism to derive the general formula for measuring the stability of overlay structures undergoing any kind of disturbances in the network. We explain the basic concept behind the development of the framework without going into mathematical details. Generating function is a formal power series whose coefficients encode information about a sequence that is indexed by the natural numbers [11]. This generating function can be used to understand different properties of graphs. For example, let the generating function $G_0(x)$ generates the probability distribution of the vertex degrees k. Therefore, $G_0(x) = \sum_{k=0}^{\infty} p_k x^k$ where p_k is the probability that a randomly chosen vertex in the graph has degree k. The importance of the generating function lies in the convenient way it can be used to understand various properties of the graph – for instance, the average degree z of a vertex in the case of $G_0(x)$ is given by $z = \langle k \rangle = \sum_k k p_k = G_0'(1)$. Higher moments can be calculated from higher derivatives also. Let q_k be the probability that a vertex of degree k be present in the network after the removal of a fraction of nodes. In our formalism f_k $(=1 - q_k)$ and p_k specify the failure model and network topology respectively whose stability is subjected to examination. The formalism helps us to locate the transition point where the giant component [16] breaks down into smaller components. $p_k \cdot q_k$ specifies the probability of a node having degree k to be present in the network after the process of removal of some portion of nodes is completed. Hence

$$F_0(x) = \sum_{k=0}^{\infty} p_k \cdot q_k x^k$$

becomes the generating function for this distribution. Distribution of the outgoing edges of the first neighbour of a randomly chosen node can be generated by

$$F_1(x) = \frac{\sum_k k p_k q_k x^{k-1}}{\sum_k k p_k} = \frac{F_0'(x)}{z},$$

where z is the average degree [12]. Let $H_1(x)$ be the generating function for the distribution of the component sizes that are reached by choosing a random edge and following it to one of its ends. Except when we are precisely at the phase transition where giant component appears, typical component size is finite. Moreover, as chance of a component containing a closed loop of edges goes down exponentially with size of the graph, it becomes negligible for large graph [11]. Therefore, the component may be conceptualized as a tree-like structure that contains zero node if the node at the other end of the randomly selected edge is removed, which happens with probability $1 - F_1(1)$, or the edge may lead to a node with k other edges leading out of it other than the edge we came in along, distributed according to $F_1(x)$ (figure 1). That means $H_1(x)$ satisfies a self-consistency condition of the form [12]

$$H_1(x) = 1 - F_1(1) + xF_1(H_1(x)). \tag{1}$$

The distribution for the component size to which a randomly selected node belongs to is similarly generated by (figure 1) $H_0(x)$ where

$$H_0(x) = 1 - F_0(1) + xF_0(H_1(x)). \tag{2}$$

Therefore the average size of the components becomes

$$H_0'(1) = \langle s \rangle = F_0(1) + \frac{F_0'(1)F_1(1)}{1 - F_1'(1)}$$

which diverges when $1 - F_1'(1) = 0$, that is, the size of the component becomes infinite. Hence

$$F_1'(1) = 1 \Rightarrow \sum_{k=0}^{\infty} kp_k(kq_k - q_k - 1) = 0. \tag{3}$$

Significance of eq. (3) lies in the fact that it states the critical condition for the stability of giant component with respect to any type of graph (characterized by p_k) undergoing any type of failure (characterized by q_k). Using this formalism, we investigate the stability of peer-to-peer networks in the face of peer churn.

3. Modelling p2p overlay networks and peer churn

In this section we formally model the p2p overlay networks and peer churn and define the stability metric which are the parameters of our analytical framework.

3.1 *Modelling peer-to-peer overlay networks*

The different types of p2p overlay networks can be modelled using the uniform framework of probability distribution p_k, where p_k is the probability that a randomly chosen node has degree k. In this paper, we consider bimodal network as a simple model of superpeer network. We believe that the bimodal network is simple enough to understand and analyse; at the same time it captures the essence of commercial superpeer networks [6,15]. In bimodal network, superpeer topology can be modelled by bimodal degree distribution where a large fraction (r) of peer nodes with small degree k_l are connected with superpeers and few superpeer nodes $(1-r)$

with high degree k_m are connected to each other. Therefore, only two separate degrees are allowed in this kind of network. Formally

$$p_k > 0 \quad \text{if } k = k_l, k_m;$$
$$p_k = 0 \quad \text{otherwise.} \tag{4}$$

k_l and k_m are degrees of peers and superpeers respectively. Therefore, $p_{k_l} = r$ and $p_{k_m} = (1 - r)$.

