Chapter 4

Attack and stability of superpeer networks

In the previous chapter, we have reported the impact of peer churn on superpeer network with the help of an analytical framework. In this chapter, we propose another analytical framework to understand the impact of different types of attacks on superpeer networks. From chapter 4, we can calculate the stability of uncorrelated large graphs in the same fashion as the previous, however this is more sophisticated than the framework of chapter 3 in different aspects.

1. In addition to the stability of overall network, the framework of this chapter gives more insights regarding the topology of the network. For instance, the removal of nodes along with their adjacent edges changes the topology of the network. The degree distribution of this deformed network after attack can be calculated with the help of this framework.

2. There are many results that have been derived for infinite networks (similar to previous framework), however, little is known about the stability of finite size networks. The framework developed in this chapter sheds some light on finite size network by proposing an alternative expression for the percolation threshold.

3. Most of the real world networks like Gnutella exhibit degree-degree correlation in the topological structure. Hence understanding the stability of these networks needs to include degree-degree correlation in the calculation (which was not possible in the previous framework). We show that, a little modification of the current framework makes it suitable for the analysis of correlated networks also.

The chapter is organized in the following way. In section 4.1, we develop the analytical framework for stability analysis. In section 4.2 we use the framework to analyze the stability of superpeer networks in face of degree independent attack as well as degree dependent attack modeled in chapter 3. We show that the degree dependent attack can be used as an unified attack model as other node disturbances may be reproduced by regulating some parameter [109]. We validate our theoretical framework with the help of stochastic simulation. The validation is done in two ways depending upon the generation of superpeer networks, as illustrated in chapter 3. We start with simple models of superpeer networks, namely bimodal network and mixed poisson network which are simple enough to understand and analyze while at the same time they capture the essential features of the superpeer networks (section 4.2.2). Our framework unfolds various issues such as (i) the available knowledge regarding the topology that helps attackers to breakdown the network (section 4.2.3) (ii) the effect of finiteness of network size on the network stability (section 4.2.4). Afterwards we implement the attack dynamics on the commercial peer-to-peer networks namely Gnutella (section 4.3). Gnutella network is simulated both from the bootstrapping protocol followed by the different Gnutella clients like limewire, mutella etc [81] and from the topological snapshots obtained from [1]. We identify some deviations between theoretical and simulation results due to the presence of degree-degree correlation in Gnutella network. In section 4.4, we further refine our framework to include the degree-degree correlation factor and show that the modified theoretical model gives good agreement with simulated results.

4.1 Development of the analytical framework

In this section, we present the detail derivation of the critical condition for measuring the stability of peer to peer networks undergoing any kinds of attacks [107]. We start out by repeating some definitions mentioned before. Let p_k be the probability of finding a node chosen uniformly at random with degree k. Let f_k be the probability that a node of degree k is removed after the attack. Correspondingly $1 - f_k$ is the probability that a node of degree k survives the attack. In our framework, degree distribution p_k models the ensemble of p2p topologies and f_k models the disruptive events that take place in the network. We are going to establish the relationship between stability, p_k and f_k. This is done as a two step process; in the first step, we calculate the degree distribution of the deformed network after attack. Subsequently in the second step, we use this expression to derive the critical condition of stability of p2p networks against attack.

4.1.1 Deformed topology after attack

In this subsection, we theoretically compute the degree distribution of the deformed topology p'_k after performing an attack on the p2p network of size N with initial degree distribution p_k. The attack in the network can be thought of in the following way. The first step in the attack is to select the nodes that are going to be removed according to the probability distribution f_k. After the selection of the nodes, we divide the network into two subsets, one subset contains the surviving nodes (S) while the other subset comprises of the nodes that are going to be removed (R). This is illustrated in Fig. 4.1. The degree distribution of the surviving subset S is $(1 - f_k)p_k$ while the subset of nodes to be removed R (that is the edges connecting set S and set R) still exist. However, when these nodes are actually removed, the degree distribution of the surviving nodes S is changed due to the removal of the E edges that run between these two subsets.

To calculate the degree distribution after the attack, we have to estimate E. The

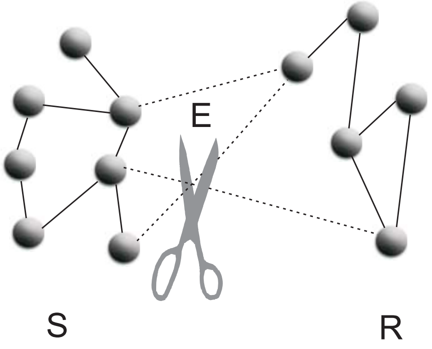


Figure 4.1: The scheme illustrates an attack as consisting of two steps: selection of nodes to be removed (set of removed nodes, R), and cutting of the edges E that run from the surviving nodes (set of surviving nodes, S) to the set of removed nodes R. As the scheme shows, the attack affects the degree of the surviving nodes.

total number of edge tips[1] in the surviving subset S including E links that are going to be removed can be expressed by the sum $\sum_{j=0}^{\infty} j\, n_j\, (1 - f_j)$ where $n_j = Np_j$ is the total number of nodes in the network having degree j. Now kn_kf_k gives the total number of edge tips connected with all the k degree nodes in the removed subset R. Therefore $\sum_k kn_kf_k$ becomes the total number of tips in R. Hence the probability of a randomly chosen tip of an edge to be removed becomes $\frac{\sum_k kn_kf_k}{\sum_k kn_k}$. Subsequently the probability of a randomly chosen tip of an edge to be removed (i.e. member of set R) and another tip of that edge being connected to either set S or R becomes $\frac{\sum_k kn_kf_k}{\sum_k kn_k - 1}$ (since a tip cannot be connected to itself). As the network is uncorrelated, it is equally probable that the other end of the removed tip (member of set R) is connected to the nodes of set S or set R. Assuming this unbiasness, the total number of edge tips in set R connected to the nodes of the set S can be expressed as

$$E = \left(\frac{\sum_{i=0}^{\infty} i\, n_i\, f_i}{(\sum_{k=0}^{\infty} k\, n_k) - 1} \right) \sum_{j=0}^{\infty} j\, n_j\, (1 - f_j) \tag{4.1}$$

Knowing this, the probability ϕ of finding an edge in the surviving subset S, that is

[1] We assume that each edge consists of two end tips. Hence the total number of tips in the network is twice the number of edges.

connected to a node of the other subset R can be expressed as

$$\phi = \frac{E}{\sum_{i=0}^{\infty} i\, n_i\, (1 - f_i)} = \frac{E}{N \sum_{i=0}^{\infty} i\, p_i\, (1 - f_i)} = \frac{\sum_{i=0}^{\infty} i\ p_i\ f_i}{(\sum_{k=0}^{\infty} k\, p_k) - 1/N}\,. \tag{4.2}$$

In large scale networks, $\lim_{N\to\infty} \phi = \frac{\sum_{i=0}^{\infty} i\ p_i\ f_i}{\sum_{k=0}^{\infty} k\, p_k}$

The probability p_q^s of finding a node with degree q in the surviving subset S (before cutting the E edges) simply becomes

$$p_q^s = \frac{(1 - f_q)p_q}{1 - \sum_{i=0}^{\infty} p_i f_i}\,. \tag{4.3}$$

The removal of nodes can only lead to a decrease in the degree of a survived node. If we find a node of degree k that has survived, it can be due to the fact that originally its degree was $k + q$ and k of its edges survived while q (q may be zero also) got removed. For example, the fraction of nodes having degree k after attack i.e. p'_k is given by the fraction of p_k^s nodes, who did not lose any link, and a fraction of p_{k+1}^s nodes who lost one link but rest k links survived, a fraction of p_{k+2}^s nodes who lost two links but rest k links survived and so on. Hence using the concept of binomial distribution and from the equations (4.2) and (4.3), we obtain the following expression for p'_k:

$$p'_k = \sum_{q=k}^{\infty} \begin{pmatrix} q \\ k \end{pmatrix} \phi^{q-k} (1 - \phi)^k\, p_q^s\,. \tag{4.4}$$

Eq. (4.4) can be iteratively evaluated by replacing p_k with p'_k into Eqs. (4.1) to (4.4).

4.1.2 Critical condition for stability

In this section, we derive the critical condition for stability of the peer to peer networks after attack. In order to do that, we utilize the expression of the deformed degree distribution p'_k after removal of nodes. According to [28, 115], the critical condition for the stability of giant component can be expressed as

$$\kappa' = \frac{\langle k^2 \rangle'}{\langle k \rangle'} > 2\,, \tag{4.5}$$

where $\langle k\rangle'$ and $\langle k^2\rangle'$ refer to the first and second moments of the degree distribution after the attack. The critical condition $\kappa' = 2$ determines the point at which the network breaks down. To compute $\langle k\rangle'$ and $\langle k^2\rangle'$ of the modified network, we utilize the generating function $G_0(x) = \sum_k p'_k\, x^k$, which reads:

$$G_0(x) = \sum_{k=0}^{\infty}\sum_{q=k}^{\infty} \left(\begin{array}{c} q \\ k \end{array} \right) \phi^{q-k}(1-\phi)^k p^s_q\, x^k \,. \tag{4.6}$$

After exchanging the order of the sum, the Binomial theorem can be applied, and we obtain:

$$G_0(x) = \sum_{k=0}^{\infty} p^s_k \left((x-1)(1-\phi)+1\right)^k . \tag{4.7}$$

From Eq. (4.7), the first two moments can be easily computed as $\langle k\rangle' = dG_0(1)/dx$ and $\langle k^2\rangle' = d^2G_0(1)/dx^2 + dG_0(1)/dx$, and the critical condition given by Eq. (4.5) takes the form:

$$(1-\phi)\,\frac{\langle k^2\rangle - \sum_{q=0}^{\infty} f_q\, p_q\, q^2}{\langle k\rangle - \sum_{q=0}^{\infty} f_q\, p_q\, q} + \phi = 2\,, \tag{4.8}$$

where $\langle k\rangle$ and $\langle k^2\rangle$ refer to the first and second moments of the degree distribution before the attack. Replacing ϕ by Eq. (4.2) and assuming $N >> 1$, we obtain

$$\sum_{k=0}^{\infty} kp_k(k(1-f_k) - (1-f_k) - 1) = 0 \tag{4.9}$$

which is the critical condition of stability in any large scale uncorrelated peer to peer networks. Comparing Eqs. 4.9 and 3.15, we conclude that, this critical condition is exactly same as that developed in chapter 3.

