Resilience of Networks

Vishwa Deepak 04CS1029 {vishd_iitkgp@yahoo.com}

May 5, 2008

1 Introduction

The ability for communication to take place in a network is dependant on the links between members or nodes that make up the network. In complex network theory, it is a matter of concern to know how data travels through the network or how nodes are related. So, Loss of edges or nodes in a network may affect the relation among nodes or in other words flow of data. Here comes the term ***Resilience***.

A network is generally labeled as being ***resilient***,if it can sustain the loss of many edges or nodes without disrupting pathways between remaining nodes. ***Resilience*** is one of many metrices considered when studying and analyzing complex networks. The susceptibility and survivability of a network to faults and attacks can be estimated by looking at resilience statistics. **Resilience is also referred to as the robustness of a network.**

It is found that most networks in the real world are susceptible to failures or attacks. So, being able to effectively estimate how a specific network will respond to these scenarios can allow us to generate response scenarios to negate network failures.

To measure the ***resilience of a network***, ***Percolation theory*** plays a major role. So,In this lecture, first of all we will discuss Percolation theory . After that , we will get into failures and attacks and stability of *Erdos-Renyi network* and *Scale free network* in those scenarios.

We know that ***Giant component*** *is a term referring to a connected subgraph that contains a majority of the entire graph's nodes and* ***Percolation theory*** *is based on adding nodes and connections to an empty graph until a giant component surfaces*,so, there will be drastic effect on the size of giant component,if some random failures happen. Therefore, we will be discussing those issues also in next sections.

Random failures affect the stability of finite and infinite networks in different fashions.So, we need to derive some formulations for those two scenarios.As

well as, Degree distribution of the *surviving network* also gets affected in case of node or edge removal.

It is evident that in recent years, the topology of *scale-free networks* ,one in which the degree distribution follows a *power law* ,has emerged as a useful model for studying real-world networks. This practice has brought about the study of resilience in real-world networks such as *internet* and *World wide web*.So, we will use scale-free networks to understand various results.

2 Percolation Theory

2.1 Definition

Percolation theory began in 1957, when ***Broadbent and Hammersley*** introduced the bond percolation model. They did this to model a porous stone on a microscopic level in order to study the question of whether the center of the stone gets wetted when it is immersed into a bucket of water. Since then, percolation theory has attracted an enormous amount of interest. For example,In earlier time, this theory was used to find the spread of an epidemic. So, In real world, this theory has been vastly used.

In mathematics, ***Percolation theory*** describes the behavior of connected clusters in a random graph. To understand this theory, assume that we have a three dimensional network of $n \times n \times n$ points or vertices.The connections or edges between each two neighbors may be open with probability p (allowing the liquid through) or closed with probability $1 - p$ and we assume that they are independent. This theory asks, for a given p, what is the probability that an open path exists from one vertex to the other vertex at different corners.Thus,percolation theory deals with methods of randomly marking nodes or edges as being opened or closed.

If n is very large, then this question reduces to a well known question, **does an infinite cluster exists in the network?** In this case,Using ***Kolmogorov's zero-one law***,we can say that, for any given p, the probability that an infinite cluster exists is either zero or one. So, there must exist a critical p, denoted by p_c, below which the probability is always 0 and above which the probability is always 1.This means that the behavior of network is quite different for $p < p_c$ and for $p > p_c$. Such a sharp transition in global behavior of a system at some parameter value is called a ***phase transition or a critical phenomenon.***

Even if, n is not so large, i.e. the graph is finite, then also this theory or model makes perfect sense.The only requirement is that the graph should be

connected.

This theory talks of a number of models. The most popular modes are (a)Site percolation model and (b)Bond percolation mode. These models will be discussed in the next subsection. Since, in case of infinite graph, the problem reduces to find **infinite cluster**,one may ask **how many infinite clusters may exist in a graph?**.So, We will discuss the importance and some important facts about these in upcoming subsection.

