How do Superpeer Networks Emerge?

Bivas Mitra, Abhishek Kumar Dubey, Sujoy Ghose, Niloy Ganguly
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur, India
Email: {bivasm,abhish,sujoy,niloy}@cse.iitkgp.ernet.in

***Abstract*—In this paper, we develop an analytical framework which explains the emergence of superpeer networks on execution of the commercial peer-to-peer bootstrapping protocols by incoming nodes. Bootstrapping protocols exploit physical properties of the online peers like resource content, processing power, storage space, connectivity etc as well as take the finiteness of bandwidth of each online peer into consideration. With the help of rate equations, we show that execution of these protocols results in the emergence of superpeer nodes in the network - the exact degree distribution is evaluated. We validate the framework developed in this paper through extensive simulation. The analysis of the results shows that the amount of superpeers produced in the network depends on the protocol as well as the properties of the joining nodes. Interestingly, our analysis reveals that increase in the amount of resource and the number of resourceful nodes do not always help to increase the fraction of superpeer nodes in the network. The impact of the frequent leaving of the peers on the topology of the emerging network is also evaluated. As an application study, we show that our framework can explain the topological configuration of commercial Gnutella networks. The developed model can almost perfectly match the degree distribution of Gnutella network.**

***Index Terms*—Superpeer networks, bootstrapping protocols, bimodal network, GWebCache, preferential attachment.**

I. Introduction

Currently superpeer networks have proved to be the most influencing peer-to-peer topologies [1], [2] and form the underlying architecture of various commercial peer-to-peer (p2p) based systems like KaZaA, Gnutella, Skype etc [3]. The constituent nodes in these networks maintain a two layer overlay topology where the top layer consists of the high speed superpeers and the bottom layer consists of the ordinary peers [4]. Superpeer network is formed mainly as a result of the bootstrapping or joining protocol followed by incoming peers[1]. The bootstrapping protocol selects some online peers (nodes) that are already part of the network and sends connection requests to them [5]. The selection of online peers is guided by two specific criteria. First of all, a peer connects to some selected online peers with the purpose of maximizing the quality of services like minimizing the search time, fast downloading of files etc [6]. Secondly, the time spent by the peer to perform bootstrapping needs to be minimized because until the peer gets connected to the network, it cannot initiate normal p2p activities. In order to attain the above two objectives, peers try (prefer) to join to 'good' (resourceful) nodes; all existing bootstrapping protocols are essentially directed towards fulfilling these basic objectives [7], [8]. An incoming peer joining the p2p networks like Gnutella collects the list of 'good' online peers from hardcoded address and/or from GWebCache which is a distributed repository for maintaining the information of 'good' online peers in the network [5]. The superpeer networks emerged following these bootstrapping protocols exhibit two regimes or 'bimodality' in their degree distribution; one regime consists of the large number of low degree peer nodes and another one consists of the small number of high degree superpeers [9]. The emergence of bimodal network due to the bootstrapping of nodes is an *interesting* observation. This is not obvious why bootstrapping of nodes leads to the emergence of bimodal network, hardly there is any explanation found in the literature. [10], [11], [12], [13] have shown the emergence of scale free networks as a result of the preferential attachment of incoming nodes with the 'good' existing nodes[2]. In line with that, one can reasonably expect that the additional constraint of finiteness of bandwidth present here may lead to power law networks with an exponential cutoff degree. However, we observe that the superpeer networks follow bimodal degree distribution that sharply deviates from the power law behavior of scale free networks [4], [9].

The performance of the superpeer networks mainly depends upon the topological properties of the emerging networks [2], [4], [14], [15]. This includes the network diameter, amount of superpeers in the network, peer-superpeer ratio etc. Hence, regulating these topological properties and subsequently improving the performance of various p2p services will prove to be an *useful* step for p2p research community. Due to its decentralized nature of formation, controlling the topological structure of the superpeer network is not a trivial task. However, to the best of our knowledge, little work has been done to calculate these network parameters that emerges as a result of the bootstrapping process.

This *useful* step can be initiated if we carefully interpret the reasons behind the *interesting* observation. In that line, we develop a theoretical framework to explain the appearance of superpeer networks due to the execution of servents like limewire, mutella etc [5], [16]. In order to develop the framework, we model the bootstrapping protocols by the node attachment rule where the probability of joining of an incoming

[1]Some other minor factors are also involved in the formation of the superpeer network like peer churn, rewiring of links, upgradation of the peers to the superpeers etc. However bootstrapping takes the major role in the formation of the network topology [8].

[2]The 'goodness' of a node has focused both towards link property (such as degree centrality) and node property (such as 'fitness') [13].

peer to an online node is proportional to the 'goodness' of the online node. 'Goodness' of a peer can be characterized by the *node property* (later quantified as *node weight*) like amount of resource, processing power, storage space etc that a particular peer possesses as well as its *current degree*. Beyond this, we identify that in p2p networks, bandwidth of a node is finite and restricts its maximum degree (*cutoff degree*). A node, after reaching its maximum degree, rejects any further connection requests from incoming peers. Our framework shows that the interplay of finite bandwidth with node property plays key role in the emergence of bimodal network. Through suitable mathematical treatment on the framework, we calculate the amount of superpeers in the network, the impact of different parameters like resource, processing power etc on the superpeer-peer ratio etc. This understanding may further help network engineers to appropriately tune the servent programs for improving the p2p services like minimizing search time, fast downloading of files etc. As a practical application, we show that our formalism (with a small modification) almost accurately explains the topological structure of the Gnutella network, obtained from the real data taken in September 2004. The outline of the paper is as follows. In section II, we state and model the bootstrapping protocol followed by peer servents. Section III proposes a formal framework considering that all the peers join with fixed cutoff degree. In section IV, we generalize the theory for the case where different peers join the network with individual (variable) cutoff degrees. In section V, we report the change in the superpeer topology due to the frequent departure of the online nodes, termed as peer churn [17]. In light of the framework developed, an empirical analysis of the global nature of the Gnutella 0.6 network is provided in section VI. Some suggestions to the network engineers in order to improve the p2p services is provided in section VII after which we conclude our paper.

II. Bootstrapping protocol

Input: Nodes, where each node i comes with individual node weight w_i and cutoff degree $k_c(i)$
Output: Emergence of the network due to joining of nodes
foreach *Incoming node* i **do**
  Node i preferentially chooses m' ($m' > m$) online nodes based on their weights and degrees
  while *m online nodes are not connected with i* **do**
    j = select an online node among the chosen m' nodes
    Node i sends the connection request to j
    if *degree(j)*$< k_c(j)$ **then**
      Node i connects with node j
    end
    else
      Node j rejects the connection request
    end
  end
end

In this section, we illustrate and model the bootstrapping protocols that is executed by different servent programs [18]. Servents like limewire and gnucleus maintain a list of 'good' hosts in the GWebCache and give priority to them during connection initiation [5]. We model bootstrapping protocols through node attachment rules where probability of attachment of the incoming peer to an online node is proportional to the node property (weight) and degree of the online node. The generalized bootstrapping protocol is mentioned above. The cutoff degree $k_c(i)$ is same for all peers i in the analysis of section III while it is varied in section IV.