4.2 Effect of attacks upon the superpeer networks

In this section, we formally analyze the effect of attacks on the superpeer networks with the help of the developed framework. Two kinds of attacks, namely deterministic attack and degree dependent attack are discussed separately. The attack models are already described in chapter 3. First of all, we show the effect of these attacks on the topological deformation of the network. This phenomenon has been modeled

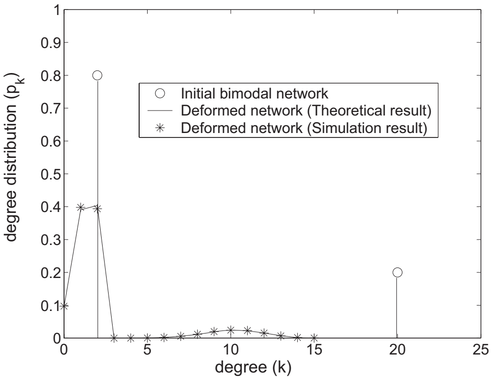


Figure 4.2: Topological deformation of the superpeer networks in face of deterministic attack. After the attack, 10% of nodes are removed. This 10% of nodes correspond to the 50% of the superpeer nodes whose degree is 20. The initial bimodal network and deformed network after attack are shown in the figure. The theoretically calculated degree distribution (p'_k) is verified through simulation.

using Eq. (4.4) and validated through simulations. Next we evaluate the stability of superpeer networks against these kinds of attacks and establish a relationship between them.

4.2.1 Analysis of deterministic attack

We consider superpeer networks with peer degree $k_l = 2$ and superpeer degree $k_m = 20$ and assume that 80% of nodes in the network are peers. Suppose 10% of nodes are removed through deterministic attack which signifies that 50% of superpeers get removed. We calculate the new degree distribution after attack (p'_k) by Eq. (4.4) and compare the results with simulation. Fig. 4.2 shows the good agreement between the theoretical and simulation results which confirms the success of our model.

Stability of the superpeer networks is challenged by attack on prominent peers or superpeers. In this section, we analyze the effect of this kind of targeted attack upon superpeer networks where r is the fraction of peers and rest are superpeers. In the case of targeted attack two cases may arise:

Case 1 Removal of a fraction of superpeers is sufficient to disintegrate the network.

Case 2 Removal of all the superpeers is not sufficient to disintegrate the network. Therefore, we need to remove some of the peer nodes along with the superpeers.

We analyze these two cases separately with the help of our analytical framework. First we consider the bimodal networks as our superpeer networks model. Next we extend the analysis for the more sophisticated mixed poisson networks.

Bimodal Networks

From Eq. (4.9) the critical condition for the stability of the superpeer networks can be rewritten as

$$\sum_{k=k_l,k_m} k(k-1)p_k q_k = \langle k \rangle \tag{4.10}$$

The equation can be further expanded as below to differentiate between peers and superpeers

$$k_l(k_l-1)p_{k_l}q_{k_l} + k_m(k_m-1)p_{k_m}q_{k_m} = \langle k \rangle \tag{4.11}$$

Case 1: In this case, removal of a fraction of superpeers is sufficient to disintegrate the network. If f_{sp} be the critical fraction of superpeer nodes, removal of which disintegrates the giant component, then $q_k = 1$ for $k = k_l$ and $q_k = 1 - f_{sp}$ for $k = k_m$. Hence according to Eq. (4.11),

$$\sum_{k=k_l} k(k-1)p_k + \sum_{k=k_m} k(k-1)p_k(1-f_{sp}) = \langle k \rangle$$

$$\Rightarrow f_{sp} = 1 - \frac{\langle k \rangle - k_l(k_l-1)p_{k_l}}{k_m(k_m-1)p_{k_m}}$$

As the fraction of superpeer nodes in the network is $(1-r)$, then percolation threshold for case 1 becomes $f_{tar} = (1-r) \times f_{sp}$

$$\Rightarrow f_{tar} = (1-r)\left(1 - \frac{\langle k \rangle - k_l(k_l-1)r}{k_m(k_m-1)(1-r)}\right) \tag{4.12}$$

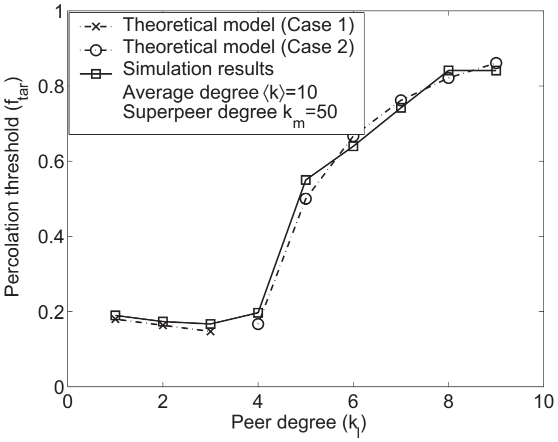


Figure 4.3: Stability of the superpeer networks in face of deterministic attack (Comparative study between theoretical and simulation results). Here X-axis represents the peer degree (k_l) and Y-axis represents the corresponding percolation threshold (f_{tar}). We keep the average degree $\langle k \rangle = 10$ and mean superpeer degree $\langle k_{sp} \rangle = 50$ fixed. Case 1 and case 2 of the theoretical model represent Eqs. (4.12) and (4.15) respectively.

Case 2: Here we have to remove f_p fraction of peer nodes along with all the superpeers to breakdown the network. Therefore $q_k = 1 - f_p$ for $k = k_l$ and $q_k = 0$ for $k = k_m$. Hence according to Eq. (4.11),

$$k_l(k_l - 1)p_{k_l}(1 - f_p) = \langle k \rangle \tag{4.13}$$

$$\Rightarrow f_p = 1 - \frac{\langle k \rangle}{k_l(k_l - 1)p_{k_l}} \tag{4.14}$$

Therefore the total fraction of nodes required to be removed to disintegrate the network for case 2 becomes $f_{tar} = rf_p + (1 - r)$.

$$\Rightarrow f_{tar} = r\left(1 - \frac{\langle k \rangle}{k_l(k_l - 1)r}\right) + (1 - r) \tag{4.15}$$

Transition point: The transition from case 1 to case 2 can be easily marked by observing the value of percolation threshold f_{tar}. While calculating using Eq. (4.12) (case 1), if the value of f_{tar} exceeds the fraction of superpeers in the network $(1-r)$, it indicates that removal of all the superpeers is not sufficient to disrupt the network. Hence subsequently we enter into case 2 and start using Eq. (4.15) to find percolation threshold.

We validate our theoretical model of attack on superpeer network with the help of simulation. During simulation, initially only high degree superpeer nodes in the network are removed gradually until the percolation point is reached. If the percolation point is not reached even after removing of all the superpeers, we remove a fraction of peers along with the superpeers to breakdown the network. We perform each experiment for 500 times and take the average of the percolation threshold obtained in each of them. Superpeer networks with average degree $\langle k \rangle = 10$ and superpeer degree $k_m = 50$ are considered for case study. We increase the peer degree k_l gradually (the peer fraction changes accordingly) and observe the change in the percolation threshold f_{tar} (Fig. 4.3).

Observations:

a. In the networks with peer degree $k_l = 1, 2$ and 3, the removal of only a fraction of superpeers causes breakdown thus making these networks vulnerable. Moreover, increase of peer degree from 1 to 2 and 3 further reduces the fraction of superpeers in the network which makes networks with $k_l = 2, 3$ more vulnerable. Normal wisdom would expect the attack vulnerability of the network to reduce with the decrease in the fraction of superpeers. But the opposite happens here. The slope of the Eq. (4.12) with respect to k_l becomes

$$\frac{\triangle f_{tar}}{\triangle k_l} = \frac{1}{M_2} \frac{(M_1 - k_l M_3 + M_4 k_l^2) - (M_5 - k_l)(2M_5 k_l - M_3)}{(M_5 - k_l)^2} \tag{4.16}$$

where M_1, M_2, M_3, M_4, M_5 are constants dependent on superpeer degree k_m and average degree $\langle k \rangle$. The slope of the curve at the points $k_l = 1, 2$ and 3 becomes negative which signifies that the attack vulnerability of the network increases with k_l. Along with the theoretical justification, this can also be explained by looking into the micro dynamics. In this zone (at $k_l = 2, 3$), although peers have a larger share in the network, yet it is not large enough to form effective connections within themselves. Therefore the stability of the network is still entirely dependent on the high degree superpeers, hence now attacking even a smaller fraction breaks down the network.

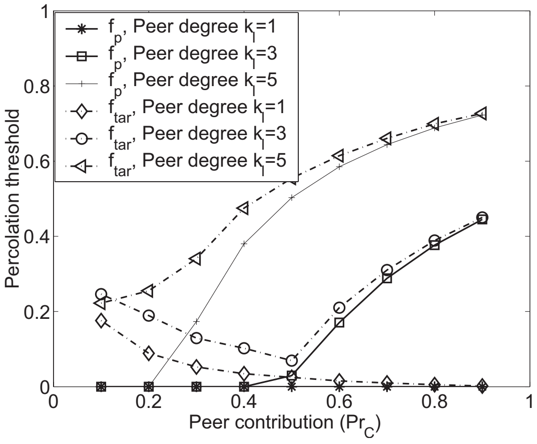


Figure 4.4: The plot represents the impact of peer contribution Pr_C upon the stability of the network against attack. f_p represents the fraction of peers required to be attacked to dissolve the network and f_{tar} indicates the corresponding percolation threshold.

b. However as peer degree increases beyond 4, the transition from case 1 to case 2 occurs. In this region a fraction of peers is required to be removed even after removal of all the superpeers to dissolve the network. The slope of the Eq. (4.15) with respect to k_l becomes

$$\frac{\triangle f_{tar}}{\triangle k_l} = \frac{k}{k_l^2(k_l - 1)} + \frac{k}{k_l(k_l - 1)} \tag{4.17}$$

Hence the slope of the Eq. (4.15) becomes positive for any peer degree $k_l > 1$ which indicates that stability of the network increases with the increase of peer degree. In practice, the high degree peers connect among themselves and they are not entirely dependent on superpeers for connectivity. This results in the steep increase of stability of the network with peer degree $k_l \geq 5$.