2.2 Site and Bond Percolation models

Let's assume that G=(V,E) denotes the underlying infinite graph on which the percolation process takes place. G is always assumed connected and for the sake of uniqueness of infinite cluster we are assuming infinite graph here. So, **bond percolation** with retention parameter p on the graph G is a random element X, whose distribution is product measure with marginals (1-p,p). We identify X with the subgraph of G containing all vertices $v \in V$ and precisely those edges $e \in E$ taking value $X(e) = 1$. When $X(e) = 1$, we speak of e as an ***open edge***,whereas if $X(e) = 0$ we say that e is ***closed edge.***

One can also consider the model in which vertices are independently open or closed, but all the edges are assumed open. This model or process is as natural as bond percolation model. This version of percolation is called **site percolation.**So, In formal terms,**Site percolation** with retention parameter p on the graph G is a random element X, whose distribution is product measure with marginals (1-p,p). We identify X with the subgraph of G containing all edges $e \in E$ and precisely those vertices $v \in V$ taking value $X(v) = 1$. When $X(v) = 1$, we speak of v as an ***open vertex***,whereas if $X(v) = 0$ we say that v is ***closed vertex.***.

As far as qualitative results are concerned,it is usually of little importance whether bond or site percolation is considered.Here, we will focus on bond percolation.Most results and proofs have obvious analogues for site percolation.

2.3 Infinite Clusters and Giant Component

Since, In case of infinite graphs ,percolation model results in finding infinite clusters. So, we need to know the total number of possible infinite clusters in a graph.

Let us take a motivational example from real world to find infinite clusters. Take $G = (V, E)$ be the graph whose vertex set consists of all mathematicians and connect any two of them by an edge $e \in E$ if they have ever coauthored a mathematical problem. The distance in this graph between a given mathematician v and Paul Erdõs is colloquially referred to as $v's$ ***Erdõs number***.If we define a mathematician to be a person who has authored at least one piece of work that by May 2004 had found its way into *MathSciNet database* and two such mathematicians to be coauthors if they have a joint publication in that database, then the graph consists of about 401,000 vertices. About 83,000 of the nodes are isolated, corresponding to mathematicians with no coauthors. The largest connected component - the one containing Erdõs - contains about 268,000;these are all the mathematicians that have finite Erdõs numbers. The second largest connected component contains only 32 vertices. This striking contrast between the size of the largest and the second largest connected component may be phrased as ***uniqueness of the giant component.***This means that only one infinite cluster exists in a graph.Now we come to the definition of giant component.

Giant Component is the large connected component of a random graph,whose size is proportional to the size of the whole graph.So, it increases linearly as the size of the graph increases. That's why giant component is assumed to be an infinite cluster.

3 Failure and Attacks

We know that random deletion of nodes and edges may result in disconnecting the network. So, we need to analyze different measures to find out the resilience or robustness of network. The simplest indicator of resilience in a network is the variation (or lack of variation)in the fraction of vertices in the largest component of the network, which we equate with giant components in our models.

Deletion of a node may be due to random failure of that particular node or caused by some attacks done intentionally. Effect of random failures and intentional attacks in various kind of graphs has been discussed by a number of people.

Cohen *et al.* has observed from the results that Internet,which can be modeled by power law network is more resilient to random failure than E-R graphs.

Albert found that scale free networks are highly sensitive to intentional attacks.This result was backed by **Cohen *et al.***

Newman *et al.* developed the theory of random graphs with arbitrary degree distribution with the help of generating function formalism.Using this formalism, **Callaway** found the exact analytic solutions for percolation on random graphs with any degree distribution where failure has been modeled by an arbitrary function of node degree.

Tanizawa *et al.* have discussed a situation when a network is subjected to simultaneous targeted and random attacks. They have modeled this situation as a sequence of *"waves"*of targeted and random attacks which removes fractions p_t and p_r of the nodes of the network.