3.2 *Modelling peer churn*

Peer churn can be modelled by different kinds of node failures in the network. Let q_k be the probability that a vertex of degree k is present in the network after the removal of a fraction of nodes. In our framework q_k is used to specify the various failure models.

- In degree-independent failure, the probability of removal of any randomly chosen node is constant, degree-independent and equal for all other nodes in the graph. Therefore, the presence of any randomly chosen node having degree k after this kind of failure is $q_k = q$ (independent of k).
- In degree-dependent failure, probability of failure of a node (f_k) having degree k is inversely proportional to k^γ, i.e. $f_k \propto 1/k^\gamma \Rightarrow f_k = \alpha/k^\gamma$ where $0 \le \alpha \le 1$ and γ is a real number. Therefore, probability of the presence of a node having degree k after this kind of failure is $q_k = (1 - \frac{\alpha}{k^\gamma})$.

3.3 *Stability metric*

The stability of overlay networks are primarily measured in terms of certain fraction of nodes (f_c) called percolation threshold [16], removal of which disintegrates the network into smaller, disconnected components. Below that threshold, there exists a connected component which spans the entire network also termed as giant component. The value of percolation threshold f_c theoretically signifies the stability of the network, e.g. higher value indicates greater stability against failure.

3.4 *Experimental set-up*

In order to generate the overlay network, every node is assigned a degree according to the topology being simulated. In the case of bimodal network, the nodes are assigned the degrees depending on the k_l and k_m values and the fraction of these nodes in total. Thereafter the edges are generated using the 'switching method' and the 'matching method' referred to in [17]. Failure of a peer effectively means deletion of the node and its corresponding edges. In the case of degree-independent failure, nodes are randomly selected using a time-seeded pseudo-random number generator and its edges removed from the adjacency list. In degree-dependent failure, first the fraction of nodes having a certain degree that need to be removed is calculated, thereafter that many nodes are selected from the total set of all such nodes randomly and its corresponding edges are removed from the adjacency list.

4. Stability against peer churn

We model the peer churn by two kinds of failures – degree-independent and degree dependent. In the next two subsections, we deal with these two kinds of failures and investigate their effect on the stability of overlay networks.

4.1 *Degree-independent failure*

In this section, we discuss the effect of degree-independent failure in generalized random graph. If $q = q_r$ is the critical fraction of nodes whose presence in the graph is essential for the stability of the giant component after this kind of failure, then according to eq. (3)

$$\sum_{k=0}^{\infty} kp_k(kq_r - q_r - 1) = 0,$$
$$\Rightarrow q_r = \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}.$$

Now if f_r is the critical fraction of nodes whose random removal disintegrates the giant component then $f_r = 1 - q_r$. Therefore, percolation threshold

$$f_r = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle} - 1}. \tag{5}$$

This is the well-known condition [8] (derived differently) for the disappearance of the giant component due to random failure. Note that, we have reproduced it to show that it can also be derived from the proposed general formula (eq. (3)).

4.2 *Degree-dependent failure*

In p2p networks, the peers (or superpeers) having higher connectivity are much more stable and reliable than the nodes having lower connectivity. Therefore, probability of the presence of a node having degree k after this kind of failure is

$$q_k = \left(1 - \frac{\alpha}{k^\gamma}\right). \tag{6}$$

Using eqs (3) and (6), we obtain the following critical condition for the stability of giant component after degree-dependent breakdown:

$$\langle k^2\rangle - \alpha\langle k^{2-\gamma}\rangle + \alpha\langle k^{1-\gamma}\rangle - 2\langle k\rangle = 0,$$

where percolation threshold is

$$f_d = \sum_{k=0}^{\infty} \frac{\alpha}{k^\gamma} p_k.$$

Considering $\alpha = 1$, where the fraction of nodes removed due to this kind of failure becomes maximum, the condition for percolation becomes

$$\langle k^{2-\gamma}\rangle - \langle k^{1-\gamma}\rangle = \langle k^2\rangle - 2\langle k\rangle. \tag{7}$$

Thus the critical fraction of nodes removed is given by

$$f_d = \sum_{k=0}^{\infty} \frac{1}{k^\gamma} p_k, \tag{8}$$

where γ satisfies eq. (7). Thus from eqs (7) and (8), we can determine the variation of percolation threshold f_d for various networks due to degree-dependent failure. In the next section, we apply these formalism for superpeer networks and compare with simulation results.