Impact of peer contribution

Similar to churn, we investigate the impact of (pure) peer contribution upon stability of the network due to attack. In order to understand the influence of the degree of pure peers, we consider the networks with $k_l = 1, 3, 5$. Three sets of networks are generated having $k_l = 1, 3$ and 5 respectively for individual peer contribution Pr_C ($0.1 \leq Pr_C \leq 0.9$). In order to do that, we choose fraction of peers r uniformly at random and adjust superpeer degree k_m accordingly to keep the peer contribution Pr_C and peer degree k_l constant. This procedure is followed to generate one hundred networks for each set. We restrict superpeer degree $k_m \geq 20$ in order to generate realistic superpeer networks. We theoretically compute the percolation threshold (f_{tar}) and fraction of peers and superpeers required to be removed (f_p and f_{sp} respectively) for individual network and calculate their average for individual k_l. This expected fraction of peers required to be removed f_p and percolation threshold f_{tar} is plotted with respect to the peer contribution Pr_C (Fig. 4.4). The theoretical model is sufficient for analysis as the model has been already validated through simulation.

Observations:

1. It can be observed from Fig. 4.4 that superpeer networks having peer degree $k_l = 1$ can be disintegrated without attacking peers at all for any peer contribution Pr_C. This kind of attack belongs to case 1 of the attack model.

2. The peers of the superpeer networks having peer contribution $Pr_C \leq 0.2$ does not have any impact upon the stability of the network. This is true for low as well as high degree peers.

3. The influence of high degree peers increases with the increase of peer contribution. At $Pr_C = 0.3$, a fraction of peers is required to be removed to disintegrate the networks having peer degree $k_l = 5$. The impact of high degree peers upon the stability of the network becomes more eminent as peer contribution $Pr_C \geq 0.5$. In this region, a significant fraction of peers is required to be removed for all the networks having peer degree $k_l = 3, 5$. This kind of attack belongs to case 2 of the attack model.

4. Increase in peer contribution $Pr_C \geq 0.4$ brings the percolation threshold f_{tar} and fraction of peers needed to be attacked f_p close to each other which implies that stability of these networks is primarily dependent upon the stability of the peers.

5. It is interesting to observe that peer contribution Pr_C has two opposite effects upon stability of the networks depending on the peer degree k_l. The percolation threshold f_{tar} increases with peer contribution Pr_C for $k_l = 3, 5$, but gradually reduces for $k_l = 1$. The reason behind this is, stability of the networks with peer degree $k_l = 1$ is entirely dependent upon superpeers. Since

increase in peer contribution decreases superpeer contribution, it decreases stability of these networks also. On the other hand, peers having degree $k_l \geq 3$ have many connections among themselves, hence stability of these networks is more dependent upon peer contribution. Therefore, percolation threshold f_{tar} increases with peer contribution Pr_C.

6. Peer degree $k_l = 3$ exhibits some kind of trade off between the impact of peer and superpeer contribution upon stability. Superpeer contribution becomes more predominant for lower values of Pr_C ($Pr_C < 0.5$) which degrades the percolation threshold against attack. However as peer contribution Pr_C increases beyond 0.5, superpeer contribution reduces hence attacking peers along with superpeers is necessary to destroy the network. This increases the percolation threshold f_{tar} i.e. the stability of the network as well.

Mixed Poisson Networks

Similar to bimodal networks, in mixed poisson networks also we have two different cases. We analyze these two cases separately with the help of our analytical framework. From Eq. (4.9) the critical condition for the stability of the giant component can be rewritten as

$$\sum_{k=0}^{\infty} k(k-1)p_k q_k = \langle k \rangle$$

The equation can be further expanded as below to differentiate between peers and superpeers

$$\sum_{k=0}^{k_{max}-1} k(k-1)p_k q_k + \sum_{k=k_{max}}^{\infty} k(k-1)p_k q_k = \langle k \rangle \tag{4.18}$$

where all the nodes having degree less than k_{max} are peers and rest are superpeers.

Case 1: In this case, removal of a fraction of superpeers is sufficient to disintegrate the network. If f_{sp} be the critical fraction of superpeer nodes, removal of which disintegrates the giant component then $q_k = 1$ for $k < k_{max}$ and $q_k = 1 - f_{sp}$ for $k \geq k_{max}$. Hence according to Eq. (4.18),

$$\sum_{k=0}^{k_{max}-1} k(k-1)p_k + \sum_{k=k_{max}}^{\infty} k(k-1)p_k(1 - f_{sp}) = \langle k \rangle$$

$$\Rightarrow f_{sp} = 1 - \frac{\langle k\rangle - \sum_{k=0}^{k_{max}-1} k(k-1)p_k}{\sum_{k=k_{max}}^{\infty} k(k-1)p_k}$$

As the fraction of superpeer nodes in the network is $(1-r)$, then percolation threshold for case 1 becomes $f_t = (1-r) \times f_{sp}$

$$\begin{aligned}\Rightarrow f_t &= (1-r)\left(1 - \frac{\langle k\rangle - \sum_{k=0}^{k_{max}-1} k(k-1)p_k}{\sum_{k=k_{max}}^{\infty} k(k-1)p_k}\right)\\ &= (1-r)\left(1 - \frac{\langle k\rangle - r\sum_{k=0}^{\langle k_p\rangle+\delta} k(k-1)\frac{\langle k_p\rangle^k e^{-\langle k_p\rangle}}{k!}}{(1-r)\sum_{k=\langle k_p\rangle+\delta+1}^{\infty} k(k-1)\frac{\langle k_{sp}\rangle^k e^{-\langle k_{sp}\rangle}}{k!}}\right) \qquad (4.19)\end{aligned}$$

where mean peer degree $\langle k_p\rangle = \frac{\langle k\rangle-(1-r)\langle k_{sp}\rangle}{r}$ and we choose suitable value of δ depending on the standard deviation of the Poisson distribution. δ ensures the inclusion of all peer and superpeer degrees around their respective means $\langle k_p\rangle$ and $\langle k_{sp}\rangle$ during the calculation of above equations.

Case 2: Here we have to remove f_p fraction of peer nodes alongwith all the superpeers to breakdown the network. Therefore $q_k = 1 - f_p$ for $k < k_{max}$ and $q_k = 0$ for $k \geq k_{max}$. Hence according to Eq. (4.18),

$$\sum_{k=0}^{k_{max}-1} k(k-1)p_k(1-f_p) = \langle k\rangle$$

$$\Rightarrow f_p = 1 - \frac{\langle k\rangle}{\sum_{k=0}^{k_{max}-1} k(k-1)p_k}$$

Therefore the total fraction of nodes required to be removed to disintegrate the network for case 2 becomes $f_t = rf_p + (1-r)$.

$$\begin{aligned}\Rightarrow f_t &= r\left(1 - \frac{\langle k\rangle}{\sum_{k=0}^{k_{max}-1} k(k-1)p_k}\right) + (1-r)\\ &= r\left(1 - \frac{\langle k\rangle}{r\sum_{k=0}^{\langle k_p\rangle+\delta} k(k-1)\frac{\langle k_p\rangle^k e^{-\langle k_p\rangle}}{k!}}\right) + (1-r) \qquad (4.20)\end{aligned}$$

where mean peer degree $\langle k_p\rangle = \frac{\langle k\rangle-(1-r)\langle k_{sp}\rangle}{r}$.

Transition point: The transition from case 1 to case 2 can be easily marked by observing the value of percolation threshold f_t. While calculating using Eq. (4.19) (case 1), if the percolation threshold f_t exceeds the fraction of superpeers in the

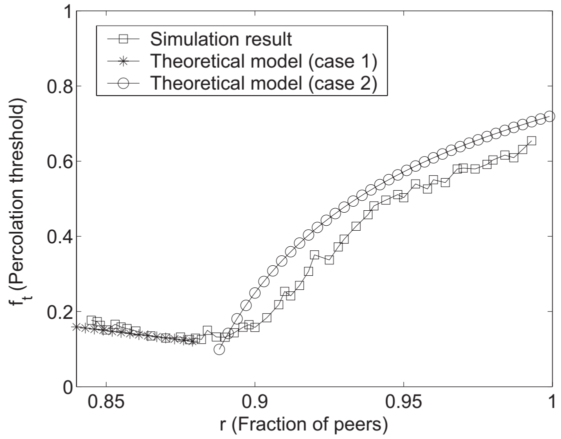


Figure 4.5: The above plot represents the behavior of the mixed poisson network in face of deterministic attack found experimentally and compares it with the proposed theoretical model. Here X-axis represents the fraction of peer nodes (r) that exist in the network and Y-axis represents the corresponding percolation threshold (f_t). We keep the average degree $\langle k \rangle = 5$ and mean superpeer degree $\langle k_{sp} \rangle = 30$ fixed. Case 1 and case 2 of the theoretical model represent Eqs. (4.19) and (4.20) respectively.

network $(1 - r)$, it indicates that removal of all the superpeers is not sufficient to disrupt the network. Hence subsequently we enter into case 2 and start using Eq. (4.20) to find percolation threshold.

We validate our theoretical model of attack on mixed poisson network with the help of simulation. In simulation, we consider a mixed poisson network with average degree $\langle k \rangle = 5$ and mean superpeer degree $\langle k_{sp} \rangle = 30$. We increase the fraction of peers gradually keeping average degree $\langle k \rangle = 5$ fixed and observe the change in the percolation threshold f_t (Fig. 4.5). It is important to note that when the fraction of superpeers in the network is high, it is possible to breakdown the network only by removing a fraction of superpeers and modeled as case 1 (Eq. (4.19)). But when the fraction of superpeers is below some threshold, a fraction of peers should be attacked alongwith the superpeers to stop percolation in the network and modeled as case 2 (Eq. (4.20)).

Summarization: In this section, the impact of deterministic attack on the stability of superpeer networks has been analyzed. We have shown that the networks having peer degree $k_l \leq 3$ are very much vulnerable and removal of only a small fraction of superpeers causes the breakdown of the network. But as the peer degree increases, the stability of the network increases as well. We have observed that peer contribution plays a major role in the network stability, specially for the networks with high peer degree (say $k_l \geq 3$). In this case, a fraction of peers are required to be removed along with all the superpeers in the network. However, depending upon the peer degree k_l, peer and superpeer contributions exhibit two opposite forces in percolation threshold due to their individual influence on the connectivity of the network. This phenomenon becomes much more predominant for the networks with $k_l \geq 3$. We have observed that both bimodal networks and mixed poisson networks qualitatively exhibit similar kinds of behavior against deterministic attack. We henceforth use bimodal network as the representative superpeer network for the analysis of degree dependent attack.