III. Development of analytical framework: peers joining with fixed cutoff degree

In this analysis, we assume that each incoming peer joins the network at a timestep n with some node weight and connects to m online nodes in the network following the bootstrapping protocol. The minimum and maximum weight of a node in the network can be w_{min} and w_{max} respectively. The probability of attachment of the incoming peer to an online node is proportional to the weight and current degree of the online node. The probability that an incoming peer has weight w_i is f_{w_i} and all the nodes have some fixed cutoff degree k_c. Any node upon reaching the degree k_c rejects any further connection request from the incoming peer.

We introduce the term set_{w_i} to denote the set of nodes in the network with weight w_i. Initially we intend to compute p_{k,w_i}, the fraction of k degree nodes in set_{w_i} and then sum it over all sets (weights) to find degree distribution p_k. These values of p_{k,w_i} can be computed by observing the shift in the number of k degree nodes to $k+1$ degree nodes as well as $k-1$ degree nodes to k degree nodes due to the attachment of a new node at timestep n. Let the fraction of nodes in set_{w_i} having degree k at some timestep n be p_{k,n,w_i}, then the total number of k degree nodes in set_{w_i} before addition of a new node is $nf_{w_i}p_{k,n,w_i}$ and after addition of the node becomes $(n+1)f_{w_i}p_{k,n+1,w_i}$. Hence, the change in the number of k degree nodes in set_{w_i} becomes

$$\Delta n_{k,w_i} = (n+1)f_{w_i}p_{k,n+1,w_i} - nf_{w_i}p_{k,n,w_i} \quad (1)$$

We formulate rate equations depicting these changes for some arbitrary set_{w_i}. By solving those rate equations, we calculate p_{k,w_i} and subsequently the degree distribution p_k (fraction of nodes having degree k) of the entire network.

Methodology

In order to write the rate equations [11], we need to know the probability A_{w_i} that an online node x with weight w_i (i.e. in set_{w_i}) will receive a new link from the incoming peer. The probability that an online node will receive an incoming link is proportional to the node weight w_i and current degree k and can be depicted as

$$\begin{aligned} A_{w_i} &= \frac{w_i f_{w_i} \sum_{k=m}^{k_c-1} k p_{k,w_i}}{\sum_{i'=min}^{max} w_{i'} f_{w_{i'}} \sum_{k_1=m}^{k_c-1} k_1 p_{k_1,w_{i'}}} \\ &= \frac{w_i f_{w_i} m_{w_i} \beta_i}{\sum_{i'=min}^{max} w_{i'} f_{w_{i'}} m_{w_{i'}} \beta_{i'}} \quad degree(x) < k_c \quad (2) \\ &= 0 \qquad degree(x) \geq k_c \end{aligned}$$

where $\beta_i = 1 - \frac{k_c p_{k_c,w_i}}{2m_{w_i}}$ (p_{k_c,w_i} is the fraction of nodes in set_{w_i} that have reached their cutoff degree k_c hence stopped accepting new links) implies the fraction of nodes in set_{w_i} capable of accepting new links from the incoming peer and normalizing constant $2m_{w_i} = \sum_{k=m}^{k_c} kp_{k,w_i}$ denotes the average degree of the nodes in set set_{w_i}. The numerator of Eq. (2) represents the total amount of weight of nodes in set_{w_i} that are allowed to take incoming links. The denominator normalizes the fraction by the total amount of weight of all the nodes in the network that are allowed to take incoming links.

The joining of a new node of degree m at timestep $n+1$ changes the total number of k degree nodes in set_{w_i}. Since all the nodes in the set_{w_i} contain equal weight w_i, the chance of getting a new link for the online nodes depends upon their current degree k and fraction present in the set at that timestep, hence can be expressed as $\frac{kp_{k,n,w_i}}{2m_{w_i}\beta_i}$. The β_i in denominator takes care of the fact that the nodes, that have reached the cutoff degree k_c do not participate in the formation of new link. Due to the joining of a new node of degree m in the network, some k degree nodes in set_{w_i} acquire a new link and become nodes of degree $k+1$. So the amount of decrease in the number of nodes of degree k, ($m \leq k < k_c$) in set_{w_i} due to this outflux is

$$\delta_{k\to(k+1)} = \frac{kp_{k,n,w_i}}{2m_{w_i}\beta_i} \times A_{w_i} m \quad (3)$$

Similarly a fraction of nodes having degree $k-1$ get a new link and move to the degree k. We now write the rate equations in order to formulate the change in the number of k degree nodes in an individual set_{w_i} due to the attachment of a new node of degree m. Three pertinent degree ranges $k = m$, $m < k < k_c$ and $k = k_c$ are taken into consideration.

Rate equation for $k = m$

Since the probability of joining of a node having weight w_i in the network is f_{w_i}, the joining of one new node of degree m on average increases f_{w_i} fraction of nodes of degree m in the set_{w_i}. Hence net change in the number of nodes having degree $k = m$ can be expressed as

$$\begin{aligned}\Delta n_{m,w_i} &= (n+1) f_{w_i} p_{m,n+1,w_i} - n f_{w_i} p_{m,n,w_i} \\ &= f_{w_i} - \frac{mp_{m,n,w_i}}{2m_{w_i}\beta_i} \times A_{w_i} m \quad (4)\end{aligned}$$

Therefore assuming the stationary condition, $p_{k,n+1,w_i} = p_{k,n,w_i} = p_{k,w_i}$ [11] we find

$$p_{m,w_i} = \frac{1}{(1+\frac{m}{\alpha_i})} \quad (5)$$

where $\alpha_i = \frac{2\sum_i w_i f_{w_i} m_{w_i} \beta_i}{w_i m}$

Similarly from Eq. (3), **rate equation for** $m < k < k_c$

$$\begin{aligned}\Delta n_{k,w_i} &= (n+1) f_{w_i} p_{k,n+1,w_i} - n f_{w_i} p_{k,n,w_i} \\ &= \left(\frac{(k-1)p_{k-1,n,w_i} - kp_{k,n,w_i}}{2m_{w_i}\beta_i}\right) \times A_{w_i} m (6)\end{aligned}$$

Subsequently the recurrence relation becomes

$$p_{k,w_i} = \frac{(k-1)}{(k+\alpha_i)} p_{k-1,w_i} \quad (7)$$

Rate equation for $k = k_c$

Since the nodes having degree k_c are not allowed to take any incoming links, nodes are only accumulated at degree $k = k_c$. Hence

$$\Delta n_{k_c,w_i} = \frac{(k_c-1)p_{k_c-1,n,w_i}}{2m_{w_i}\beta_i} \times A_{w_i} m \quad (8)$$

Hence the corresponding recurrence equation becomes

$$p_{k_c,w_i} = \frac{(k_c-1)}{\alpha_i} p_{k_c-1,w_i} \quad (9)$$

Computing the degree distribution

Solving the above stated rate equations, we obtain the degree distribution of the entire network.