5. Stability of superpeer networks against churn

In this section we study the stability of the superpeer networks with the help of our analytical framework. We investigate the change of percolation threshold (f_c) due to the change of fraction of peers (r) and the connectivity of the superpeers (k_m) in the networks for various types of failures. To ensure fair comparisons, we keep the average degree $\langle k\rangle$ constant for all graphs. We verify our theoretical results with the help of simulation.

5.1 *Degree-independent failure*

We model superpeer networks with the help of bimodal degree distribution. Let r be the initial fraction of peers in the superpeer networks having degree k_l and rest are superpeers having degree k_m where $k_l \ll k_m$. Therefore, in bimodal degree distribution p_k becomes nonzero only at k_l and k_m (eq. (4)). Mathematically, $k_l p_{k_l} + k_m p_{k_m} = \langle k\rangle$ and $p_{k_l} + p_{k_m} = 1$ which provides

$$p_{k_m} = \frac{\langle k\rangle - k_l}{k_m - k_l}, \quad p_{k_l} = \frac{k_m - \langle k\rangle}{k_m - k_l}$$

$\Rightarrow \langle k^2\rangle = k_m^2 p_{k_m} + k_l^2 p_{k_l} = \langle k\rangle(k_l + k_m) - k_l k_m$ and using eq. (4) we get

$$f_r = 1 - \frac{\langle k\rangle}{\langle k\rangle(k_l + k_m - 1) - k_l k_m}.$$

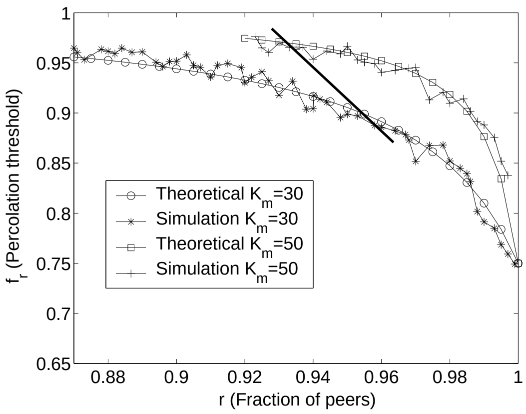

Figure 2. The above plots represent a comparative study of theoretical and simulation results of stability for two bimodal networks undergoing churn. Here x-axis represents the initial fraction of peer nodes (r) existing in the network and y-axis represents the corresponding percolation threshold (f_r). We keep the average degree $\langle k\rangle = 5$ fixed and vary the superpeer degree $k_m = 30, 50$ for two plots. The tangential line indicates the change in peer degree due to change in the peer fraction r.

The initial fraction of peers in the network having degree k_l is r, and therefore the average degree of the network $\langle k\rangle = k_l r + k_m(1-r)$ implies that $k_l = \frac{\langle k\rangle-(1-r)k_m}{r}$. Hence percolation threshold

$$f_r = 1 - \frac{\langle k\rangle r}{\langle k\rangle^2 - 2\langle k\rangle k_m + 2rk_m\langle k\rangle - r\langle k\rangle + k_m^2 - rk_m^2}. \tag{9}$$

Using eq. (9), we study the variation of percolation threshold (f_r) due to the change in the initial fraction of peers (r) in the networks with two different superpeer degrees k_m and compare the results experimentally (figure 2). It can be observed from figure 2 that simulation results match closely with theoretical predictions which shows the success of our theoretical framework.

Observations

1. It is important to observe that for the entire range of peer fractions r, the percolation threshold f_r is greater than 0.7 which implies that superpeer networks are quite robust against churn. In practice also, superpeer networks exhibit stable behaviour against churn and consequently the possible breakdown of the network is a rare event [18].