4.2.2 Analysis of degree dependent attack

In this kind of attack, the probability of removal of a node of degree k is directly proportional to k^γ where $\gamma \geq 0$ is a real number and represents the information available to the attacker about the topological structure of the network. Similar to the deterministic attack, in this case also we compute the deformed degree distribution p'_k after attack and validate the results through simulations. Without the loss of generality, we use bimodal network as the representative topology to model superpeer networks. We consider a superpeer network with peer degree $k_l = 2$ and superpeer degree $k_m = 10$ where 80% of the nodes are peers. The probability of removal of a node is proportional to its degree, i.e. $f_k = \frac{k}{k_m+1}$ (so $\gamma = 1$). The theoretically computed p'_k (using Eq. (4.4)) and simulation results are shown in Fig. 4.6. Next we analyze the effect of degree dependent attack upon the stability of the superpeer networks. With proper normalization, probability of removal of a node having degree k becomes $f_k = \frac{k^\gamma}{C}$ where C is the normalization constant.

As mentioned in bimodal degree distribution, let r be the fraction of peers with degree k_l while rest are superpeers of degree k_m. If $\langle k \rangle$ is the average degree of the network, then

$$p_{k_l} = r = \frac{k_m - \langle k \rangle}{k_m - k_l} \qquad p_{k_m} = (1-r) = \frac{\langle k \rangle - k_l}{k_m - k_l}$$

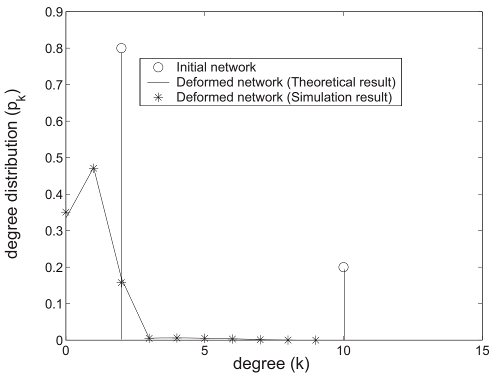


Figure 4.6: Topological deformation of the superpeer networks in face of degree dependent attack. The nodes are removed from the network with $f_k = \frac{k}{k_m+1}$. The initial bimodal network and the deformed network after attack p'_k are shown in the figure.

From Eq. (4.9) the critical condition for the stability of the giant component can be rewritten as

$$\sum_{k=k_l,k_m} k(k-1)p_k(1-f_k) = \langle k\rangle$$

$$\begin{aligned} \Rightarrow\ & \langle k^{\gamma+2}\rangle - \langle k^{\gamma+1}\rangle = C(\langle k^2\rangle - 2\langle k\rangle) \\ \Rightarrow\ & rk_l^{\gamma+1}(k_l-1) + (1-r)k_m^{\gamma+1}(k_m-1) = \\ & C(\langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle) \end{aligned} \tag{4.21}$$

where θ^{th} moment of the bimodal degree distribution can be written as $\langle k^\theta\rangle = k_m^\theta p_{k_m} + k_l^\theta p_{k_l}$. The solution of Eq. (4.21) yields a particular value of γ, say γ_c (termed as critical exponent) and the percolation threshold becomes

$$f_c^{\gamma_c} = r\frac{k_l^{\gamma_c}}{C} + (1-r)\frac{k_m^{\gamma_c}}{C} \tag{4.22}$$

In order to evaluate the disintegration point, proper assignment of the value of normalizing constant C is necessary. Since f_k should be $\leq 1\ \forall k$, hence the minimum value of $C = k_m^\gamma$. Assuming this condition, Eq. (4.21) becomes

$$\begin{aligned} rk_l^{\gamma+1}(k_l-1) + (1-r)k_m^{\gamma+1}(k_m-1) \ &\geq \\ k_m^\gamma(\langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle) & \end{aligned} \tag{4.23}$$

The solution set of the above inequality (say S_{γ_c}) can be bounded (where $0 \leq \gamma_c \leq \gamma_c^{bd}$) or unbounded (where $0 \leq \gamma_c \leq +\infty$). Each critical exponent $\gamma_c \in S_{\gamma_c}$ specifies the fraction of peers and superpeers required to be removed to breakdown the network. Assuming equality of Eq. (4.23) and hence obtaining minimum value of C, each γ_c results in the corresponding normalizing constant

$$C_{\gamma_c} = \frac{rk_l^{\gamma_c+1}(k_l-1)+(1-r)k_m^{\gamma_c+1}(k_m-1)}{\langle k\rangle(k_m+k_l)-k_m-2\langle k\rangle} \tag{4.24}$$

Hence the fraction of peers and superpeers need to be attacked are

$$f_p^{\gamma_c} = \frac{k_l^{\gamma_c}}{C_{\gamma_c}} \qquad f_{sp}^{\gamma_c} = \frac{k_m^{\gamma_c}}{C_{\gamma_c}} \tag{4.25}$$

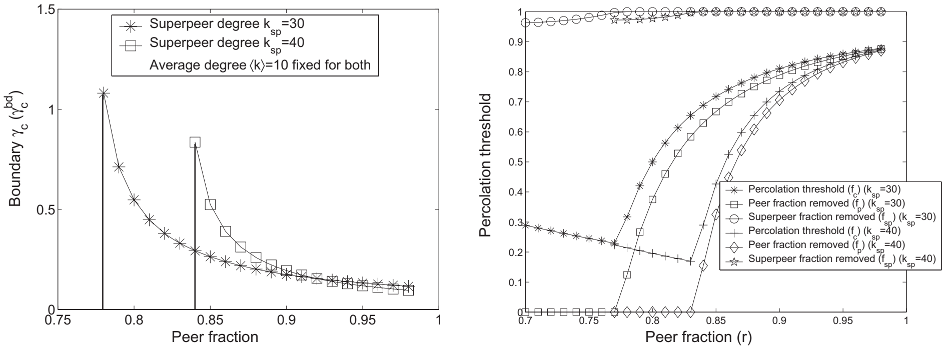
respectively and the total fraction of removed nodes $f_c^{\gamma_c}$ is obtained from Eq. (4.22). The $f_c^{\gamma_c}$ depends upon the critical exponent $\gamma_c \in S_{\gamma_c}$ and normalizing constant C_{γ_c}. The nature of the solution set S_{γ_c} has profound impact upon the behavior of $f_p^{\gamma_c}$, $f_{sp}^{\gamma_c}$ as well as $f_c^{\gamma_c}$. The breakdown of the network can be due to one of the three situations noted below.

1. The removal of all the superpeers along with a fraction of peers.
2. The removal of only a fraction of superpeers.
3. The removal of some fraction of both superpeers and peers.

The above mentioned three cases are discussed one by one with example.

Case 1 : Removal of all superpeers along with a fraction of peers

Networks having bounded solution set S_{γ_c} where $0 \leq \gamma_c \leq \gamma_c^{bd}$ exhibit this kind of behavior at the maximum value of the solution $\gamma_c = \gamma_c^{bd}$. Here the fraction of superpeers removed become $f_{sp}^{\gamma_c^{bd}} = 1$ and fraction of peers removed $f_p^{\gamma_c^{bd}} = \frac{k_l^{\gamma_c^{bd}}}{C_{\gamma_c^{bd}}}$. We consider superpeer networks with superpeer degrees $k_m = 30, 40$ and average degree $\langle k\rangle = 10$ and theoretically study the stability of the networks due to the change in the peer fraction r. The results of the case study are noted in Fig. 4.7. It can be



(a) Behavior of γ_c^{bd} with respect to the change in peer fraction (r).

(b) Fraction of peers and superpeers required to be removed to breakdown the network and its impact upon percolation threshold f_c.

Figure 4.7: Case 1 of the degree dependent attack. The superpeer degree k_m is adjusted with the change of peer fraction r to keep the average degree fixed.

observed that the solution set of these networks upto a threshold peer fraction r_c, ($r_c = 0.78$ and 0.84 for $k_m = 30$ and $k_m = 40$ respectively) remains unbounded. The bounded solution set is observed for the networks with $r \geq r_c$ and the behavior of the boundary critical exponent γ_c^{bd} due to the change of peer fraction r is shown in Fig. 4.7(a). The fraction of peers and superpeers needed to be attacked for these networks is presented in Fig. 4.7(b). These networks exhibit the properties of case 1 of degree dependent attack, hence the removal of all the superpeers is necessary to disintegrate the network along with a fraction of peers. Fig. 4.7(b) also represents some instances of case 2 where only some fraction of superpeers are needed to be removed ($r < r_c$).

The main findings are listed below

a. Impact upon the fraction of peers removed

The increase in peer fraction slowly decreases γ_c^{bd} (Fig. 4.7(a)) which in turn gradually increases the fraction of peers removed $f_p^{\gamma_c^{bd}}$ (Fig. 4.7(b)). The amount of removal of peers also depends upon the superpeer degree k_m. The increase in the superpeer degree reduces the role of peers in determining the stability of the network. Hence fraction of peers required to be removed $f_p^{\gamma_c^{bd}}$ reduces with increase in k_m.

b. Impact upon percolation threshold

Let the percolation threshold for the networks having peer fraction r_1 and r_2 (where $r_1 < r_2$) be $f_{c1}^{\gamma_c^{bd}}$ and $f_{c2}^{\gamma_c^{bd}}$ respectively. Hence the percolation threshold for these two networks are

$$f_{c_1}^{\gamma_c^{bd}} = r_1 f_{p_1}^{\gamma_c^{bd}} + (1 - r_1) \tag{4.26}$$

$$f_{c_2}^{\gamma_c^{bd}} = r_2 f_{p_2}^{\gamma_c^{bd}} + (1 - r_2) \tag{4.27}$$

Therefore the change in the percolation threshold when the peer fraction changes from r_1 to r_2 is

$$\begin{aligned} f_{c_1}^{\gamma_c^{bd}} - f_{c_2}^{\gamma_c^{bd}} = \triangle f_c^{\gamma_c^{bd}} &= r_1 f_{p_1}^{\gamma_c^{bd}} - r_2 f_{p_2}^{\gamma_c^{bd}} - (r_1 - r_2) \\ &= \triangle \left(r f_p^{\gamma_c^{bd}} \right) - \triangle r \end{aligned} \tag{4.28}$$

The Eq. (4.28) shows that the change of percolation threshold $f_c^{\gamma_c^{bd}}$ is influenced by two opposite forces; on one hand the increase of peer fraction r (from r_1 to r_2) in the network makes $\triangle r < 0$ that increases $\triangle f_c^{\gamma_c^{bd}}$. On the other hand, this increase in r increases the fraction of peers required to be removed (Fig. 4.7(b)) which makes $\triangle \left(r f_p^{\gamma_c^{bd}} \right) < 0$. Depending upon the weightage of influence, $\triangle f_c^{\gamma_c^{bd}}$ (and subsequently $f_c^{\gamma_c^{bd}}$) either decreases or increases. For $r < r_c$, the $r f_p^{\gamma_c^{bd}}$ remains 0, hence $f_c^{\gamma_c^{bd}}$ decreases with r. When peer fraction $r \geq r_c$, due to the finite value of $f_p^{\gamma_c^{bd}}$, the $f_c^{\gamma_c^{bd}}$ increases.