$$\begin{aligned} p_k &= \sum_{i=min}^{max} p_{k,w_i} f_{w_i} \quad (10) \\ &= \begin{cases} \sum_{i=min}^{max} \frac{1}{(1+\frac{k}{\alpha_i})} f_{w_i} & k = m \\ \sum_{i=min}^{max} \frac{f_{w_i}}{(1+\frac{m}{\alpha_i})} \times \prod_{j=1}^{k-m} \left(\frac{k-j}{k-j+1+\alpha_i}\right) & m < k < k_c \\ \sum_{i=min}^{max} f_{w_i} \prod_{j=m}^{k-1} \frac{j}{(j+\alpha_i)} & k = k_c \end{cases}\end{aligned}$$

A. Emergence of superpeer nodes

We are now in the position to theoretically understand the emergence of bimodal distribution as well as the accumulation of superpeer nodes. A closer look at the equations reveals that two modes appear in the degree distribution, one at $k = m$ around which the degree of most of the nodes are concentrated and another at $k = k_c$. **Conditions: a.** In order to show the appearance of mode or spike at $k = k_c$, we have to satisfy the condition $p_{k_c} > p_{k_c-1}$ and $p_{k_c} > p_{k_c+1}$. **b.** In order to show the modal behavior at $k = m$, we have to satisfy the condition $p_k < p_{k-1}$ for $m \leq k < k_c$. This also confirms that no other modes have emerged in the network.

Fulfilling condition a: First of all, we show that the fraction of nodes having degree k_c, p_{k_c} is greater than p_{k_c-1}. From Eq. (9), we find that for the set_{w_i}

$$\frac{p_{k_c,w_i}}{p_{k_c-1,w_i}} = \frac{(k_c-1)}{\alpha_i} = (k_c-1)\frac{w_i m}{\sum_i w_i m_{w_i} f_{w_i} \beta_i} \quad (11)$$

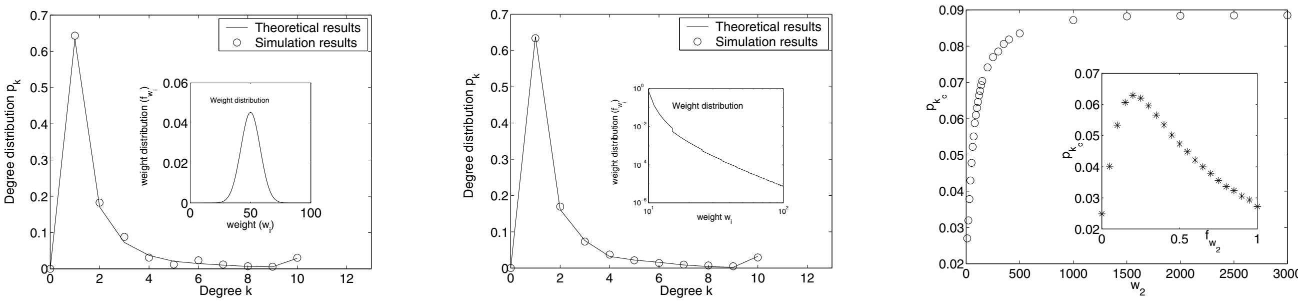
Then, for the entire network,

$$\frac{p_{k_c}}{p_{k_c-1}} \approx \sum_{i=min}^{max} \frac{p_{k_c,w_i}}{p_{k_c-1,w_i}} = (k_c-1)\frac{\sum_i w_i m}{\sum_i w_i m_{w_i} f_{w_i} \beta_i} > 1 \quad (12)$$

since $\sum_i m_{w_i} f_{w_i} = m$, $\beta_i < 1$ therefore $m_{w_i} f_{w_i} \beta_i < m$ and $k_c >> 1$. Secondly, the bootstrapping protocol gives $p_k = 0$ for $k > k_c$. Hence, we conclude the presence of a spike at degree k_c.

Fulfilling condition b: We find for $m \leq k < k_c$, the probability p_k continuously decreases. This can be understood from Eq. (7) of the set_{w_i}

$$\frac{p_{k,w_i}}{p_{k-1,w_i}} = \frac{(k-1)}{(k+\alpha_i)} < 1 \quad (13)$$



(a) Degree distribution of the network. Weight distribution is taken from normal distribution (inset).

(b) Degree distribution of the network. Weight distribution is taken from power law distribution (inset in log-log scale).

(c) The plot illustrates the change in p_{k_c} at $k_c = 10$ with the change of node weight w_2. The inset illustrates the change in p_{k_c} with the change of f_{w_2}.

Fig. 1. The plot represents the degree distribution of the network emerged following bootstrapping protocol with fixed cutoff degree $k_c = 10$ and $m = 1$. The nodes can join the network with weights taken from normal distribution (mean=50 and standard deviation 8, Fig 1(a)) and power law distribution (exponent=2.5, Fig 1(b)). Fig 1(c) shows the change in p_{k_c} due to change in w_2 and f_{w_2} for the bimodal weight distribution (simulation results). Inset of Fig 1(c) indicates the presence of optimum f_{w_2} (i.e. $f^*_{w_2}$) at which p_{k_c} becomes maximum.

i.e. $p_{k,w_i} < p_{k-1,w_i}$. Hence for the entire network, $p_k < p_{k-1}$. These two observations confirm the presence of two distinct modes in the degree distribution and lead to the emergence of high degree superpeer nodes at degree k_c (Fig 1(a), 1(b)). Note that, this feature is independent of the weight distribution f_w.

B. Simulation results and inference derivation

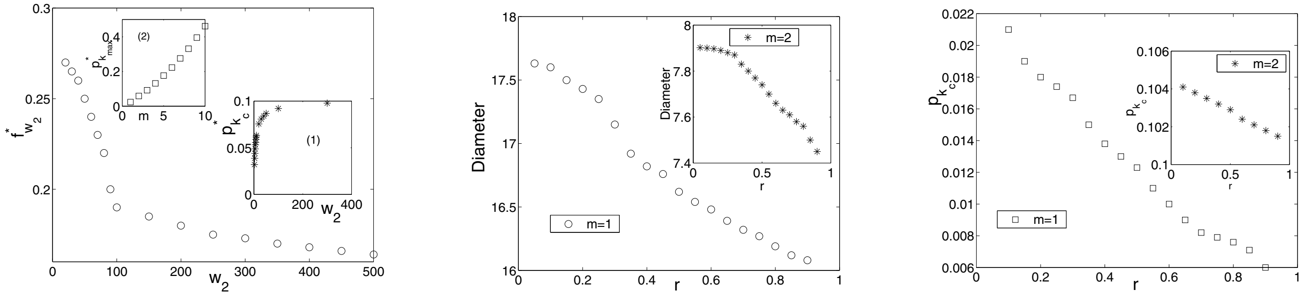
We validate the theoretically obtained degree distribution (Eq. (10)) by simulating the emergence of the network (Fig. 1). In these simulations, we follow the exactly same procedure and assumptions that we have taken for theoretical modeling. The stochastic simulation set up is as follows. During bootstrapping, each node joins the network with some weight ($10 \leq w \leq 100$) taken from a weight distribution f_w. A 'fitness' value is assigned to each online node based upon its weight and current degree. The incoming new node gets connected with an online node depending upon the 'fitness' of that online node. In our simulation, we consider two weight distributions, namely normal distribution and power law distribution [15], [19]. The total number of nodes in the system is considered to be 5000 and we perform 500 individual realizations and plot the average of that. Fig. 1 shows that the agreement between the theoretical and simulation results is exact which validates the correctness of the theoretical model. Figs 1(a), 1(b) produce the evidence of the emergence of two distinct regions in the degree distribution - the peer and superpeer regions; the accumulation of the superpeer nodes occurs at degree $k_c = 10$. Fig 1 confirms that the weight distribution hardly changes the nature (i.e. bimodalilty) of the degree distribution. In the following, we investigate the influence of different parameters on the amount of superpeers in the network. In order to gain more insights, we consider a simple bimodal weight distribution where nodes join with two weights w_1 (low) and w_2 (high) with individual fractions f_{w_1} (high) and f_{w_2} (low) respectively.