Thus all the researchers have considered some particular types of networks such as E-R, scale free or bimodal networks and analyzed the effect of a few specific kinds of failures like random,intentional or mixed upon them. From their works, it have been found that in a random graph, node failures can be divided into two models.

3.1 Models of node failures and their representation

In random graphs, some node failures depend upon the degree distribution of network whereas others are independent of degree distribution. This criterion is used to model two kinds of random failures in a random graph.

The most common type of failures are denoted as **degree independent failure.**In this model, the probability of removal of any randomly chosen node is constant, degree independent and equal for all other nodes in the graph. Therefore, the presence of any randomly chosen node having degree k after this kind of failure is $q_k = q$ independent of k .

In some networks,nodes having higher connectivity are more robust in the network than the nodes having lower connectivity because those loosely connected nodes enter and leave the network quite frequently.These observations lead to a new kind of failure where probability of removal of a node is inversely proportional to the degree of that node. We denote this kind of failure as **degree dependent failure.** Here , probability of failure of node (f_k) having degree k is inversely proportional to k^γ i.e. $f_k = \alpha/k^\gamma$ where $0 \leq \alpha \leq 1$ and γ is a real number. Therefore, probability of the presence of a node having degree k after this kind of failure is $q_k = 1 - f_k = 1 - \alpha/k^\gamma$.

4 Stability and robustness of infinite network

The stability and robustness of networks are mainly measured in terms of certain fraction of nodes f_c called **percolation threshold or critical fraction**,removal of which disintegrates the network into smaller,disconnected components. Below that threshold, their exists a connected component which spans the entire network also termed as **giant component**.The valued of f_c signifies the stability of the network,higher the value, higher is the stability of network against failure or attack.

To measure the stability of network, **Callaway *et al.***expanded the research on **stability or resilience metrics** using percolation theory by studying graphs with general degree distributions,which was formalised as generating function by **Newman *et al.***

Assume that p_k and q_k specifies the network topology and failure model respectively. We need to find out the transition point where the giant component breaks down into smaller components. Here p_k shows the probability that a node chosen at random has degree k and q_k shows the probability that a node with degree k is open or occupied.Thus, $p_k q_k$ specifies the probability of a node having degree k to be present in the network after the process of removal of some portion of nodes is completed. Let, the generating function for this distribution be $F_0(x)$.So,

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

Distribution of the outgoing edges of the first neighbor of a randomly chosen node can be generated by

$$F_1(x) = \frac{\sum_{k=0}^{+\infty} k p_k q_k x^{k-1}}{\sum_{k=0}^{+\infty} k p_k} = F_0'(x)/z$$

Here z is the average degree of the network.

Now,before moving further,we present **sum rule for the connected component of vertices reached by following a randomly chosen edge.** This rule states that, *"Probability of each such component can be represented as the sum of the probabilities of having no vertex, only a single vertex , having a single vertex connected to one other component, or two other components and so forth. "*

Let's assume that $H_1(x)$ be the generating function for the distribution of the component sizes that are reached by choosing a random edge and following it to one of its ends. The component may contain zero node if the node at the

other end of the randomly selected edge is removed, which happens with probability 1- $F_1(1)$, or the edge may lead to a node with k other edges leading out of it other than the edge we came along,distributed according to $F_1(x)$. That means that $H_1(x)$ satisfies a self-consistency condition,which may be shown as

$$H_1(x) = 1 - F_1(1) + xF_1(H_1(x))$$

Similarly, if we select a random node then the distribution for the component size to which that node belongs is generated by $H_0(x)$. It may be shown that

$$H_0(x) = 1 - F_0(1) + xF_0(H_1(x))$$

We know that average size of any component having generating functin $H_0(x)$ is $H_0'(1)$. Here,

$$H_0'(x) = xF_0(H_1(x)) + F_0'(H_1(x)).H_1'(x)$$

$$H_1'(x) = xF_1(H_1(x)) + F_1'(H_1(x)).H_1'(x)$$

$$\implies H_1'(x(1 - F_1'(H_1(x))) = xF_1(H_1(x))$$

$$\implies H_0'(1) = F_0(1) + \frac{F_0'(1)F_1(1)}{1 - F_1'(1)}$$

So, Average size of components diverges when $1 - F_1'(1) = 0$. This means that, at that condition, size of the component becomes infinite. This implies that the entire network joins together forming one giant component.

