Resilience Measurements on Internet Topology Generators

Bachelor-Thesis von Nam Truong Le
September 2012

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Fachbereich Informatik
P2P Networks Group

Resilience Measurements on Internet Topology Generators

Vorgelegte Bachelor-Thesis von Nam Truong Le

Supervisor: Prof. Dr. Thorsten Strufe
1. Coordinator: Benjamin Schiller
2. Coordinator: Stefanie Roos

Tag der Einreichung: 17.09.2012

Erklärung zur Bachelor-Thesis

Hiermit versichere ich, die vorliegende Bachelor-Thesis ohne Hilfe Dritter nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 17.09.2012

(N. T. Le)

Abstract

As the Internet becomes increasingly important to all aspects of society, its disruption has many bad consequences. Hence it is necessary to increase the resilience of the Internet. We define *resilience* as the ability to work well, although a set of nodes of the network are removed. But it was not well analyzed how the Internet behaves against the node removals. The purpose of this thesis is to create many attack scenarios based on various types of node removals and resilience metric. Then we use the results to analyze the resilience of the Internet.

In this thesis we analyze the network's behavior against node removals using two types of metric. The first is based on the partition, such as the giant component, and the second on the path lengths of the networks, such as the average shortest path length. There are also two types of node removals, using a random attack and intentional attack. In the intentional attack scenario, we have different ways to choose the deleted nodes, one is based on the network global properties and the other is based on the local properties of each node on the network.

The result shows that we must combine these two types of metrics to have a best view on the networks. The partition based metrics give us the information when the network is fragmented, we call this point as the critical point, and the shortest paths based metrics show us more information about the network behaviors from the beginning to this critical point. The result also indicates that the degree and betweenness centrality removals are the most efficient methods. They can find the most vital nodes in the network. The random node removal nearly has no impact on the network.

Such an analysis is not analyzed. In this thesis, all of the metrics, the attack scenarios and the network generators are taken together to get a better knowledge about them.

Contents

1 Introduction

The *resilience* of a network is the ability to operate well when failures occur. Internal failures can be systems errors and external failures are attacks from outsides. The resilience of a network depends on which structure the network has, how the network is attacked and how we measure it.

The structure of the Internet has a strong meaning on management and performance. Good models of the Internet are necessary to develop and analyze it. A good model of the Internet should have the same structure properties as the Internet. It should have the phenomena that the Internet also has, such as the richclub phenomenon [11], the power law degree distribution [2] and so on. It should also react to the failures in the same way the Internet does, so that the results from its analysis can be applied to the Internet. The Barabasi Albert Model (the BA model) was long time ago designed to model the Internet scale-free properties, see [1]. Currently there are a number of network models which were introduced to improve the ability of approximating the Internet, such as the Interactive Growth Model (the IG model), the Positive-Feedback Preference Model (the PFP model) and the Generalized Linear Preference Model (the GLP model), see [10], [12] and [2]. Besides, node removals were also well analyzed. It is divided into two types: random and intentional attacks. The degree centrality was often chosen for the intentional attack [6] [4]. The other types of intentional attacks, that use eigenvector, closeness, betweenness and effective eccentricity centrality, were not well analyzed. They were studied separately with the major purpose to rank the nodes in a network. Now we combine all of them together. But not only this, we also use different methods to quantify the resilience of networks. From simple metrics as the largest component to the complicated metrics such as the effective diameter, the largest biconnected component.

Our goals are to confirm and widen the results of related papers by running more simulations and then try to analyze and explain the new results. To do this we must develop tools to automate the deployment of the resilience measurement. After all we see which types of network model are closer to the Internet (it can depend on which metrics are used) and better against attacks, which type of attacks is the best way to partition a network. In Section 2 we recall some basic background of graphs and graph models. Then some popular resilience metrics and the ways used to attack the networks are introduced. In Section 3, we discuss the results of the simulation.

2 Background

In this section, we go through all concepts and definitions used in the thesis. First, we provide a briefly worded introduction of the graphs theory and how the Internet is modeled as a graph. Second, we discuss about the recently introduced network generators in an effort to recreate the AS level network. Finally, we choose some meaningful metrics and node removals that can be implemented with an acceptable runtime and then recall their definitions.