Case 2 : Removal of only a fraction of superpeers

Some networks have unbounded solution set S_{γ_c} where $0 \leq \gamma_c \leq +\infty$. As $\gamma_c \to \infty$, $f_p^{\gamma_c}$ converges to 0 and $f_{sp}^{\gamma_c}$ converges to some x where $0 < x < 1$. This illustrates the case 2 of degree dependent attack where removal of only a fraction of superpeers is sufficient to disintegrate the network. The case study is performed with a network having superpeer degree $k_m = 25$, average degree $\langle k \rangle = 5$ and peer degree $k_l = 2$. The results are validated with the help of simulation. We plot the theoretically calculated (Eqs. (4.24), (4.25)) fraction of peers and superpeers required to be removed to breakdown the network for each critical exponent γ_c (Fig. 4.8). In simulation, we initially remove the fraction of superpeers $f_{sp}^{\gamma_c}$ which has been predicted theoretically and then start removing peers gradually to breakdown the network. The minimum peer fraction, removal of which causes the breakdown of the network corresponds to the simulated $f_p^{\gamma_c}$. We perform the simulation on graphs of 5000 nodes and repeat each experiment for 500 times and take the average of the removed peer fraction. We compare simulated results with theoretically calculated $f_p^{\gamma_c}$ (Fig. 4.8). The interesting findings are noted below.

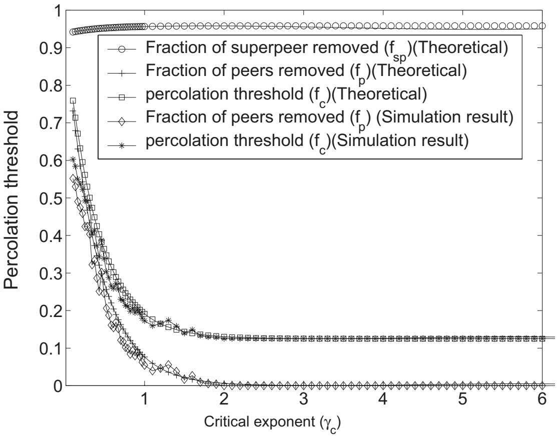


Figure 4.8: The above plot illustrates the case 2 of degree dependent attack.

a. The fraction of peers removed $f_p^{\gamma_c}$ gradually decreases with the increase of the critical exponent γ_c, which in turn decreases the value of $f_c^{\gamma_c}$. As $\gamma_c \to \infty$, the $f_p^{\gamma_c} \to 0$ with $f_{sp}^{\gamma_c} \to x$ (where $0 < x < 1$) and $f_{sp}^{\gamma_c}$, $f_c^{\gamma_c}$ both converges to some steady value. This signifies that the removal of only a fraction of superpeers is sufficient to breakdown the network (Fig. 4.8).

b. In Fig. 4.7(a), the nonexistence of the boundary critical exponent γ_c^{bd} for the networks having peer fraction $r < r_c$ signifies that the solution set of these networks is unbounded and the percolation process belongs to case 2. It can be observed that the fraction of peers required to be removed for these networks becomes zero (Fig. 4.7(b)) and removal of only a fraction of superpeers disintegrates the network.

c. It is important to note that removal of only a fraction of superpeers is sufficient to disintegrate any network with peer degree $k_l = 1$ and 2 irrespective of the superpeer degree and its fraction. Mathematically it can be explained as follows. For $k_l \leq 2$,

$2k_l \geq k_l^2$

$$\begin{aligned}
&\Rightarrow \quad 2rk_l \geq rk_l^2 \\
&\Rightarrow \quad (1-r)k_m + 2rk_l - rk_l^2 \geq 0 \\
&\Rightarrow \quad (1-r)k_m(k_m-1) \geq \langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle \\
&\Rightarrow \quad rk_l^{\gamma+1}(k_l-1) + (1-r)k_m^{\gamma+1}(k_m-1) \geq \\
&\qquad\; k_m^{\gamma}(\langle k\rangle(k_m+k_l) - k_m - 2\langle k\rangle)
\end{aligned}$$

This is exactly the inequality that we get in Eq. 4.23. This inequality is essentially the condition for breakdown of the superpeer network. Since the above inequality holds for any values of γ, it indicates that any network with $k_l = 1, 2$ has unbounded solution set.

Case 3 : Removal of some fraction of both peers and superpeers

Degree dependent attack allows to disintegrate the network by removing a fraction of both peers and superpeers. Intermediate critical exponents ($\gamma_c \in S_{\gamma_c}$ and $\gamma_c \neq \gamma_c^{bd}$) signify the fractional removal of both peers and superpeers. We calculate the amount of peers and superpeers needed to be removed to dissolve the network due to the change in γ_c. We deduce the results for a network having superpeer degree $k_m = 25$, average degree $\langle k\rangle = 5$ and peer degree $k_l = 3$. Results are also validated with the help of simulation (Fig. 4.9). The simulation set up is same as that described for case 2 of the degree dependent attack.

Observations:

a. Our analytical results show that this network has bounded solution set S_{γ_c} of the inequality (4.23) and all the critical exponents γ_c less than the boundary critical exponent $\gamma_c^{bd} = 1.171$ results in this kind of breakdown. It is evident from both theoretical and simulation results that the removal of any combination of $f_p^{\gamma_c}, f_{sp}^{\gamma_c}$ (obtained from the curves in Fig. 4.9) where $0 \leq \gamma_c < \gamma_c^{bd}$, results in the breakdown of the network.

b. Networks with unbounded solution set (Fig. 4.8) have finite values of γ_c ($\gamma_c < 2$) where the removal of both fraction of peers and superpeers are necessary to disintegrate the network.

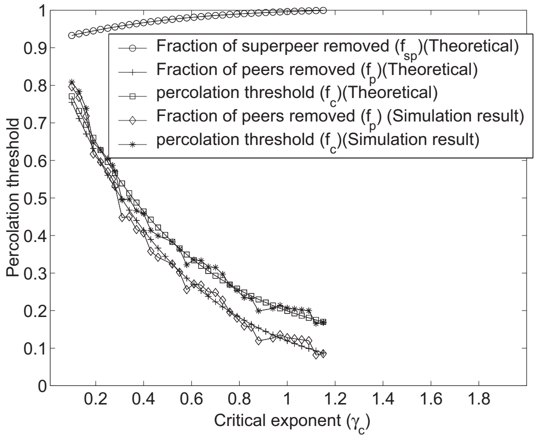


Figure 4.9: The above plot illustrates the case 3 of the degree dependent attack.

Summarization: In this section, the impact of degree dependent attack on the stability of the superpeer networks has been discussed in details. We have formulated the critical condition for network stability and subsequently obtained the critical exponent γ_c. This critical exponent γ_c and the normalizing constant C_{γ_c} determine the amount of peers and superpeers required to be removed to breakdown the network. Interestingly, we also find that the removal of only a fraction of superpeers is sufficient to disintegrate any network with peer degree $k_l = 1$ and 2 irrespective of the superpeer degree and its fraction [112].

One of the major contributions of this section is that, we have been able to provide a ***uniform attack framework*** (through degree dependent attack $f_k \sim k^\gamma$) which besides providing a flexibility in deciding attack strategy (through γ) also captures the essential features of deterministic attack. Case 1 and case 2 of the degree dependent attack resemble exactly the case 2 and case 1 of the deterministic attack respectively. In addition, $\gamma = 0$ and $\gamma < 0$ essentially model the degree independent and degree dependent failures respectively which have been illustrated in chapter 3.

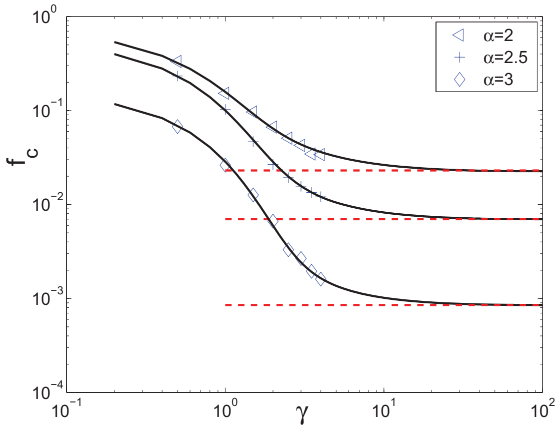


Figure 4.10: The above plot illustrates the change in percolation threshold f_c with the change of attack exponent γ. Three different scale free networks ($p_k \sim k^{-\alpha}$) with $\alpha = 2, 2.5$ and 3 have been considered. Curves represent the theoretical results whereas the symbols show the simulation results. The agreement between theoretical and simulation results (with $N = 10^5$) shows the success of Eq. (4.31). The dashed lines indicate the line of convergence of f_c calculated using Eq. (4.31) at $\gamma \to \infty$.