$$\implies F_1'(1) = 1$$

$$F_1'(x) = \frac{\sum_{k=0}^{+\infty} k(k-1)p_k q_k x^{k-2}}{\sum_{k=0}^{+\infty} kp_k}$$

$$\implies 1 - \frac{\sum_{k=0}^{+\infty} k(k-1)p_k q_k}{\sum_{k=0}^{+\infty} kp_k} = 0$$

$$\implies \sum_{k=0}^{+\infty} k(k-1)p_k q_k = \sum_{k=0}^{+\infty} kp_k$$

$$\implies \sum_{k=0}^{+\infty} kp_k(kq_k - q_k - 1) = 0$$

The above equality shows the critical condition for the stability of giant component with respect to any type of graphs, be it E-R, scale-free or bimodal (characterized by p_k) undergoing any type of failure be it degree dependent or degree independent (characterized by q_k).

In the upcoming subsections, we will investigate the stability situation under various special conditions.

4.1 Stability of network at degree independent random failure

We know that mainly there are two kinds of random failures of nodes– degree independent random failure and degree independent random failure.In this subsection we will deal with degree independent random failure. We will derive the stability condition and give an example of scale free network.

In generalized random graph, assume that q_k is constant for all the nodes. So,$q_k = q_c$ is the critical fraction of nodes whose presence in the graph is essential for the stability of the giant component after this kind of failure.

Now, the stability condition is

$$\sum_{k=0}^{+\infty} kp_k(kq_c - q_c - 1) = 0$$

$$\implies q_c[\sum_{k=0}^{+\infty}(k^2p_k - kp_k)] = \sum_{k=0}^{+\infty} kp_k$$

$$\implies q_c = \frac{\sum_{k=0}^{+\infty} kp_k}{\sum_{k=0}^{+\infty}(k^2p_k - kp_k)}$$

$$\implies q_c = \frac{< k >}{< k^2 > - < k >}$$

$$\implies q_c = \frac{1}{\frac{<k^2>}{<k>} - 1}$$

Now, If f_c is the critical fraction of nodes whose random removal disintegrates the giant component, then percolation threshold $f_c = 1 - q_c$. Therefore, percolation threshold will be

$$f_c = 1 - \frac{1}{\frac{<k^2>}{<k>} - 1}$$

This is the condition for the disappearance of the giant component due to random failure. **Now, we derive the value of f_c for scale- free network.**

A scale-free network is one in which certain nodes have high degree distribution and act as hubs in the network. The distribution follows a power law

$$P(k) = ck^{-\gamma}, k = m, m+1, ...K$$

where m is the smallest degree value and $K \approx mN^{1/\gamma-1}$ represents the largest degree value. For random failures,we use a continuum approximation in the limit $K \gg m \gg 1$ to obtain

$$\frac{< k^2 >}{< k >} = \frac{2-\gamma}{3-\gamma} \times \begin{cases} m & \text{if } \gamma > 3; \\ m^{\gamma-2}k^{3-\gamma} & \text{if } 2 < \gamma < 3; \\ K & \text{if } 1 < \gamma < 2. \end{cases}$$

For $\gamma > 3$ there is a transition (formation of a giant component or disappearance of a giant component) at

$$f_c = 1 - \frac{1}{\frac{\gamma-2}{\gamma-3}m - 1}$$

4.2 Stability of network at degree dependent random failure

In some networks such as p2p networks, the nodes having higher connectivity are much more stable and reliable than the nodes having lower connectivity. Therefore the probability of the presence of a node having degree k after this kind of failure is

$$q_k = 1 - \frac{\alpha}{k^\gamma}$$

This means that

$$f_k = \frac{\alpha}{k^\gamma}$$

Now $\gamma > 0$ shows that,there is degree dependent failure, whereas $\gamma < 0$ shows that, there is an attack.