1) Impact of node weight w_2 on p_{k_c} : In order to examine the impact of node weight, we perform the simulation with $w_1 = 10$ and $f_{w_1} = 0.8$. The node weight w_2 is varied from 10 to 3000 and we observe how it affects p_{k_c} (k_c=10). It can be observed from Fig. 1(c) that, initial increase in w_2 increases the fraction of superpeer nodes (p_{k_c}) in the network rapidly. However, after a certain threshold, the p_{k_c} stabilizes and further increase in weight does not increase p_{k_c}. Mathematically from Eq. (10), as $w_2 \to \infty$, p_{k_c} becomes

$$\lim_{w_2 \to \infty} p_{k_c} = f_{w_2} \prod_{j=m}^{k_c-1} \frac{j}{(j + \frac{2}{m} f_{w_2} m_2 \beta_2)} \tag{14}$$

and converges to some finite value. Hence, we conclude that after some threshold limit, increase in the node weight does not increase the amount of superpeers in the network.

2) Impact of fraction of high weighted nodes (f_{w_2}) on p_{k_c}: In order to observe the impact of f_{w_2} on p_{k_c}, we simulate the bootstrapping protocol for two weights $w_1 = 10$ and $w_2 = 100$ and gradually increase the f_{w_2} (i.e. decrease f_{w_1}). Common intuition is that increase in f_{w_2} in the network should increase p_{k_c} (number of superpeers) as well. However inset of Fig. 1(c) shows that the initial increase in f_{w_2} increases p_{k_c}. But after reaching some maximum value ($p^*_{k_c}$), p_{k_c} decreases. We are interested in understanding the reason behind the presence of an optimum f_{w_2} ($f^*_{w_2}$, at which p_{k_c} becomes maximum). This can be understood by looking into the opposite forces performing at two ends (high and low) of $f^*_{w_2}$. **In low** f_{w_2} : During the joining of a new node of degree m, the existing nodes in the network acquire the links from the new node and scale their own degrees. Low f_{w_2} (i.e. high f_{w_1}) makes the $w_1 f_{w_1}$ quite significant and subsequently increases A_{w_1} in Eq. (2). In effect, out of m links of the incoming node, some of them get connected to w_1 weight nodes. However, since w_1 is small, any individual w_1 weighted node rarely becomes capable to reach k_c for contributing to p_{k_c}. But, collectively



(a) Change in the $f^*_{w_2}$ due to the increase in w_2. Inset(1) - the corresponding $p^*_{k_c}$ calculated at $f^*_{w_2}$. Inset(2) - $p^*_{k_c max}$ (using $f^*_{w_2}$ and $w_2 \rightarrow \infty$) with m.

(b) The plot illustrates the change in the diameter of the network with the change in bootstrapping protocol (r).

(c) The plot illustrates the change in the amount of superpeers (p_{k_c}) of the network with the change in bootstrapping protocol (r).

Fig. 2. In Fig 2(a), $f^*_{w_2}$ is f_{w_2} at which p_{k_c} becomes maximum ($p^*_{k_c}$). The figure shows the change in $f^*_{w_2}$ and $p^*_{k_c}$ due to w_2. In Fig 2(b) and 2(c), r is the fraction of incoming nodes which have joined the network purely based on the degree sequence of the online nodes. The results are obtained through stochastic simulation.

they restrict the w_2 weighted nodes from taking new links, hence reduce the rate of degree scaling of those nodes. This results in low value of p_{k_c}. **In high** f_{w_2} : However in high f_{w_2}, all the nodes of weight w_2 compete with each other to get the new links. This results in slowdown in the rate of increase of the degrees of the individual w_2 weighted nodes and gradually reduces p_{k_c}. The interaction of these two opposite effects result in an optimal $f^*_{w_2}$.

3) Impact of w_2 on $f^*_{w_2}$: Fig. 2(a) shows that the increase in w_2 sharply decreases the $f^*_{w_2}$. Increase in w_2 increases A_{w_2}, hence most of the links of the incoming node get attached with the nodes with high weight w_2 even if f_{w_2} is small. At the same time, low f_{w_2} restricts competition for the incoming links among the w_2 nodes and helps the small fraction of high degree nodes to quickly scale towards the cutoff degree k_c. Inset(1) of Fig. 2(a) shows that the interplay of these two factors increases $p^*_{k_c}$ (i.e. p_{k_c} at $f^*_{w_2}$). However, after reaching the saturated w_2, all the incoming links are joined to the w_2 nodes hence further increase in w_2 does not reduce the $f^*_{w_2}$ (or increase $p^*_{k_c}$) much.

Increase in m increases the amount of superpeers:
Eq. (14) calculates the maximum amount of superpeers in the network as $w_2 \rightarrow \infty$ for different f_{w_2}. The optimum fraction $f^*_{w_2}$ can be calculated from Eq. (14) by taking $\frac{dp_{k_c}}{df_{w_2}} = 0$. Substituting that $f^*_{w_2}$ in Eq. (14) gives the maximum possible amount of superpeers $p^*_{k_c max}$ for a particular m. Inset(2) of Fig 2(a) shows that with the increase in m, the $p^*_{k_c max}$ increases almost linearly.

4) Impact of the bootstrapping protocol on p2p services:
In this subsection, we investigate the implications of some modifications in the bootstrapping protocols on the various network properties like diameter, amount of superpeers etc. Let us assume that the bootstrapping protocol of the incoming peer can be controlled such that probability of connecting with only high degree online nodes is r and probability of connecting with an online node based upon both its weight and degree is $1 - r$. In simulation, we assume that the weight distribution of the incoming nodes follow power law distribution [15]. Fig. 2(b) shows that increasing r slowly decreases the diameter of the network. Reducing the diameter of the network improves the search efficiency of the network [20]. On the other hand, increasing r reduces the amount of superpeers in the network p_{k_c} (Fig 2(c)). As the file download latency is primarily dependent on the nature of the neighboring peers, the increase in the amount of superpeers results in fast downloading of files. Hence we conclude that carefully modifying the bootstrapping protocol to sieve appropriate nodes from the GWebCache may improve the p2p services by reducing search latency and fast file downloading etc.