2.1 Graphs

In this paper, we focus on the network of the Internet, particularly the AS network of the Internet. In fact, there are many factors that affect this network. For example, how far it is between any two AS of the network, how fast the data can be transferred on the line, etc.. Here we are working with an idealized modeling of this network. An AS is considered as a node and the lines between them are the edges. All nodes are the same and so are all edges. The set of them is called a graph. This graph models the network of the AS layer of the Internet. First of all let's recall some definitions as well as some notations which are used throughout this article.

A graph $G = (V, E)$ consists of a nonempty set of nodes, V, and a set of pairs of elements of V representing edges, E. There are directed and undirected graph as the example in Figure 1. Directed edges can be seen as two-way roads and

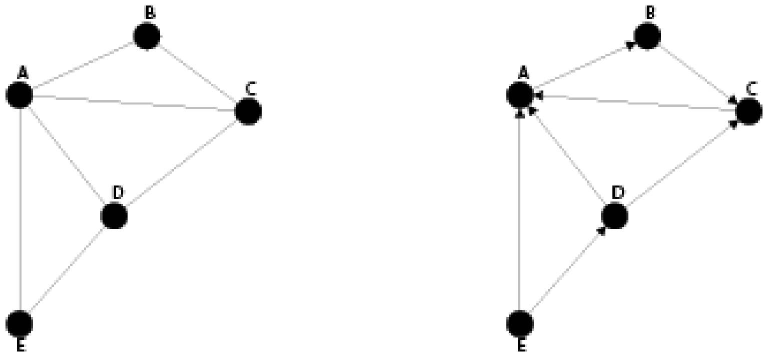


Figure 1: Undirected graph (left) and directed graph (right)

undirected edges as one-way road. An undirected graph can also be represented by a directed graph. For our purpose, we use the undirected graph as the modeling of the Internet. And the two following restrictions are also used: there are no loops, which means that two nodes of an edges are different; It can not have multiple edges joining nodes. So, between two nodes there are at most one edge.

When two nodes $u, v \in V$ are endpoints of an edge, we say u and v are *adjacent*.

A *path* is a subset of the graph whose nodes can be ordered so that two consecutive nodes are adjacent. A path that begins at node u and ends at nodes v is called a u, v- path. In Figure 2 a path from A to B is highlighted.

A *geodesic* from u to v is a path of minimum length. The *geodesic distance* $d(u, v)$ between u and v is the length of the

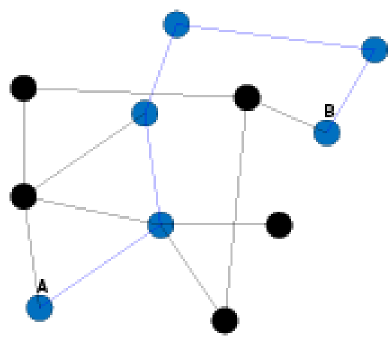
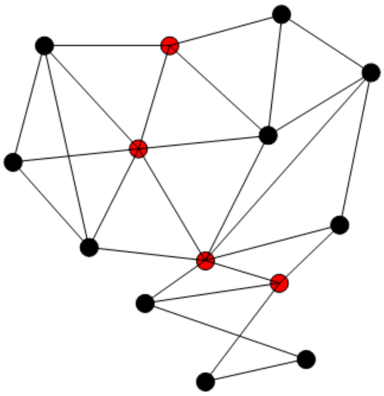


Figure 2: A path from A to B

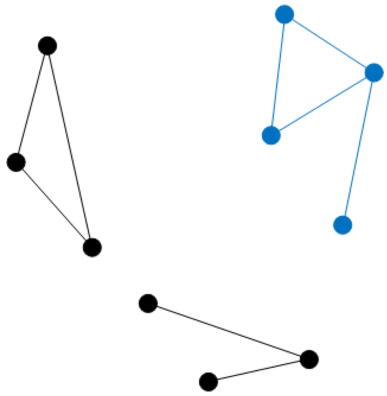
geodesic (the number of edges in this geodesic). If there is no path from u to v, the geodesic distance is *infinite*.

A graph $G = \{V, E\}$ is *connected* if for every $u, v \in V$ there exists an u, v- path in G. Otherwise G is called *disconnected*.

A *component* of G is a maximal connected subgraph. For example, we have a network in Figure 3a. We delete the red-marked nodes, the network is not connected any more and the result is three components Figure 3b. The blue component is the largest component.