We derived in the earlier subsection

$$\sum_{k=0}^{+\infty} kp_k(kq_k - q_k - 1) = 0$$

$$\implies \sum_{k=0}^{+\infty} kp_k[(k-1)(1-\frac{\alpha}{k^\gamma}) - 1] = 0$$

$$\implies \sum_{k=0}^{+\infty} kp_k(k-1) - \alpha \sum_{k=0}^{+\infty} \frac{p_k(k-1)}{k^{\gamma-1}} = \sum_{k=0}^{+\infty} kp_k$$

$$\implies \sum_{k=0}^{+\infty} kp_k(k-2) = \alpha \sum_{k=0}^{+\infty} \frac{p_k(k-1)}{k^{\gamma-1}}$$

$$\implies \alpha = \frac{\sum_{k=0}^{+\infty} kp_k(k-2)}{\sum_{k=0}^{+\infty} \frac{p_k(k-1)}{k^{\gamma-1}}}$$

$$\implies < k^2 > -\alpha < k^{2-\gamma} > +\alpha < k^{1-\gamma} > -2 < k >= 0$$

$$\implies \alpha = \frac{< k^2 > -2 < k >}{< k > -1} for \gamma = 1$$

Here the percolation threshold is

$$f_c = \sum_{k=0}^{+\infty} \frac{\alpha}{k^\gamma} p_k$$

and for $\alpha = 1$

$$f_c = \sum_{k=0}^{+\infty} \frac{1}{k^\gamma} p_k$$

Thus we can determine the variation of percolation threshold f_c for various networks due to degree dependent random failure.

5 Stability and robustness of finite network

While studying the stability and robustness of finite network, we develop an alternative derivation for the percolation network. Instead of applying a generating function formalism to find an analytic expression for percolation threshold, we use the fact that during an attack or a failure the degree distribution of the network changes. In other words, we study network topology after the disturbance has happened.

Assume that in a finite network, p_k be the degree distribution of the network before the attack. Let us say that f_k represents the probability for a node of degree k being removed from the network and q_k represents the probability for a node of degree k surviving in the network. Here $0 \leq f_k \leq 1$. After the node selection, we divide the network into two subsets, one subset contains the surviving nodes (S) while the other subset comprises the nodes that are going to be removed (R). At the moment the nodes in R are actually removed, the degree distribution of the surviving nodes is changed due to the removal of the E edges that run between these two subsets. Let us assume that total number of nodes is N.

Total number of tips in the surviving subset, including the E links that are going to be removed is

$$P_S = N \sum_{k=0} k p_k (1 - f_k)$$

Now, we find number of edges E that run between these two subsets.It is

$$E = \frac{N \sum_{k=0} k p_k (1 - f_k)}{N \sum_{k=0} k p_k - 1} N \sum_{k=0} k p_k f_k$$

If Φ be the probability of finding an edge in the surviving subset S that is connected to a node of the subset R .Then

$$\Phi = \frac{E}{P_S} = \frac{N \sum_{k=0} k p_k f_k}{N \sum_{k=0} k p_k - 1} = \frac{\sum_{k=0} k p_k f_k}{\sum_{k=0} k p_k - 1/N}$$

In other words, we can say that Φ is the probability of an edge hanging after some nodes have been removed.