4.2.3 Physical interpretation of the attack exponent γ

The availability of the generalized attack model $f_k \sim k^\gamma$ immediately points to the importance of analyzing the attack parameter γ which signifies the information available to the attacker to breakdown the network [54]. As we know, the generalized attack can be represented as $f_k = \frac{k^\gamma}{C}$ where C is the normalizing constant. Clearly in the case of $\gamma > 0$, high degree nodes are removed with higher probability. Under this kind of generalized attack, the critical condition for stability of the large scale networks ($N \to \infty$) with degree distribution p_k can be expressed according to Eq. (4.9) as follows:

$$\langle k^2 \rangle - 2\langle k \rangle + \frac{[\langle k^{1+\gamma} \rangle - \langle k^{2+\gamma} \rangle]}{C} = 0\,, \tag{4.29}$$

where $\langle k^\omega \rangle$ is defined as $\langle k^\omega \rangle = \sum_k k^\omega\, p_k$. In consequence, the critical value of C that breaks down the network (termed as 'percolating C') simply reads:

$$C = \frac{\langle k^{2+\gamma} \rangle - \langle k^{1+\gamma} \rangle}{\langle k^2 \rangle - 2\langle k \rangle}. \tag{4.30}$$

The fraction of removed nodes f after an attack becomes $f = \sum_k p_k f_k$. Interestingly, for a given value of γ, the value of C obtained from Eq. (4.30) may not be feasible if $f_k = \frac{k^\gamma}{C} > 1$. This implies that an attack of the form $f_k = \frac{k^\gamma}{C}$ is unable to destroy the network. Given an attack characterized by an exponent γ, and using Eq. (4.30), the critical fraction of nodes that is required to remove in order to destroy the network is given by

$$f_c = \frac{\langle k^2 \rangle - 2\langle k \rangle}{\langle k^{2+\gamma} \rangle - \langle k^{1+\gamma} \rangle} \langle k^\gamma \rangle \,. \tag{4.31}$$

Eq. 4.31 is a generalized expression and can be applicable for any kind of network. However, the concept of topology information γ becomes more relevant for the network with continuous degree distribution, rather than the network consisting only two distinct degrees. Hence, next we perform a case study for the scale free networks where degree distribution follows $p_k \sim k^{-\alpha}$ with a maximum degree k_M. Fig 4.10 illustrates the behavior of the percolation threshold f_c of the scale free networks due to the change in the attack exponent γ. It also shows a comparison between Eq. (4.31) and stochastic simulations performed on the networks of size 10^5 with 500 realizations. In order to find the simulated value of percolating C as well as percolation threshold f_c, we have followed the method described in chapter 3. As expected, random failure ($\gamma = 0$) requires high attack intensity that increases percolation threshold. However as $\gamma \to \infty$,

$$\begin{aligned} f_c &\to (\langle k^2 \rangle - 2\langle k \rangle) \lim_{\gamma \to \infty} \frac{\langle k^\gamma \rangle}{\langle k^{2+\gamma} \rangle - \langle k^{1+\gamma} \rangle} \\ \Rightarrow f_c &\to h(\alpha) \frac{1}{k_M(k_M - 1)} \end{aligned} \tag{4.32}$$

where $h(\alpha)(= \langle k^2 \rangle - 2\langle k \rangle)$ is a constant function of power law exponent α and maximum degree of the network k_M. Hence as information about the network (γ) increases, f_c decreases and converges to some constant value. The analysis of this attack has revealed that in scale free networks an increase of γ leads to a decrease of the critical fraction of nodes that must be removed to disintegrate the network; i.e. a decrease in the percolation threshold f_c. However, after a threshold γ, the percolation threshold f_c reaches to some constant value and does not decrease further.

4.2.4 Impact of network size on the percolation threshold

Till now, our work has focused on analyzing the stability of large scale networks; this is in line with the general trend. Hence, the percolation threshold f_c remains independent of the network size N. However, the framework developed in this chapter provides us the flexibility to understand the stability of small scale networks also. In this section, we illustrate the effect of network size N upon the percolation threshold $f_c(N)$. In section 4.1.1, we compute the probability ϕ of finding an edge in the surviving subset S that is connected to a node of other subset R (Fig. 4.1) as

$$\phi = \frac{E}{\sum_{i=0}^{\infty} i\, n_i\,(1-f_i)} = \frac{\sum_{i=0}^{\infty} i\ p_i\, f_i}{(\sum_{k=0}^{\infty} k\, p_k) - 1/N}\,. \tag{4.33}$$

Following section 4.1.2, we find that the critical condition for the disintegration of the finite size networks can be expressed as

$$\left(\sum_k kp_k(1-f_k)\right)\left(\sum_k p_k k^2(1-f_k) + \sum_k kp_k(f_k-2)\right) + \frac{1}{N}\left(\sum_k kp_k(1-f_k)(2-k)\right) = 0 \tag{4.34}$$

Next we customize Eq. 4.34 for random failure by substituting $f_k = f$. Subsequently the percolation threshold for finite size network becomes

$$f_c(N) = \left(1 - \frac{1}{\frac{\langle k^2\rangle}{\langle k\rangle} - 1}\right) + \frac{1}{N}\left(\frac{2 - \langle k^2\rangle/\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}\right) \tag{4.35}$$

As network size $N \to \infty$, the expression of percolation threshold for random failure reduces to

$$f_c^{\infty} = 1 - \frac{1}{\frac{\langle k^2\rangle}{\langle k\rangle} - 1} \tag{4.36}$$

which converges to Eq. (3.17) of chapter 3.

Although Eq. (4.35) is a generalized expression, we show the results for Erdos-Renyi graph where the distinction between the finite and infinite size networks becomes nicely evident. We perform analysis on the E-R graph of finite size N with average degree $\langle k\rangle = 3$. Fig. 4.11 shows a comparative study between the percolation thresholds calculated from Eq. 4.35 (where we consider the network size N) and from

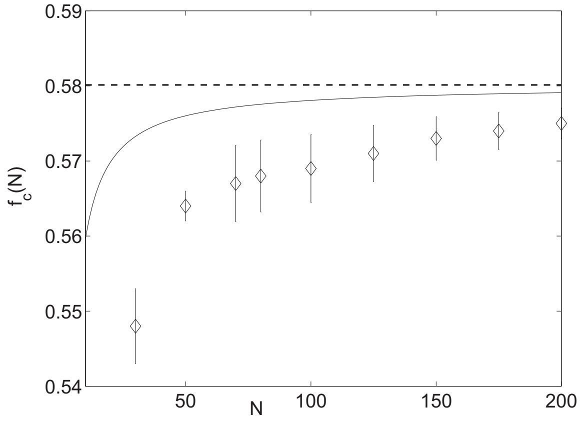


Figure 4.11: The figure illustrates the impact of network size N upon the percolation threshold f_c. The symbols represent the f_c obtained from stochastic simulation with a large number of realizations. The dashed line shows the percolation threshold calculated by Eq. (3.17) first proposed in [28] where f_c remains invariant with network size. The solid line shows the f_c calculated according to Eq. (4.35). The nature of the curve of Eq. (4.35) matches with the simulation however the results are not exact.

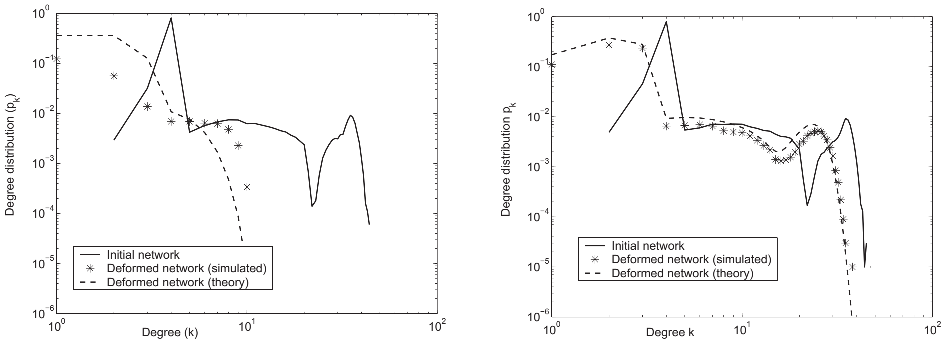
Eq. 4.36 (where f_c is invariant of network size) and results obtained from stochastic simulation. As Eq. 4.36 does not take the network size under consideration, f_c^{∞} takes a constant value for a specific network configuration. However, $f_c(N)$ calculated from Eq. 4.35 takes a lower value for small sized networks and gradually increases with increase in N. The observed deviation between $f_c(N)$ and simulation results can be arguably attributed to clustering effects, which have been ignored in the current approach.

4.3 Effect of attacks upon the commercial Gnutella Networks

In the previous sections, we have modeled the superpeer networks as various theoretical random graphs and validated our theoretically derived results through stochastic simulation. In this section, we choose the commercially popular peer-to-peer network, Gnutella as a case study and examine its stability in face of attacks. In section 4.1.2, we have shown that the measurement of network stability primarily depends upon the deformed degree distribution p'_k after attack. Hence, in this section we focus on the accurate calculation of p'_k for Gnutella networks. We perform a comparative study of the p'_k obtained from the experiments on Gnutella networks with the results calculated from the analytical framework.

4.3.1 Attacks on Gnutella networks

In chapter 3, we have described the generation of Gnutella networks following (a) bootstrapping protocol (b) topological snapshot. In this section, we refer the Gnutella network generated from bootstrapping protocol as 'Gnutella A' and Gnutella network generated from the topological snapshot as 'Gnutella B' and simulate deterministic attack and random failure on these two networks. We simulate the 'Gnutella A' net-

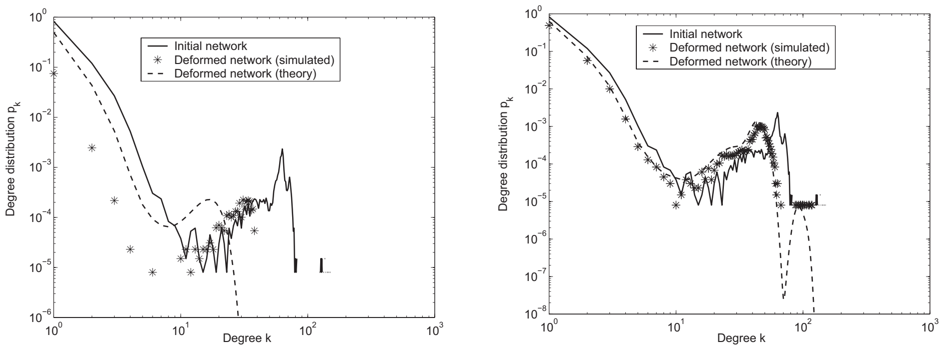


(a) The degree distribution of the deformed Gnutella network after deterministic attack. Here all the nodes in the network having degree greater that $k_{cut} = 10$ are removed.

(b) The degree distribution of the deformed Gnutella network after random failure. Here 20% of the nodes are randomly removed from the network.