(a) Original graph



(b) Graph after some removals

Figure 3: Components of a graph

The degree of a node v, $d(v)$ is the number of edges connected with this node v.
For directed graphs, indegree and outdegree are used instead of degree. The *indegree* $d^{-}(u)$ of u is the number of edges that end at u, the *outdegree* $d^{+}(u)$ is the number of edges that start at u. And if we model an undirected graph G as a directed graph G', the corresponding of u is u', then we have $d(u) = d^{-}(u') = d^{+}(u')$.
In undirected graphs, two nodes are connected if they have a path connecting them. But it is more complicated in directed graph. In directed graph, we can have a path from u to v but no paths from v to u. So, we say that a node u is *strongly connected* to v if two paths exist, one from u to v and another from v to u. And two nodes u and v are *weakly connected* if there's either a path from a to b or one from b to a but not necessarily both.
The clustering coefficient and the transitivity are the measure of how nodes in a graph are clustered together. The transitivity is based on triplets of nodes. A triplet is three nodes that are connected by either two or three links. They are called *open triplet* and *close triplet*. A triangle is counted as three close triplet. The transitivity is then defined as:

$$\text{Transitivity} = \frac{3 \times \text{number of triangles}}{\text{number of connected triplet of nodes}} = \frac{\text{number of closed triplet}}{\text{number of connected triples of nodes}}$$

A local clustering coefficient of a node means how close its neighbors is to be a complete graph. A node u has N neighbors and we have M links between these N nodes. Then the local clustering coefficient of u is:

$$C(u) = \frac{2M}{N(N-1)}$$

The network's clustering coefficient is the average of the local clustering coefficient over all nodes of the network:

$$C(G) = \frac{1}{|V|} \sum_{u \in V} C(u)$$

2.2 Internet Topology Models

As mentioned before, we focus on analysis the Internet and in particular at the AS level. So here we aim at using the network models, which was built to recreate the Internet AS level topology. And then those results are compared with the real network structure of the Internet, taken from CAIDA - *The Cooperative Association for Internet Data Analysis*.

2.2.1 Barabasi Albert Model

Some network models were made in the past but they did not simulate two important properties of the Internet: the number of nodes increasing with time and new nodes connected to old nodes using a preferential attachment mechanism, such as the Erdos-Renyi model (ER model) which was introduced in [5]. The number of nodes in this model was determined and the edges were randomly selected among the nodes. So we do not have a mechanism to add a new node. In contrast, the number of nodes in the Internet increases over time.
Random network models assume that the probability that two nodes are connected is the same throughout the network. But the reality is not so, when a new node appears, it tends to connect to the *important* nodes. Here, we can understand the importance simply as the degree of nodes. *The Barabasi Albert model* (BA model), see [1], is designed to overcome

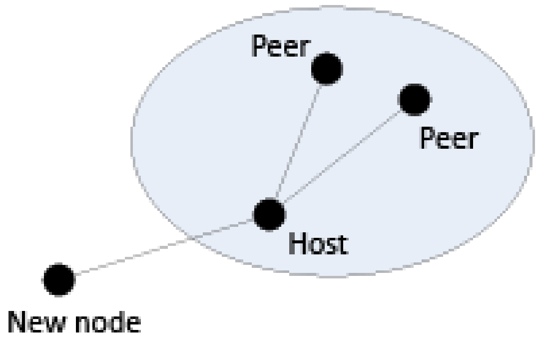
this disadvantage. In the stage that the network has N nodes, a new node appears to be connected to m other nodes of the network. The probability that a node u is connected to the new node, $\Pi(u)$, is:

$$\Pi(u) = \frac{d(u)}{\sum_{v \in V} d(v)}$$

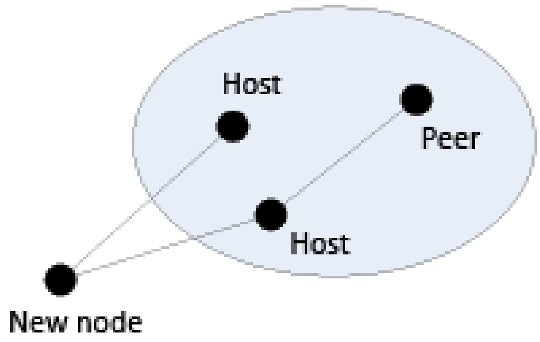
Specifically the growth of networks takes place as follows:

- We start with a small random graph with m_0 nodes.
- At each step, a new node is added and $m \leq m_0$ edges connect this new node to other nodes.