IV. Development of analytical framework: peers joining with individual/variable cutoff degrees

In reality, nodes join the network with various bandwidth connections like ISDN, ADSL, leased line etc. Subsequently, the cutoff degree of individual nodes becomes different from one another. For simplicity, we can assume that there is a fixed number of discrete cutoff degrees each representing a type of connection. Addressing this phenomena in this section, we generalize the bootstrapping in the following way. We assume that the probabilities that a node j joins the network with cutoff degree $k_c(j)$ and weight w_j are $q_{k_c(j)}$ and f_{w_j} respectively ($q_{k_c(j)}$ and f_{w_j} are independent). Let every node necessarily have cutoff degree between a specified minimum and maximum, $k_c(min)$ and $k_c(max)$ respectively. Similar to the previous section, the probability that an online node of weight w_i (i.e. in set_{w_i}) receives a new link from the incoming peer

$$\widehat{A}_{w_i} = \frac{w_i f_{w_i} (\sum_{k=m}^{k_{min}-1} k p_{k,w_i} + \sum_{k=k_{min}}^{k_{max}} k p_{k,w_i} S_{k,w_i})}{\sum_i w_i f_{w_i} (\sum_{k=m}^{k_{min}-1} k p_{k,w_i} + \sum_{k=k_{min}}^{k_{max}} k p_{k,w_i} S_{k,w_i})} = \frac{w_i f_{w_i} \widehat{\beta}_i}{\sum_i w_i f_{w_i} \widehat{\beta}_i} \quad (15)$$

where

$$\widehat{\beta}_i = 1 - \frac{\sum_{k=k_c(min)}^{k_c(max)} (1 - S_{k,w_i}) k p_{k,w_i}}{2m_{w_i}} \tag{16}$$

implies the fraction of nodes in set_{w_i} capable of accepting new links from the incoming peer. Here S_{k,w_i} is the fraction of k degree nodes in set_{w_i} whose cutoff degree is greater than k and hence are still capable of taking incoming connections. We calculate the exact expression for S_{k,w_i} later in this section.
Similar to the section III, we formulate the rate equations to characterize joining of an incoming node of degree m. Based on the behavior of S_{k,w_i}, the formulation of rate equations and subsequently the computation of degree distribution need to be done in two parts; nodes with degree $m \le k < k_c(min)$ in part A and nodes with degree $k_c(min) \le k \le k_c(max)$ in part B.

Part A : Dynamics analysis for $m \le k < k_c(min)$
In this case, none of the nodes has reached its cutoff degree. Hence S_{k,w_i} trivially becomes 1 and the rate equations for $m \le k < k_c(min)$ are similar to the Eqs. (4) and (6).

Part B : Dynamics analysis for $k_c(min) \le k \le k_c(max)$
An important difference between part B and part A is that, at each k ($k_c(min) \le k \le k_c(max)$), a fraction of nodes reach to their cutoff degree and stop taking further links from the incoming nodes. So the calculation of S_{k,w_i} becomes nontrivial and their values play a major role in formulating the rate equations. We start our analysis with the nodes having smallest cutoff degree $k = k_c(min)$.

Calculation for $k = k_c(min)$
We defined earlier that S_{k,w_i} is the fraction of nodes having degree $k = k_c(min)$ in the set_{w_i} that have not reached to their cutoff degree and still capable of taking incoming links. Hence similar to Eq. (3), $\frac{kp_{k,w_i}}{2m_{w_i}\widehat{\beta}_i}\widehat{A}_{w_i} m S_{k,w_i}$ number of nodes can move from degree $k_c(min)$ to $k_c(min)+1$ and leave the $k_c(min)$ set. On the other hand, similar to Eq. (3), the mean number of nodes with degree $k-1$ that accepts new link and moves to degree k becomes $\frac{(k-1)p_{k-1,w_i}}{2m_{w_i}\widehat{\beta}_i}\widehat{A}_{w_i} m$. The net change in the number of nodes having degree k (for $k = k_c(min)$) due to the attachment of a new node

$$\Delta n_{k,w_i} = \frac{((k-1)p_{k-1,w_i} - kp_{k,w_i} S_{k,w_i})}{2m_{w_i}\widehat{\beta}_i} \times \widehat{A}_{w_i} m \tag{17}$$

Calculation of S_{k,w_i} **for** $k = k_c(min)$
The mean number of nodes of degree $(k-1)$ that acquires the new links from the incoming node and moves from degree $k-1$ to degree k is $\widehat{\delta}_{(k-1)\to k} = \frac{(k-1)p_{k-1,w_i}}{2m_{w_i}\widehat{\beta}_i}\widehat{A}_{w_i} m$. As q_k is the probability that a node joins the network with cutoff degree $k = k_c(min)$, hence $\widehat{\delta}_{(k-1)\to k} \times \frac{q_k}{\sum_{k'=k}^{k_c(max)} q_{k'}}$ specifies the number of nodes that moves from degree $k-1$ to k and also reaches its cutoff degree $k = k_c(min)$. If the fraction of k degree nodes in set_{w_i} is p_{k,w_i}, then the fraction of nodes reaching the cutoff degree k can be normalized as

$$1 - S_{k,w_i} = \frac{\frac{(k-1)p_{k-1,w_i}}{2m_{w_i}\widehat{\beta}_i}\widehat{A}_{w_i} m q_k^*}{p_{k,w_i}} \tag{18}$$

where $q_k^* = \frac{q_k}{\sum_{k'=k}^{k_c(max)} q_{k'}}$. Substituting the value of S_{k,w_i} in Eq. (17) and rearranging p_{k,w_i}, we get

$$p_{k,w_i} = \frac{(k-1)}{(k+\widehat{\alpha}_i)}\left(1 + \frac{kf_{w_i} q_k^*}{\widehat{\alpha}_i}\right) p_{k-1,w_i} \tag{19}$$

where $\widehat{\alpha}_i = \frac{2\sum w_i f_{w_i} m_{w_i}\widehat{\beta}_i}{w_i m}$.

Generalization : Continuing the calculations for $k_c(min) < k \le k_c(max)$, we obtain the generalized equation

$$p_{k,w_i} = \frac{(k-1)}{(k+\widehat{\alpha}_i)}\left(1 + \frac{kf_{w_i} q_k^*}{\widehat{\alpha}_i}\right) \left(p_{k-1,w_i} + \sum_{j=1}^{k-k_c(min)} (-1)^j p_{k-j-1,w_i} \prod_{t=1}^{j} \frac{(k-t-1) f_{w_i} q_{k-t}^*}{\widehat{\alpha}_i}\right) \tag{20}$$

The degree distribution of the entire network p_k is calculated by summing up p_{k,w_i} over all w_i's, i.e. $p_k = \sum_{i'=min}^{max} p_{k,w_{i'}}$.