2. Lower fraction of superpeers in the network (specifically when it is below 5%) results in a sharp fall of f_r. That is, the vulnerability of the network drastically increases when the fraction of superpeers is below 5%. Higher fraction of superpeers in the network results in low peer connectivity. Therefore, most of the peers are only connected to superpeers (and not within themselves). Hence stability of the

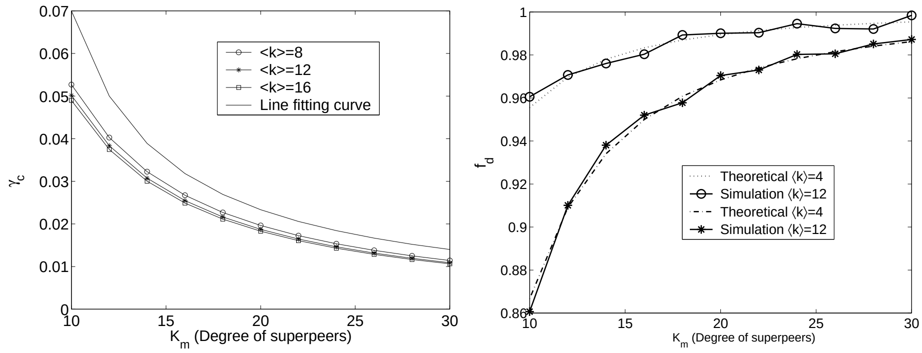

Figure 3. Change of γ_c and percolation threshold f_d with respect of superpeer degree k_m for superpeer networks undergoing degree-dependent failure. Here mean degree $\langle k\rangle$ varies from 8 to 16. x-axis represents the superpeer degree (k_m) and y-axis represents the corresponding γ_c and f_d.

network depends entirely on superpeers. As the fraction of superpeer reduces below 5%, peer degree becomes quite high (4 to 5). This gives rise to situations where some peers are not connected to the superpeers at all, but only connected to fellow peers. Hence removal of individual peers also results in the removal of fellow peers. This produces an avalanche effect which results in a drastic reduction of stability of the network in this region.

5.2 *Degree-dependent failure*

In degree-dependent failure, the bimodal network percolates if

$$\langle k^{2-\gamma}\rangle - \langle k^{1-\gamma}\rangle = \langle k^2\rangle - 2\langle k\rangle. \tag{10}$$

If the value of $\gamma = \gamma_c$ satisfies this equation then removal of $f_d = \sum_{k=0}^{\infty} \frac{1}{k^{\gamma_c}} p_k$ fraction of nodes destroys the giant component. However, for $\gamma > \gamma_c$, the network survives after node removal. In most of the commercial superpeer networks like KaZaA [19], peers are only directly connected to the local superpeer making their degree $k_l = 1$. In that case, the value of γ_c which percolates the bimodal network can be derived from eq. (10) as

$$\gamma_c = 1 - \frac{\ln \frac{\langle k\rangle(k_m+1) - k_m - 2\langle k\rangle}{\langle k\rangle - 1}}{\ln k_m}. \tag{11}$$

We plot the variation of γ_c and percolation threshold f_d with respect to the superpeer degree k_m for various average degrees $\langle k\rangle$ (figure 3). It is important to notice that the increase in the superpeer degree k_m increases peer fraction r to keep the average degree $\langle k\rangle$ fixed. However, here we are interested in understanding the impact of superpeer degree upon stability of the networks. It can be observed from figure 3 that simulation result matches closely with theoretical prediction which shows the success of our theoretical framework.

Observations

1. It can be easily identified from figure 3, that with the increase of superpeer degree k_m, the value of γ_c that percolates the network decreases. This increases the necessary fraction of superpeers required to be removed to breakdown the network. The nature of γ_c can be approximated by the polynomial $a/(x-b)$ ($0 < a < 1$ and b is some positive integer). Thus the decrease of γ_c follows hyperbolic curve. Since the increase of k_m increases the fraction of peers r, the removal of most of the low degree peers along with a fraction of superpeers increases the percolation threshold f_d.

2. It is interesting to observe that the percolating γ_c remains quite low and less than 0.1 for the entire range of k_m because small values of γ_c result in the removal of higher fraction of superpeer nodes from the network. Since the degree-dependent failure mainly removes the lower degree nodes, which are not so useful to break the network down, removal of a significant amount of superpeers becomes necessary. On the other hand, increase in the value of $\gamma_c > 0.1$ removes relatively small amount of superpeers that is insufficient to destroy the network.