2.2.2 Interactive Growth Model



(a) Strategy 1



(b) Strategy 2

Figure 4: The growth of a new node in IG model

There is a property of the AS graph that BA model does not express, the rich-club phenomenon. As described in the article [11], the rich-club connectivity is defined as follows: the nodes are sorted in descending order, using the degree of node. For n nodes with rank less than r_{max}, the rich-club connectivity $\phi(r)$ is defined as the ratio of the total actual number of edges to the maximum possible number of edges between these n nodes. Here, r is normalized by the total number of nodes. And the author also found that the BA model has no rich-club. The reason is because the new edges are connected only to the new node, so the rich nodes are not well connected together. It leads to a suggestion for a new network model: while the network is growing, new edges are also added between the old nodes. *The Interactive Growth Model* (IG Model) is described in [10]:
At each step, a new node is connected to the old nodes by two mechanisms:

- With the probability p, the new node is connected to one old node called *host*. This host is also connected to two other old nodes called *peers*, see Figure 4a.
- With the probability $1 - p$, the new node is connected to two hosts. One of these hosts is connected to one peer, see Figure 4b.

This model uses the same linear preference as the BA model. Adding new edges between old nodes has the significant effect, the rich-nodes are connected to each other better and they have higher degree than the nodes in BA model.

2.2.3 Positive-Feedback Preference Model

As stated in [12], the IG model still has a limit, the highest degree in AS is greater than the highest degree in the IG model. To overcome this limitation, we have to make up the highest degree in the network. Note that, new edges are selected based on a defined preference. In BA and IG model the preference is linear. We can change the preference, then the high degree nodes are more priority when choosing new edges. The IG model was modified by using a nonlinear preference:

$$\Pi(u) = \frac{d(u)^{1+\delta \log_{10} d(u)}}{\sum_{v \in V} d(v)^{1+\delta \log_{10} d(v)}}$$

and the new model is called *The Positive-Feedback Preference Model* (PFP model). The growth of the network is also improved as follows [12]:

- With probability $p \in [0, 1]$, a new node is connected to one host node. One edge appears between the host node and a peer node.
- With probability $q \in [0, 1-p]$, a new node is connected to one host node. Two edges appear between the host node and two peer nodes.
- With probability $1-p-q$, a new node is connected to two host nodes. One edge appears between one of the host nodes and a peer node.

With this improvement we can adjust the network's average degree. In IG model, the average degree is always 6. Now we can bring it down close to the value of the AS.

2.2.4 Genralized Linear Preference Model

Tian Bu *et al.* [2] observe that some network generators have difficulty to reproduce the characteristic path length and the cluster coefficient. As defined in Section 2.3.4, the characteristic path length is the average value of all shortest paths in the network. Tian Bu *et al.* [2] use the following definitions:

- The characteristic path length L is the median of the means of the shortest path lengths connecting each vertex u to all other nodes.
- The clustering coefficient $\gamma = \frac{\sum_{u \in V} \gamma(u)}{|V|}$. Here $\gamma(u) = \frac{\text{The number of edges between all neighbours of u}}{d(u)(d(u)-1)/2}$

They modified the BA model to get more accuracy in reproducing the characteristic path length and the cluster coefficient. First the linear preference was modified as follows:

$$\Pi(u) = \frac{d(u) - \beta}{\sum\limits_{v \in V} (d(v) - \beta)}$$

where $\beta \in (-\infty, 1)$ is a parameter. Starting with a small graph from m_0 nodes and $m_0 - 1$ edges, this model has two growth mechanisms:

- With the probability p, $m \leq m_0$ edges is added using the defined preference.
- With the probability $1-p$, a new node u is added. m new edges connect u with m already existed node using the defined preference.

This model is called *the Generalized Linear Preference Model* (GLP model).