Figure 4.12: The above plots show the topological impact of deterministic attack and random failure upon the simulated Gnutella A network of 5000 nodes. A comparative study of the simulation results with our theoretical model is performed.

work of $N = 5000$ nodes and all nodes in the network having degree more than 10 are removed in deterministic attack scenario. In random failure, 20% nodes in the network are removed randomly. The experiment is performed for 500 realizations and the average of the deformed degree distribution ($p'_{k_{sim}}$) and percolation threshold (f_{sim}) are calculated. We plot the degree distribution of the initial (p_k) and deformed network ($p'_{k_{sim}}$) in Fig. 4.12 and compare the simulation results with the theoretically calculated $p'_{k_{theory}}$ according to Eq. (4.4). Similarly, we mount a deterministic attack on 'Gnutella B' network where all the nodes in the network having degree more than 40 are removed. In random failure, 20% nodes in the network are removed randomly. The comparative study of the deformed degree distribution $p'_{k_{sim}}$ obtained from simulation with the theoretical model (Eq. (4.4)) has been done for these two kinds of node disturbances (Fig. 4.13). We observe that in both topologies (Gnutella A and B), the proposed theoretical model provides a reasonable approximation of the topological changes in the network under random failure (Fig. 4.12(b), Fig. 4.13(b)) however there is a deviation in case of deterministic attack (Fig. 4.12(a), Fig. 4.13(a)). We quantify the deviation of the theoretically predicted result from simulation in two different perspectives. First, we calculate the deviation in the individual $p_k \forall k$ (micro level deviation), second, the deviation in the average degree (macro level deviation). In order to quantify the deviation of individual $p_k, \forall k$ for Gnutella A network, we calculate the deviation parameter dev_A in the following manner. We compute $p'_{k_{sim}}$ and $p'_{k_{theory}}$ for individual degree k and subsequently derive their difference $diff_k = |p'_{k_{sim}} - p'_{k_{theory}}|$. The overall deviation (dev_A) is calculated from $\frac{\sum_k diff_k}{max(k)}$. Similarly we calculate the deviation parameter dev_B for the Gnutella B network. We find that the deviation parameter $dev_A = 0.0284$ in the Gnutella A network is higher than the Gnutella B network, $dev_B = 0.0219$. Next we show the deviation in the theoretically and experimentally calculated average degree of the Gnutella network after deterministic attack. In Gnutella A and B networks, the average degree of the initial network is 5.6191 and 2.4359 respectively. After attack, the new average degree obtained from simulation becomes $Avg\_deg^A_{sim} = 0.4858$ and $Avg\_deg^B_{sim} = 0.1608$ respectively for Gnutella A and B network. However the theoretically calculated average degree for these two networks show higher values than simulation ($Avg\_deg^A_{theory} = 1.5917$ and $Avg\_deg^B_{theory} = 0.6617$). We believe that the observed deviation between theoretical and simulation results are due to the presence of degree-degree correlation in the



(a) The degree distribution of the deformed Gnutella network after deterministic attack. Here all the nodes in the network having degree greater that $k_{cut} = 40$ are removed.

(b) The degree distribution of the deformed Gnutella network after random failure. Here 20% of the nodes are removed from the network randomly.

Figure 4.13: The above plots show the effect of attack and failure upon the Gnutella B network simulated from the topological snapshot taken during September 2004. The network is of the size of $1,31,869$ nodes. A comparative study of the simulation results with our theoretical model is performed.

network which was not present in the random graphs. We first formally define the degree-degree correlation and then examine its precise role.

Defining degree-degree correlation

Degree-degree correlation is defined as the probability of attachment of a source node to the target node given the present degree of the source/target node. Many networks show “assortative mixing” on their degrees, i.e., a preference for high-degree nodes to attach to other high-degree nodes in the network. Others show “dis-assortative mixing” where high degree nodes attach to low degree ones. In [123], this property has been conveniently measured by means of a single normalized index, the assortativity coefficient[2]. In our simulation, the Gnutella networks generated through the bootstrapping protocol (Gnutella A) as well as topological snapshot (Gnutella B) exhibit dis-assortativity (negative assortativity). The average assortativity of the Gnutella A for 500 realizations becomes $\alpha = -0.6749$ whereas the Gnutella B has $\alpha = -0.6318$. The deviation of the theoretical results from simulation for Gnutella A ($dev_A = 0.0284$) is more than the Gnutella B network ($dev_B = 0.0219$) as well as Gnutella A has lower assortativity than Gnutella B. This indicates some sort of relationship between the deviation and assortativity. The precise role of assortativity is investigated next.

Role of assortativity

In this section, we intuitively explain the deviation between the theoretical and simulation results in assortative network. First we explain the impact of assortativity on the average degree of the network.

Impact of assortativity on the average degree

A given attack on some assortative network changes the average degree (density) of the network, and the amount of change depends upon the assortative nature of the network. In Fig. 4.1, we find that two types of edges originate from the nodes of the removed set R; (a) one set of edges whose other end is also connected to the nodes of set R (say E_R) (b) another set of edges whose other end is connected to the nodes of set S (say E). For any given attack f_k^{atk}, the number of nodes in set R will be same for all networks. Let us assume that due to attack f_k^{atk} on a given network, the number of tips removed only from the nodes of removed set R is $\widehat{R_{tips}}$ and E is the number of tips removed from the set S. The number of edge tips removed will be the summation of $\widehat{R_{tips}}$ and E. Hence, the total number of edges removed from the network after attack becomes $\frac{E+\widehat{R_{tips}}}{2}$. $\widehat{R_{tips}}$ will be a constant across all networks

[2] Degree-degree correlation of a network is formally defined through assortativity coefficient α [123] such that

$$\alpha = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum \frac{1}{2}(j_i + k_i)]^2}$$

where j_i, k_i are the degrees of the vertices at the ends of the i^{th} edge, with $i = 1...M$ (M is the total number of edges in the network).

(it is directly dependent on the number of nodes removed); therefore the number of edges removed will be directly dependent upon the value of E. Subsequently, the number of edges survived in the network after the attack f_k^{atk} may be expressed as

$$E_{new} = E_{tot} - \frac{E + \widehat{R_{tips}}}{2} \tag{4.37}$$

The value of E (number of edges running between the set S and R) depends on the assortativity of the network. In case of deterministic attack in assortative network, most of the high degree nodes (in R) are connected among themselves (making E_R quite high), hence a very small number of edges E are connected to the set S. Using Eq. 4.37, we find that the removal of few E edges keeps the network quite dense with high average degree. However, in disassortative network, most of the edges E run between S (low degree nodes) and R (high degree nodes) and there exits few links E_R connecting the high degree nodes of set R. Subsequently, the removal of large number of E edges reduces the average degree. Hence $E_{new}(assort) > E_{new}(uncorr) > E_{new}(disassort)$.

Intuitive justification behind $Avg\_deg_{theory} > Avg\_deg_{sim}$ against attack

We simulate an attack on Gnutella networks (a disassortative network) such that most of the high degree nodes are removed. As explained, removal of high degree nodes removes the large number of edges running between set S and R , say E_{sim} (E obtained from simulation). On the other hand, in theoretically calculated E (according to the Eq. (4.1)), say E_{theory}, we assume that the network is uncorrelated in nature, hence there is an equal/uniform probability that the other end of the removed tip (in set R) is connected to the nodes in the set S and set R. Hence the total number of edges running between the set S and set R, calculated theoretically (E_{theory}) is less than E_{sim}. This difference in the estimation of E (E_{theory} and E_{sim}) affects the number of survived edges E_{new} (Eq. 4.37) in the survived network. More specifically, in the theoretical calculation, the amount of reduction of the average degree of the survived network after attack is underestimated than that of the simulation. Hence after the given attack, the simulated network ($p'_{k_{sim}}$) becomes more sparse than the theoretically calculated network ($p'_{k_{theory}}$). Subsequently, $Avg\_deg_{theory} > Avg\_deg_{sim}$. This directly answers the question why for Gnutella network, $Avg\_deg_{theory} > Avg\_deg_{sim}$ where *theory* signifies the uncorrelated network and *sim* signifies disassortative network.

Assortativity does not have any impact on random failure

However it is interesting to observe in Fig. 4.12(b) and Fig. 4.13(b) that although assortativity takes a major role in attack, it does not have any influence in random failure. In random failure, the nodes in the set S and R are placed independent of their degree, hence high and low degree nodes are uniformly distributed in those sets. Subsequently, there is an equal/uniform probability that the other end of the edge connected to a node of the removed set R is linked with either a node of set S or of set R. In this way, the effect of assortativity becomes nullified in face of random failure.

In the next section, we utilize this intuitive understanding to refine and rectify our analytical framework so that it becomes applicable to the correlated networks also.

4.4 Stability analysis for degree correlated networks

In the previous section, we find that our theoretical framework is not able to explain the exact behavior of Gnutella network in face of deterministic attack. However, we have presented an intuitive explanation for the deviation of the theoretically computed results from the simulation. In this section we refine our framework, developed in section 4.1 to include correlated networks and examine its applicability on Gnutella network.

4.4.1 Deformed topology after attack

In this section, we modify the expression (derived in section 4.1.1) of deformed degree distribution p'_k to make it suitable for degree correlated networks. The degree-degree correlation information of a network with maximum degree k_M is represented by the correlation matrix M as follows

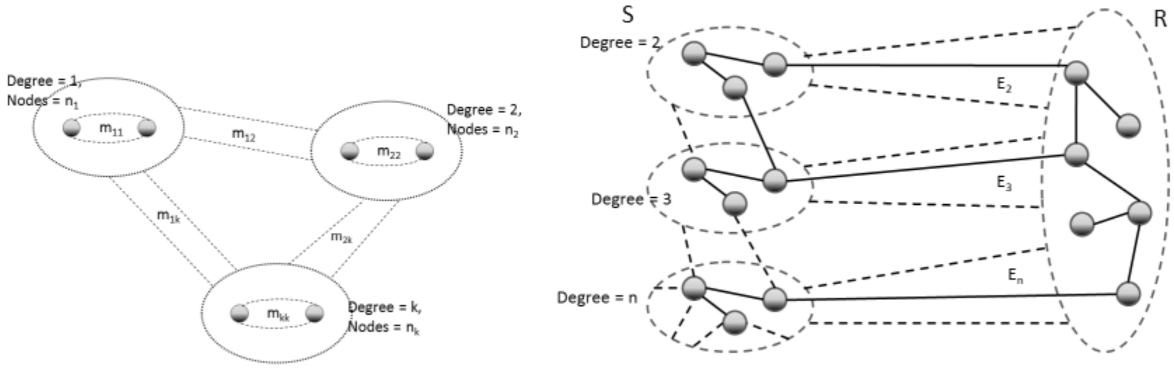
$$M = \begin{pmatrix} m_{11} & m_{12} & m_{13} & ... & m_{1k_M} \\ m_{21} & m_{22} & m_{23} & ... & m_{2k_M} \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ m_{k_M 1} & m_{k_M 2} & m_{k_M 3} & ... & m_{k_M k_M} \end{pmatrix}$$

In this correlation matrix M, each element m_{jk} represents the fraction of total edges that exist between nodes of degree j and nodes of degree k (Fig. 4.14(a)). We frame the attack on the network in the same manner as explained in the section 4.1.1. The attack on the network divides the graph into two sets of nodes: one set containing the surviving nodes S and another set containing the nodes to be removed R as shown in the Fig. (4.14(b)).