A. Simulation results and inference derivation

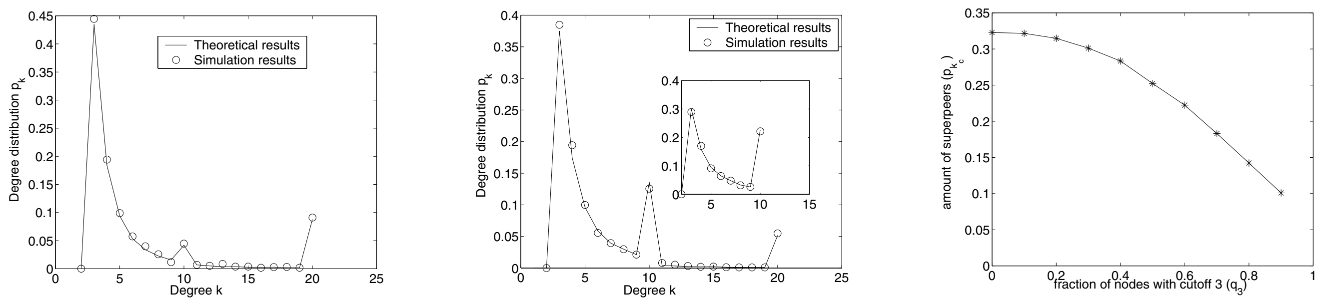
We first analyze the emerging topology and then illustrate the simulation results.

1) Emergence of superpeer nodes: Fig. 3(a) shows that if peers join with (say) v different cutoff degrees, the degree distribution of the network shows upto v (say $\widehat{v}$) spikes. We observe that the exact value of $\widehat{v}$ typically depends upon the fraction of nodes joining the network with a particular cutoff degree. Theoretically probing into the equations gives a better idea.
Let us assume that the nodes join the network with v distinct and far apart (i.e. $k_c(a_{j+1}) > k_c(a_j)+1$) bandwidths with cutoff degrees being $k_c(a_1)$, $k_c(a_2)$, $k_c(a_3)$ ... $k_c(a_v)$ respectively where $k_c(a_1)$ is the smallest cutoff and $k_c(a_v)$ is the highest one. Fraction of nodes joining with cutoff degree $k_c(a_i)$ is $q_{k_c(a_i)}$ for $1 \le i \le v$. **Condition:** $p_{k_c(a_i)-1} < p_{k_c(a_i)} > p_{k_c(a_i)+1}$ confirms the appearance of spike at degree $k_c(a_i)$. The analysis follows. Calculating $p_{k_c(a_i)+1,w_i}$ and $p_{k_c(a_i),w_i}$ and eliminating $p_{k_c(a_i)-1,w_i}$, we get $\frac{p_{k_c(a_i)+1,w_i}}{p_{k_c(a_i),w_i}} < 1$, hence for the entire network, $p_{k_c(a_i)+1} < p_{k_c(a_i)}$; that is the fraction of nodes having degree one more than some cutoff degree (say $k_c(a_i)+1$) is less than the fraction of nodes in that cutoff degree (say $k_c(a_i)$).
Similarly, from Eq. (19) we find

$$\frac{p_{k_c(a_i),w_i}}{p_{k_c(a_i)-1,w_i}} = \frac{(k_c(a_i)-1)}{(k_c(a_i)+\widehat{\alpha}_i)}\left[1 + \frac{k_c(a_i) f_{w_i} q_{k_c(a_i)}^*}{\widehat{\alpha}_i}\right] \tag{21}$$

In order to satisfy $p_{k_c(a_i)} > p_{k_c(a_i)-1}$, we find that if $q_{k_c(a_i)}$ (the fraction of nodes joined the network with cutoff degree $k_c(a_i)$) is above a threshold level, then only a mode or spike appears at degree $k_c(a_i)$.

2) Simulation results: In order to validate our theoretical framework, we simulate the bootstrapping protocol where nodes join with variable cutoff degrees. We consider that the



(a) Degree distribution of the emerging network in case 1.

(b) Degree distribution of the emerging network in case 2 and case 3 (inset).

(c) Impact of q_3 on the amount of superpeers nodes (p_{k_c}) in the network with two cutoff degrees 3 and 10.

Fig. 3. Case 1: fractions of nodes joined with cutoff degrees $3, 10$ and 20 are $0.5, 0.1$ and 0.4 respectively. Case 2: fractions of nodes joined with cutoff degrees $3, 10$ and 20 are $0.5, 0.3$ and 0.2 respectively. Inset in Fig 3(b) shows case 3 where 50% nodes joined with cutoff 3 and rest 50% joined with cutoff 10. The change in p_{k_c} (at $k_c = 10$) in the network due to the increase in q_3 (the fraction of nodes with cutoff degree 3) is plotted in Fig 3(c).

weight distribution (f_w) of the incoming nodes follows power law distribution (with exponent=2.5) [15], [19] and the nodes can have 3 different cutoff degrees $3, 10$ and 20. At the time of joining, each node establishes connections with 3 online nodes in the network i.e. $m = 3$. We assume that the 50% of nodes join through dial up connection having cutoff degrees 3. Rest 10% of nodes join through ISDN connection with cutoff degree 10 and 40% through leased line connection with cutoff degree 20. We assume that all the nodes having degree ≥ 10 can be considered as superpeer nodes [4]. The total number of nodes in the simulation system is 5000 and we perform 500 realizations. Fig 3(a) shows that the agreement between the theoretical model (Eq. (20)) and simulation is exact.

3) Measuring the amount of superpeers in the network: Fig. 3(a) shows that in case 1, total amount of superpeer nodes (i.e. degree ≥ 10) in the network is 0.1472. On the other hand, if the fraction of nodes joined with cutoff degrees $3, 10$ and 20 is $0.5, 0.3$ and 0.2 respectively (Fig 3(b), referred as case 2), the amount of superpeers in the network becomes 0.2158. If 50% of nodes join with cutoff 3 and rest 50% joins with a cutoff 10, the total amount of superpeers in the network becomes 0.2361 (inset of Fig 3(b), referred as case 3). Hence our results show that instead of joining the network through multiple high bandwidth connections, using a single bandwidth is optimal for the emergence of highest amount of superpeers in the network.

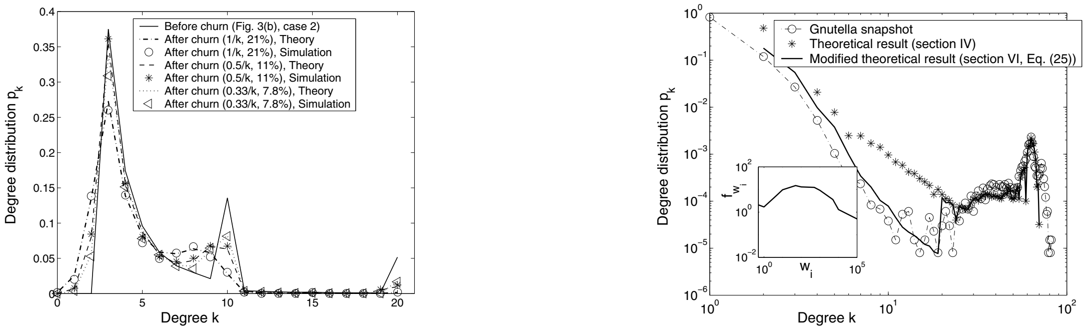
Effect of low cutoff degrees: In Fig. 3(c), we consider a situation where the nodes join with two cutoff degrees; q_3 fraction of nodes join with cutoff degree 3 and rest $q_{10} = (1 - q_3)$ fraction of nodes join with higher cutoff degree 10. In the idealistic case, when all the nodes join with cutoff degree 10 (i.e. $q_3 = 0$), the amount of superpeers in the network would be maximum ($p_{k_c} = 0.32$). The amount would decrease as some nodes with lower bandwidth hence lower cutoff degree (here 3) joins the network. The plot in Fig. 3(c) shows the rate at which p_{k_c} decreases. We find that the fraction of superpeer nodes hardly changes as long as percentage of nodes with cutoff degree 3 are less than 20%.