2.2.5 Parallel Addition and Rewiring Growth Model

In [8] M. Piraveenan *et al.* introduced a network model producing the local assortativity that can be observed in the real AS level network. The growth mechanism was defined as follows: the model start with a small network of size N. This could be a simple random graph. As growthing nodes are added to the network. At each step, a new node are connected to N_{add} number of old nodes using linear preference attachment. This means the probability of an old node i with degree d_i is $p_i = \frac{d_i}{\sum_{j \in \text{old nodes}} d_j}$.
That is the node-addition mechanism of the *Parallel Addition and Rewiring Growth Model* (PARG). This model has another mechanism is the link rearrangement mechanism. After each node deletion, N_{del} number of links in the network are deleted. These N_{del} link are chosen by the following manner: choose N_{cut} nodes with highest degree. The links in the selected node is deleted using probabilities:

$$p = \frac{N_{\text{del}}}{dd_{\text{cut}}}$$

where d is the degree of a node, that node has degree-based rank $rank_d \geq n_{\text{cut}}$, d_{cut} is the degree of the node that have exactly the rank of $rank_d$.
After each link deletion, two new links are created. From two nodes connected by the deleted link, a node with higher degree is chosen. We create new links from this node to other two nodes from the network. We then rank all the nodes decreasing. The higher the rank is, the higher the probability is to be selected. You can find concreter description in [8].

2.3 Resilience Metrics

To simulate and study the properties of the network, we need to have some measures, which can be compared to see which network is more resistant and which network is closer to the CAIDA than the others. In recent years there have been many such measures proposed by some authors. In this article we introduce some that can be applied to massive networks in term of the number of nodes and links.

2.3.1 The Largest Component

The first factor we can think of to evaluate a network is its connectivity. Two nodes u and v are called *connected* if there is a set of edges e_1, e_2,..., e_n and a set of nodes v_0, v_1,..., v_n such that e_i is the edge between v_{i-1} and v_i and $v_0 = u$, $v_n = v$. A set of nodes is called *connected* if every pairs of nodes from it are connected. The set of edges e_1, e_2,..., e_n is called a path from u to v and n is the length of this path.

We can think that when a set of nodes was isolated (i.e. we delete some edges so that these nodes is no longer connected to the network) the network does not work well anymore. But in reality, a network still works fine if it lost some nodes, such as the Internet. Some clients may be isolated as the entire Internet still works normally and is not affected by such isolation. And if we choose *the connectivity* as a measure for the sustainability of the network, we only have two values, that are *connected* and *not connected*. This is not appropriate to compare multiple networks together. Instead, we use the concept *giant component* or *the largest component*, as in [6].

We can separate the network into disjunct subgraphs so that each node belongs to exactly one subgraph, each subgraph is connected and no two nodes of different subgraphs are connected together. Each of these subgraphs is called *component*, or *connected component*. The size of a component is the number of nodes in this component. And we define the component with maximal size as *the largest component*. We use the largest component's size as our metric. Suppose that $\{G_1, G_2, \ldots, G_k\}$ is the set of components of the graph G. The largest component's size is defined as:

$$\max\{|G_i||1 \le i \le k\}$$

It is more meaningful when we use giant component to compare networks, a network perform better if its giant component is larger than the giant component of the other.

2.3.2 The Largest Biconnected Component

Now we consider an extension of the giant component, the biconnected giant component. In the actual situation, the network depends not only on a single line, a lot of network structure is built so that when a line is broken, the computer (or a cluster of computers, or an organization) are still capable of connecting to the network with other lines. From here, the term *biconnected* are given in [6] as follow: two paths connecting the same pair of nodes are *independent*, if there are no node that exists in both paths. *A biconnected component* is a subset of nodes such that every pair of nodes is connected with at least two independent paths. And *The largest biconnected component* is a biconnected component with maximal number of nodes.