E_j **instead of** E

In section 4.1.1, we have calculated E which represents the number of edges running between set S and R. It is also the number of tips that is going to be removed from the nodes of the set S. The expression of E in Eq.(4.1) gives correct approximation for an uncorrelated network as the edge connectivity between a node of set R and any node of set S is equally probable. But in case of a degree correlated network, the probability of an edge between a node of degree i and a node of degree j is given by m_{ij} element of the correlation matrix M. Hence instead of calculating E we calculate E_j which indicates the number of edges connected between nodes of degree j in the set S and the nodes of any degree in the set R (Fig. 4.14(b)). Hence the total number of edges connected between the set S and R, that are going to be removed is given by $E = \sum_{j=0}^{k_M} E_j$. The expression for E_j can be formulated in the following way.

The total number of edge tips connected to the k degree nodes in set R can be expressed as kn_kf_k. Therefore, the number of edge tips connected to the j degree nodes of the network whose other end is connected to the k degree node of set R becomes $m'_{jk}kn_kf_k$. The fraction m'_{jk} represents the fraction of edges connecting j degree nodes and k degree nodes over all the edges in the network with at least one end connected to the k degree nodes. The value of m'_{jk} can be computed from the



(a) The degree correlation in the network represented by the elements of the assortativity matrix M

(b) The dissection of a correlated network into two sets S and R due to the attack on the network.

Figure 4.14: Degree correlation present in the network and its implication on attack.

edge correlation matrix M as

$$m'_{jk} = \frac{m_{jk}}{\sum_{j=0}^{\infty} m_{jk}} = \frac{m_{jk}}{kp_k} \sum_i ip_i \tag{4.38}$$

where $\sum_{j=0}^{\infty} m_{jk}$ denotes the fraction of edge tips connected to k degree nodes in the network and may be expressed as

$$\sum_{j=0}^{\infty} m_{jk} = \frac{kp_k}{\sum_i ip_i} \tag{4.39}$$

Similar to section 4.2, we can say that the number of edge tips connected to the j degree nodes of set S whose other end is connected to the k degree node of set R becomes $m'_{jk} k n_k f_k (1 - f_j)$. This helps us to derive the total number of edges whose one end is connected to a j degree node in set S and the other end is connected to any node in the set R, which can be expressed as

$$E_j = \sum_{k=0}^{\infty} m'_{jk}\, k\, n_k\, f_k\, (1 - f_j) \tag{4.40}$$

Due to the presence of degree correlation, the probability that a surviving node of set S loses one link due to the removal of $E (= \sum_{i=0}^{k_M} E_i)$ edges is not constant (as ϕ

in Eq. 4.33). Moreover, the probability that a survived node loses one link depends upon the degree (j) of the survived node. Hence, the probability ϕ_j of finding an edge running between a j degree node in the surviving set S and any node of the other set R can be expressed as

$$\phi_j = \frac{E_j}{jn_j(1-f_j)} \tag{4.41}$$

Here ϕ_j signifies the probability that a j degree node loses one link due the removal of E edges.

Finally, using the concept of Eq. (4.4) and from the Eqs. (4.41) and (4.3), the expression of the deformed degree distribution p'_k can be expressed in binomial distribution form

$$p'_k = \sum_{q=k}^{\infty} \begin{pmatrix} q \\ k \end{pmatrix} \phi_q^{q-k} (1-\phi_q)^k \, p_q^s \, . \tag{4.42}$$

where the probability p_q^s of finding a node with degree q in the surviving subset S (before removal of the E edges) is given by Eq. (4.3) of section 4.1.1.

Random failure as a special case

In case of random failure attack the probability of attack on every node is same i.e. $f_j = f_k = f\,(constant)$. Therefore we can express E_j, which is the total number of edges whose one end is connected to a j degree node in set S and the other end is connected to any node in the set R, as the following:

$$E_j = f(1-f)\sum_{k=0}^{\infty} m'_{jk}\, k\, n_k \tag{4.43}$$

Using Eq. (4.38), Eq. (4.43) and (4.39) the expression for E_j reduces to

$$E_j = f(1-f)N\sum_{i=0}^{\infty} ip_i \sum_{k=0}^{\infty} m_{jk} = f(1-f)Njp_j \tag{4.44}$$

We substitute the expression for E_j obtained from Eq. (4.44) in Eq. (4.41) and find

$$\phi_j = \frac{f(1-f)Njp_j}{jn_j(1-f)} = f \tag{4.45}$$

Hence in case of random failure

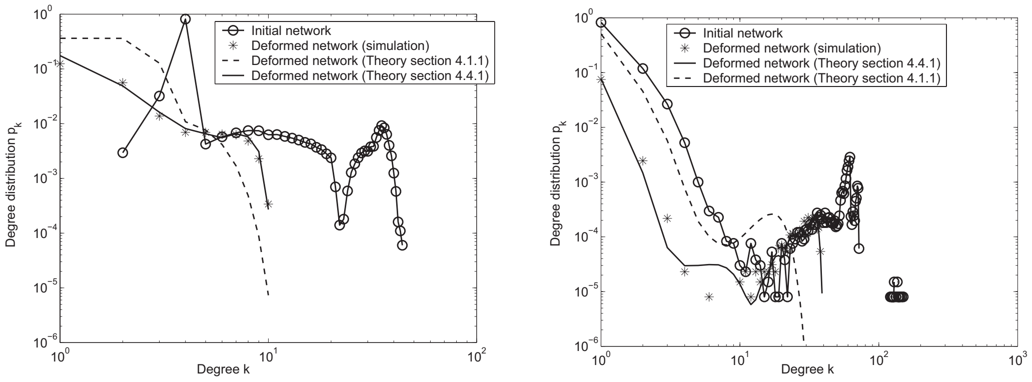
$$\phi = \phi_j = f(constant)\ independent\ of\ any\ degree\ j. \tag{4.46}$$

Substituting the value of $\phi = f$ in Eq. (4.4) and the value $\phi_q = f$ in Eq. (4.42) we find that

$$p'_k(Uncorrelated) = p'_k(Correlated) \tag{4.47}$$

$$= \sum_{q=k}^{\infty} \begin{pmatrix} q \\ k \end{pmatrix} f^{q-k}(1-f)^k\, p_q^s \tag{4.48}$$

The above expression is independent of any correlation parameter. This shows that degree-degree correlation has no role to play in case of random failure. This conclusion confirms the results shown in Figs. 4.12(b) and 4.13(b) where we observe a good agreement of p'_k obtained from the theory and simulation for Gnutella network. However, this does not hold for attacks in correlated networks. Next, we show that our refinement gives better agreement with the simulation results for the attacks on correlated Gnutella networks.



(a) Gnutella A network, correlation coefficient $\alpha = -0.6749$

(b) Gnutella B network, correlation coefficient $\alpha = -0.6318$

Figure 4.15: The impact of deterministic attack upon the degree distribution p_k of the Gnutella network. The figures show that Eq. 4.42 gives far better approximation of the deformed degree distribution than Eq. 4.4

Simulation results on Gnutella Network

We validate the theory developed for correlated network by simulating deterministic attack on 'Gnutella A' and 'Gnutella B' networks. Similar to section 4.3.1, we simulate the deterministic attack on the Gnutella networks. In 'Gnutella A' and 'Gnutella B' network, we simulate deterministic attacks by removing all the nodes with degree greater than 10 and 40 respectively. Fig. (4.15) shows the impact of the deterministic attack on the degree distribution of Gnutella network. It can be observed that the deformed degree distribution obtained from Eq. 4.42 for the Gnutella network is in good agreement with simulation results. We find that the average degree of the 'Gnutella A' and 'Gnutella B' networks obtained from simulation ($Avg\_deg^{A}_{sim} = 0.4858$ and $Avg\_deg^{B}_{sim} = 0.1608$) are quite close to the theoretically calculated values using Eq. 4.42 ($Avg\_deg^{A}_{theory} = 0.4739$ and $Avg\_deg^{B}_{theory} = 0.1514$).

4.5 Conclusion

In this chapter, we have developed a more sophisticated framework for stability analysis of superpeer networks against attacks. We have shown that this framework enables us to calculate the degree distribution of the deformed network p'_k after removal of nodes. In addition, the framework enables us to measure stability of small scale network as well as networks exhibiting strong degree-degree correlated mixing. As an application of the framework, we have analyzed the effects of two kinds of attacks namely deterministic attack and degree dependent attack and validated the results through simulation. We have shown that in deterministic attack, the increase in peer degree may be detrimental in some cases. Our framework has also revealed that the degree dependent attack provides us a more generalized attack strategy where various situations can be generated only by changing the attack parameter γ. This attack parameter γ also signifies the amount of topological information available to the attacker to breakdown the network. We have observed that increase in γ makes the attack efficient by reducing the percolation threshold. However, beyond a threshold limit, this information does not help the attackers in a significant manner. We have presented a comparative study of our theoretical analysis with real world Gnutella network. The results have shown that degree degree correlation present in Gnutella exhibits a disparity in p'_k in case of attack however the disparity is not seen in case of random failure. We have suitably modified our framework to include the degree-degree correlation factor in consideration. It is important to note that, the stability condition stated in Eq. (4.5) [128] is not applicable for degree-degree correlated network [128]. Hence, in this work we do not derive the percolation threshold of degree correlated network; rather we focus on the accurate calculation of p'_k through a generalized framework. Since degree distribution p'_k is the main ingredient for the stability condition of correlated networks [66], we claim that our work makes a significant contribution towards the understanding stability of generalized network.

In chapters 3 and 4, we have analyzed the stability of some 'existing' superpeer networks against peer churn and attacks. However, superpeer networks are generally growing networks that continuously evolve with the addition of new peers as well as realignment of peers. Hence, the formation or emergence of superpeer network due to various node and link dynamics is another interesting research problem. The next two chapters focus on the various issues related to the emergence of superpeer networks due to joining and leaving of nodes, rewiring of links etc.