So, If p_q be the degree distribution in case of hanging edges then $p_q = p_k q_k$

5.1 Degree distribution of surviving network

Now, we need to find out the degree distribution of network without the hanging edges, which means that we have to neglect the concept of hanging edges.Let us assume that p'_k be the degree distribution without hanging edges and q be the degree of a node in surviving subset S . So,

$$p'_k = \sum_{q=k}^{+\infty} \binom{q}{k} \Phi^{q-k}(1-\Phi)^k p_q$$

We know that an infinite network percolates after an attack if $k' = \frac{<k^2>'}{<k>'} > 2$. That is called the **critical condition for infinite network**. We will borrow the same critical condition to define a "percolation" criterion for finite network. So, $k' = 2$ determines the point at which the network breaks down. To compute $< k >'$ and $< k2 >'$ we use the generating function:

$$G_0(x) = \sum_{k=0}^{+\infty} \sum_{q=k}^{+\infty} \binom{q}{k} \Phi^{q-k}(1-\Phi)^k p_q x_k$$

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

$$G_0(x) = \sum_{k=0}^{+\infty} p_q((x-1)(1-\Phi)+1)^q$$

From here, we can compute easily the first two moments.

$$< k >' = dG_0(1)/dx$$

$$< k^2 >' = d^2G_0(1)/dx^2 + dG_0(1)/dx$$

After some calculations we get

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

This equation determines the stability condition according to our definition for any uncorrelated network of finite size under any arbitrary attack. If we take limit of $N \lim +\infty$,this equation reduces to

$$(\sum_k kp_k(1-f_k))(\sum_k k^2p_k(1-f_k) + \sum_k kp_k(f_k-2)) = 0$$

$$\implies \sum_k k^2 p_k(1-f_k) + \sum_k k p_k(f_k - 2)) = 0$$

$$\implies \sum_k k p_k(kq_k + (1-q_k) - 2) = 0$$

$$\implies \sum_k k p_k(q_k(k-1) - 1) = 0$$

It is amusing to know that we have derived this result in the last section through a more classical generating function formalism.

5.2 Percolation threshold in case of random failure

In case of **random and degree independent failure** $f = f_k$. we put this value in the equation derived in last subsection.

For finite network, $[(\sum_k k p_k(1-f_k))(\sum_k k^2 p_k(1-f_k) + \sum_k k p_k(f_k-2)) + \frac{1}{N}(\sum_k k p_k(1-f_k)(2-k)) = 0$

$$\implies (1-f)[< k > ((1-f) < k^2 > + (f-2) < k >) + \frac{1}{N}(2 < k > - < k^2 >)] = 0$$

$$\implies < k > ((1-f) < k^2 > + (f-2) < k >) + \frac{1}{N}(2 < k > - < k^2 >) = 0$$

$$\implies < k > f(< k^2 > - < k >) = (< k^2 > -2 < k >) < k > + \frac{1}{N}(2 < k > - < k^2 >)$$

$$\implies f = 1 - \frac{1}{\frac{<k^2>}{<k>} - 1} + \frac{1}{N}(\frac{2 - \frac{<k^2>}{<k>}}{< k^2 > - < k >})$$

We know that

$$f^\infty = 1 - \frac{1}{\frac{<k^2>}{<k>} - 1}$$

So, $f = f^\infty + \epsilon(N)$. Here, $\epsilon(N)$ is a negative quantity. So, random failure is less in finite graphs.This means that infinite graphs are more robust.

Now, If we take an example of *scale free network*,which follows *power law distribution*,then $f_k = ck^\gamma$.Let us interpret the behavior of attack or failure with varying the value of γ.

If $\gamma = 0$, this means that network is attributed to a random attack. If $\gamma < 0$, this means that attack is biased towards lower degree nodes, whereas if $\gamma > 0$, attack is biased towards higher degree nodes. So,γ shows, how much information an attacker has.

In the end,we would want to clarify that all the results which have been derived here are valid only for uncorrelated networks.The effect of correlations on the percolation threshold for finite(and infinite) networks remains as one of the major challenge.