2.3.3 Diameter

Another important thing in network is the length of paths. The distance between two nodes u and v, $d(u,v)$, is the shortest path between them. If the distance between two nodes is small, then the time needed to transmit data between them is also shorter. It suggests a different measure for network, the diameter. The diameter of a network G, $D(G)$, is the biggest distance between all nodes. If the diameter is small, the network can be considered better in the sense that it takes less time to transmit data between nodes.

$$D(G) = \max_{u,v \in V} d(u,v)$$

2.3.4 Average Shortest Path Length

There is a disadvantage when using diameter, it is just a maximum value. In many cases, the diameter does not reflect exactly the current state of the network. For example, two networks have the same diameter 5, but the first network have more edges and its nodes are well connected together, see Figure 5. The first network, the network in Figure 5a, is clearly better than the second network. Thereby, the average value of distances is also an important measure for networks. That value is called *Average Shortest Path Length* (ASPL) and is defined as follows

$$\mathrm{ASPL(G)} = \frac{2 \cdot \sum\limits_{\mathrm{u,v \in V}} \mathrm{d(u,v)}}{|\mathrm{V}|\,(|\mathrm{V}|-1)}$$

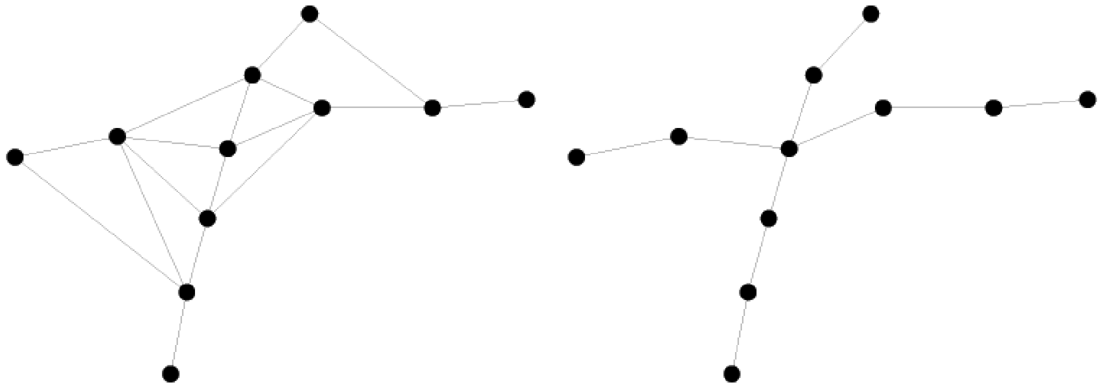


Figure 5: Two networks with the same diameter

The average shortest path length is also called *the characteristic path length*.
But what would happen if the networks are not connected anymore. Then for any two nodes u and v, which are not connected to each other, how can $g(u,v)$ be similarly defined? Obviously we then think of the definition of $g(u,v) = \infty$, but then $\mathrm{ASPL(G)} = \infty$. So we cannot compare unconnected networks because they have the same value of ASPL. To avoid this, *The Average Inverse Shortest Path Length* (AISPL) was used instead ASPL in case we want to study both unconnected networks.

$$\mathrm{AISPL(G)} = \frac{2 \cdot \sum\limits_{u,v \in V} \frac{1}{d(u,v)}}{|V|\,(|V|-1)}$$

There is another way to overcome the disadvantage of ASPL. We define P is the set of all connected pairs in G, then:

$$\mathrm{ASP(G)} = \frac{\sum\limits_{p \in P} d(p)}{|P|}$$

2.3.5 Effective Diameter

In [7], the effective diameter was introduced as an average statistic of a network. Here *effective* means that 90% *of all pairs can be connected within a distance*. It was defined as follows:

$$\mathrm{diam_{eff}(G)} = \min\{h \in \mathbb{N} | P(h) \geq 0.9P(\infty)\}$$

where P is defined:

$$P(h) = \left|\{(u,v) \in V^2 | d(u,v) \leq h\}\right|$$

2.4 Failure Scenarios and Attack Strategies

To assess the resilience, we do it in the following manner: in a certain order, each node of the network is deleted. Every time you delete a node, one of the metrics that we have described in Section 2.3 is recalculated on the new network. Then we study and analyze the network based on the results of the measures. The problem is, in which order we want to remove the nodes? Random or by a certain sort.
Of course, when we want to partition a network, we delete starting from the *most important* nodes. But we should find a way to know whether one node is *more important* than the others. The first consideration is the degree of the nodes. A node is more important than the others if it has more degree. But degree is not the only one we can use. For a communication network or the Internet, shortest paths are also a factor worth considering. Then we also describe some type of node sorting based on this factor.

2.4.1 Random

In this type of attack, we remove the nodes randomly. This type of nodes removal is already well analyzed. Results show that the Internet AS level network and other scale-free networks are stable against a random nodes removal. But we still want to take it as a node removal into our simulation to validate the previous results from other papers and to compare the effect of other node removals with the random method on various types of network.