3. Another interesting observation is that after a certain threshold k_m, the curves become parallel to the x-axis and never cut it. Thus, the value of γ_c is small but never becomes 0 (in that case $f_d = \sum_{k=0}^{\infty} \frac{1}{k^0} p_k = 1$). This implies that for any large value of k_m, although f_d becomes significantly large it is required to remove only a part of nodes (and not all the nodes) from the network to dissolve the giant component.

6. Conclusion

The basic contribution of this paper is the development of a general framework to analyse the stability of various p2p networks against peer churn. We have modelled the peer churn using degree-independent and degree-dependent node failure. As superpeer networks are currently the most promising and widely used overlay structure, we perform stability analysis of these networks as a case study of our analytical model. It has been observed that when the fraction of superpeers in the network is less than 5%, the stability of the network sharply decreases for degree-independent failure. This result points to a zone where superpeer networks are most vulnerable. Similarly for degree-dependent failure, our analysis shows that increase of superpeer degree improves the stability of the network and the improvement follows a hyperbolic curve.

References

[1] D Clark, *IEEE Computer* **34(1)**, 18 (2001)
[2] K Singh and H Schulzrinne, *Peer-to-peer internet telephony using SIP*, Columbia University Technical Report CUCS-044-04, New York, NY October (2004)

[3] W Si and M Li, On the connectedness of peer-to-peer overlay networks, *Proceedings of the 11th International Conference on Parallel and Distributed Systems* (ICPADS'05), Volume 01, July (2005)
[4] Q Lv, P Cao, E Cohen, K Li and S Shenker, Search and replication in unstructured peer-to-peer networks, *ACM International Conference on Supercomputing* (New York, USA, 2002)
[5] B Mitra, F Peruani, S Ghose and N Ganguly, Analyzing the vulnerability of superpeer networks against attack, *14th ACM Conference on Computer and Communications Security* (Alexandria, USA, 29th Oct–2nd Nov. 2007)
[6] B Yang and H Garcia-Molina, Designing super-peer networks, *Proceedings of the International Conference on Data Engineering* (*ICDE*), (Los Alamitos, CA, March 2003)
[7] Y J Pyun and D S Reeves, Constructing a balanced, log(N)-diameter super-peer topology, *Proceedings of the 4th International Conference on Peer-to-Peer Computing* (Zurich, Switzerland, August 2004)
[8] R Cohen, K Erez, D Avraham and S Havlin, *Phys. Rev. Lett.* **85(21)**, 4626 (2000)
[9] R Cohen, K Erez, D Avraham and S Havlin, *Phys. Rev. Lett.* **86(16)**, 3682 (2001)
[10] R Albert, H Jhong and A L Barabsi, *Nature (London)* **406**, 378 (2000)
[11] M E J Newman, S H Strogatz and D J Watts, *Phys. Rev.* **E64**, 026118 (2001)
[12] D S Callaway, M E J Newman, S H Strogatz and D J Watts, *Phys. Rev. Lett.* **85(21)**, 5468 (2000)
[13] M Molloy and B Reed, *Random Structures and Algorithms* **6**, 161 (1995)
[14] T Tanizawa, G Paul, R Cohen, S Havlin and H E Stanley, *Phys. Rev.* **E71**, 047101 (2005)
[15] G Paul, S Sreenivasan, S Havlin and H Stanley, *Phys. Rev.* **E72**, 056130 (2005)
[16] M Molloy and B Reed, *Combinatorics, Probability and Computing* **7**, 295 (1998)
[17] R Milo, N Kashtan, S Itzkovitz, M E J Newman and U Alon, *On the uniform generation of random graphs with prescribed degree sequences*, eprint arXiv:cond-mat/0312028 (2003)
[18] D Stutzbach, R Rejaie and S Sen, Characterizing unstructured overlay topologies in modern p2p file-sharing systems, *Proceedings of ACM SIGCOMM/USENIX Internet Measurement Conference* (Berkeley, CA, October 2005)
[19] KaZaA website. http://www.kazaa.com