V. Impact of Peer Churn on the Topology of the Emerging Network

The frequent departures of the online nodes, termed as peer churn, also takes a major role in determining the topological structure [17]. In this section, we calculate the degree distribution of the emerging network in face of peer churn. In [21], we developed a theory to calculate the degree distribution of the deformed network after a fraction of nodes are removed along with the adjacent links. Mathematically, if the initial degree distribution of the network is p_k and the probability of removal of a node having degree k is f_k, then the degree distribution of the deformed network can be expressed as

$$p'_k = \sum_{q=k}^{\infty} \left(\begin{array}{c} q \\ k \end{array} \right) \phi^{q-k} (1-\phi)^k \, p^s_q \qquad (22)$$

where in large scale networks, $\phi = \frac{\sum_{i=0}^{\infty} i \; p_i \, f_i}{\sum_{k=0}^{\infty} k \, p_k}$ and $p^s_q = \frac{(1-f_q)p_q}{1-\sum_{i=0}^{\infty} p_i f_i}$. We model peer churn as the removal of nodes from the network. In p2p networks, it has been observed that 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. Hence in peer churn, it is quite realistic to assume that the probability of removal of a node is inversely proportional to the degree of that node i.e. $f_k = \frac{c}{k}$ (c is some constant). We substitute this f_k in ϕ and find that $\phi = \frac{c}{\langle k \rangle}$ where $\langle k \rangle$ is the average degree of the network. We further substitute ϕ and the initial degree distribution p_k obtained from Eq. (20) in Eq. (22) to theoretically calculate the degree distribution of the emerging network after peer churn. Fig. 4(a) shows the change in the network topology due to peer churn (theoretically and through simulation) in the network which has emerged in case 2 of Fig. 3(b). Based upon f_k ($c = 1, 0.5, 0.33$), the total percentage of nodes removed due to peer churn (f) becomes $21\%, 11\%$ and 7.8% respectively. Similar to section IV, the network size is of 5000 nodes and during simulation,



(a) Degree distribution before and after churn in case 2 of Fig. 3(b). In simulation, we remove nodes and their adjacent links based on f_k.

(b) Degree distribution of Gnutella network taken from the topological snapshot [22]. The inset shows the weight distribution of the incoming nodes [14].

Fig. 4. Fig 4(a) shows the effect of peer churn on the network emerged in Fig 3(b). Fig. 4(b) illustrates the comparative study between the real world Gnutella networks [22] and our theoretical model.

we remove the nodes and their adjacent links based on f_k. Fig. 4(a) shows that in face of churn, the bimodality in the degree distribution is still maintained. However, the disappearance of old modes and emergence of new modes in the degree distribution is an interesting observation. For example, as we gradually increase f from 7.8% to 21%, the number of superpeers in the network decreases. At the same time, new nodes with degrees even below the first maximum degree 3 appears (at degrees 1 and 2) and their amount gradually increases with the increase of f. This increase actually results due to the removal of links from the nodes having degree more than or equal to 3.

VI. Case study with Gnutella network

We simulate Gnutella network following the snapshots obtained from the Multimedia & Internetworking Research Group, University of Oregon, USA [22]. The snapshot is collected by the research group during September 2004 and the size of the network simulated from the snapshot is of $1,31,869$ nodes. In order to check whether the degree distribution of Gnutella can be explained through the developed framework, we theoretically compute the degree distribution of the emerging network (from section IV) by taking the weights from the weight distribution of the inset of Fig 4(b). The weight distribution, obtained following [14] is explained next. We assume that the weight of a node can be determined by the amount of shared files it possesses (indicates the shared resource) and inverse of node latency (indicates the node's processing power). The cumulative distribution of the amount of shared files and latency of the Gnutella peers are available in [14]. We take a joint probability distribution of these two parameters in order to get the weight distribution (inset of Fig 4(b)). During connection initiation, most of the servents initially connect to multiple online peers [5] therefore we keep $m = 2$. The probability $q_{k_c(j)}$ of joining of a node j with cutoff degree $k_c(j)$ is adjusted accordingly to fit the calculated degree distribution close to the Gnutella network.

As can be seen from Fig 4(b), our theoretical model can mimic the degree distribution of Gnutella network with reasonable accuracy, however there are some deviations. Although the higher degree nodes match almost exactly with theory, the fraction of small degree nodes in Gnutella is less than the theoretically calculated p_k. The possible reason is, due to the finite size of the web cache, the GWebcache is totally populated by the high degree nodes in the network. Henceforth, the peers having low degree do not receive any connection from the incoming node. Thus most of the low degree peer nodes remain with the low degree and subsequently the p_k for the low degree (k) nodes becomes lower than theoretically calculated value. Next we address the finite size web cache issue and modify the formalism accordingly.

A. Modifying the formalism with finite size WebCache

In order to model the finite size web cache, we assume that the nodes having degrees greater than $m'(m' < k_c(min))$ be always present in the web cache (with probability 1). However, the probability of getting a node in the webcache having degree k, such that $m \le k \le m'$ is γ. We suitably modify the rate equations described in Eqs. (4) and (6) to incorporate these assumptions. It is important to note that, as $m' < k_c(min)$, these changes may only affect the calculations of the part A of section IV.

Similar to Eq. (3), the average number of m degree nodes in the webcache acquiring links from the incoming node becomes

$$\gamma \frac{m p_{m,n,w_i}}{2 m_{w_i} \beta_i} \times A_{w_i} m \tag{23}$$

Hence **modified rate equation** for ***k*=*m***

$$\Delta n_{m,w_i} = f_{w_i} - \gamma \frac{m p_{m,n,w_i}}{2 m_{w_i} \beta_i} \times A_{w_i} m \tag{24}$$

Similarly, the rate equation for $m \le k \le m'$

$$\Delta n_{m,w_i} = \gamma \left(\frac{(k-1) p_{k-1,n+1,w_i} - k p_{k,n,w_i}}{2 m_{w_i} \beta_i} \right) \times A_{w_i} m$$

Hence the modified degree distribution becomes

$$p_k = \sum_{i=min}^{max} p_{k,w_i} f_{w_i} \tag{25}$$

$$= \begin{cases} \sum_{i=min}^{max} \frac{1}{(1+\frac{\gamma k}{\alpha_i})} f_{w_i} & k = m \\ \sum_{i=min}^{max} \frac{f_{w_i}\gamma^{k-m}}{(1+\frac{\gamma m}{\alpha_i})} \times \prod_{j=1}^{k-m} \left(\frac{k-j}{\gamma(k-j+1)+\alpha_i}\right) & m < k < m' \\ \sum_{i=min}^{max} \frac{f_{w_i}\gamma^{k-m}}{(1+\frac{\gamma m}{\alpha_i})} \times \prod_{j=1}^{k-m} \left(\frac{k-j}{k-j+1+\alpha_i}\right) & k = m'+1 \end{cases}$$

Calculation of p_k for the nodes having degree $k > m' + 1$ remains same as section III and IV. We plot the modified equations (Eq. (25)) in Fig 4(b) with $\gamma = 0.37$ and $m' = 18$ which fits the *Gnutella snapshot almost perfectly*. This indicates that the webcache is mainly populated by the nodes with degree more than 18 and only 37% of low degree nodes ($2 \leq k \leq 18$) can be present in the webcache due to its finite size.

VII. Conclusion and design guidelines to the network engineers

This paper brings forward an important message that preferential attachment may also result in a bimodal degree distribution which superpeer topologies exhibit. This happens when preferential attachment takes into consideration three features simultaneously; the node weight (quantifies the amount of resource, processing power, storage space), current degree and the available bandwidth. The developed formalism points to the fact that accurate computation of the degree distribution of a network is possible (as shown in the Fig. 4(b) for Gnutella) only based on the bootstrapping protocol and the information about the nature of web cache. In addition to that, rigorous analysis of our formalism leads to some suggestions to the network engineers which they may use to improve the servent program. Specifically two areas of servent program - bootstrapping protocol and GWebcache updation can be improved. **1. Bootstrapping:** The bootstrapping protocols can be properly modified to control the cutoff degrees of the individual nodes in the network and subsequently the amount of superpeers. Section IV-A3 shows that instead of joining the network with different bandwidth levels, using a few (or single) cutoff degrees is optimal for the emergence of high amount of superpeers in the network. In Gnutella, different nodes join the network with individual bandwidth which widely varies. However, the bootstrapping protocols can be properly designed to restrict the number of different cutoff degrees in the network (by regulating the cutoff degrees of the individual joining nodes). This measure will result the presence of superpeer nodes in the network in a large number (Fig 3(b)). **2. Updation of GWebCache:** GWebcache is periodically populated by the online peers/superpeers nodes based on the specific servent implementation [5]. Two important results have been reported (a) high weighted node can increase the fraction of superpeers only upto a level (section III-B1) (b) presence of too many high weighted nodes may be detrimental (section III-B2). Hence instead of blindly updating the 'popular' nodes in GWebCache, proper design based upon nodes' weight and degree need to be undertaken. Simultaneously the bootstrapping protocol, unlike present scheme of blindly accepting nodes from GWebCache, should sieve appropriate nodes from the host list of the web cache (section III-B4). This selection of nodes can have a strong influence on the different services like searching and file downloading latency.

Acknowledgments

This work is partially supported by the DIT, Govt. of India project grant no. $0527/T/IITKARG/014/0809/33$ and DST, Govt. of India project grant no. $SR/S3/EECE/059/2006$. Bivas Mitra acknowledges Fernando Peruani for his valuable comments and suggestions and SAP Labs India for Doctoral Fellowship.

References

[1] M. Rodriguez-Perez, O. Esparza and J. L. Muoz, "Surework: a Superpeer Reputation Framework for P2P Networks", Proceedings of the ACM Symposium on Applied Computing, Fortaleza, Ceara, Brazil, pp 2019-2023, March 2008.

[2] Y. J. Pyun and D. S. Reeves, "Constructing a Balanced, log(N)-Diameter Super-peer Topology", Proceedings of the 4^{th} International Conference on Peer-to-Peer Computing, Zurich, Switzerland, August 2004.

[3] G. Jesi, A. Montresor, and O. Babaoglu, "Proximity-Aware Superpeer Overlay Topologies", Proceedings of the 2^{nd} IEEE International Workshop on SelfMan, Vol. 3996, LNCS, pp 4357, June 2006.

[4] B. Yang and H. Garcua-Molina, "Designing a Super-Peer Networks", Proceedings of the International Conference on Data Engineering (ICDE), Los Alamitos, CA, March 2003.

[5] P. Karbhari, M. Ammar, A. Dhamdhere, H. Raj, G. Riley and E. Zegura, "Bootstrapping in Gnutella: A Measurement Study", In PAM, April 2004.

[6] X. Li, and J. Wu, "Searching Techniques in Peer-to-Peer Networks", Handbook of Theoretical and Algorithmic Aspects of Ad Hoc, Sensor, and Peer-to-Peer Networks, CRC Press, Boca Raton, USA, 2004.

[7] P. Merz, M. Priebe, and S. Wolf, "Super-Peer Selection in Peer-to-Peer Networks Using Network Coordinates", Third International Conference on Internet and Web Applications and Services, ICIW '08, Athens, Greece, June, 2008.

[8] C. GauthierDickey and C. Grothoff, "Bootstrapping of Peer-to-Peer Networks", Proceedings of IEEE DAS-P2P, Turku, Finland, August, 2008.

[9] B. Mitra, F. Peruani, S. Ghose and N. Ganguly, "Analyzing the Vulnerability of the Superpeer Networks Against Attack", ACM CCS, Alexandria, USA, November 2007.

[10] G. Bianconi, "Emergence of Weight-Topology Correlations in Complex Scale-Free Networks", Europhys. Lett. 71 1029-1035, 2005.

[11] A. L. Barabasi and R. Albert, "Emergence of Scaling In Random Networks", Science 286, 509-512, 1999.

[12] P. L. Krapivsky and S. Redner, "Organization of Growing Random Networks", Phys. Rev. E 63, 066123, 2001.

[13] G. Bianconi and A.-L. Barabasi, "Competition and Multiscaling in Evolving Networks", Europhys. Lett. 54, pp 436-442, 2001.

[14] P. Saroiu, K. Gummadi, S. D. Gribble, "A Measurement Study of Peer-to-Peer File Sharing Systems", In Proceedings of Multimedia Computing and Networking (MMCN), January 2002

[15] L. Ronga and I. Burnett, "Dynamic Resource Adaptation in a Heterogeneous Peer-to-Peer Environment", Second IEEE Consumer Communications and Networking Conference, pp 416-420, 3-6 January 2005.

[16] Clip2 DSS. Gnutella Protocol Specification v0.4. http://www.limewire.com, 2007.

[17] B. Mitra, F. Peruani, S. Ghose and N. Ganguly, "Brief Announcement: Measuring Robustness of Superpeer Topologies", ACM PODC, Portland, USA, August 2007.

[18] M. Conrad and H. J. Hof, "A Generic, Self-organizing, and Distributed Bootstrap Service for Peer-to-Peer Networks", Proceedings of the Second International Workshop on Self-Organizing Systems, IWSOS 2007, The Lake District, UK, September 11-13, 2007.

[19] S. Meng, C. Shi, D. Han, X. Zhu, and Y. Yu, "A Statistical Study of Today's Gnutella", Proceedings of APWeb, pp.189-200, Harbin, China, January 16-18, 2006.

[20] G. Pandurangan, P. Raghavan, and E. Upfal, "Building Low-Diameter P2P Networks", IEEE Journal on Selected Areas in Communications, Vol. 21, pp. 995-1002, Aug 2003.

[21] B. Mitra, N. Ganguly, S. Ghose and F. Peruani, "Generalized Theory for Node Disruption in Finite-size Complex Networks", Physical Review E, 78, 026115, 2008.

[22] "Gnutella snapshot", http://mirage.cs.uoregon.edu/P2P/info.cgi.