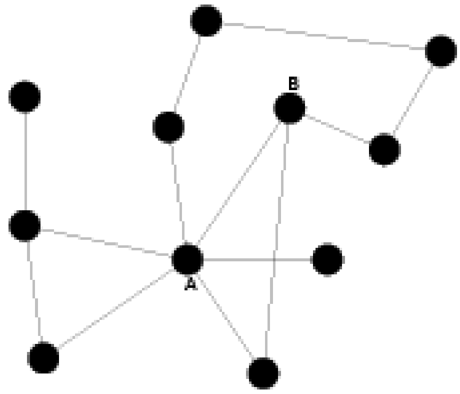


Figure 6: Deleting node A split the network into many components

2.4.2 Degree

We consider first the sorting based on the degree of the nodes. This is the sorting that we first thought as the simplest sorting. In social network, if a person has many friends, he *may* be an important one. The same thing can be seen in a communication network, where a node that connected to a large number of other nodes could be one of the most important nodes, too. And obviously deleting it has a large effect to the network, such as the connectivity of the network. In Figure 6, if we delete node A, the network is not connected anymore and is split into three connected subgraphs. However, if we delete node B, the network is still connected. Later we compare this simple method with other complicated ones.

2.4.3 Closeness Centrality

Now we study centralities based on distances between nodes. Yet, the distances tell us only a individual property between two nodes together. Thus, to have an assessment of a node in terms of the whole network, the following concepts are defined: the *farness* of a node is the sum of its distances to all other nodes, and its *closeness* is defined as the inverse of the farness

$$\mathrm{CC(u)} = \frac{1}{\sum\limits_{\mathrm{v \in G, v \neq u}} \mathrm{d(u,v)}}$$

The larger the closeness of a node is, the more it is in the middle of the network. That means if we take another node randomly and send a packet from the node to this random node, the time (on average) is shorter.

2.4.4 Betweenness Centrality

Usually when u want to send a packet to v, u try to find a shortest path to send. Yet, we have more than one shortest path from u to v. So if t appears in these shortest paths many times, then with high probability u must send the packet through t. More specifically, σ_{uv} is the number of shortest paths from u to v and $\sigma_{uv}(t)$ is the number of shortest paths from u to v through t, then the probability is $\frac{\sigma_{uv}(t)}{\sigma_{uv}}$. Applying it to all pairs of nodes, we have *the betweenness centrality* of t:

$$\mathrm{BC(t)} = \sum_{\mathrm{u \neq v \neq t}} \frac{\sigma_{\mathrm{uv}}(\mathrm{t})}{\sigma_{\mathrm{uv}}}$$

A node having a high value of betweenness centrality lies on many shortest paths. If it is removed, many shortest paths are lost. So it lengthen some paths between nodes.

2.4.5 Effective Eccentricity

Next we introduce the final sorting based on shortest paths, which was described in [7], *the effective eccentricity*. The effective eccentricity of node u is the least distance h, such that 90% of u's reachable nodes are connected to u with path lengths less than or equal to h

$$\mathrm{EE(u)} = \min_{\mathrm{h}} \{\mathrm{N(u,h)} \geq \mathrm{N(u,\infty)}\}$$

here $N(u,h)$ is the number of nodes that are connected to u with path lengths less than or equal to h. As indicated in [7], *nodes with high eccentricity are important for the network; their failures create problems to the connectivity of the network. At the same time, we see that eccentricity has a different effect than the degree regarding the connectivity*. Later we see if this observation is correct.

2.4.6 Eigenvector Centrality

Finally we look at *the eigenvector centrality*. It is based on the influence of a node in the network. A node has a higher score if it is connected to high-scoring nodes. Let EVC(u) be the eigenvector centrality score of node u and $u_1, u_2, \dots, u_N$ be the nodes of the network. Then

$$\mathrm{EVC}(\mathrm{u_i}) = \frac{1}{\lambda} \sum_{\mathrm{u_j} \in \Gamma(\mathrm{u_i})} \mathrm{EVC}(\mathrm{u_j})$$

here λ is a constant. And let A be the adjacency matrix of the network. The equation can be rewritten as

$$\mathrm{EVC}(\mathrm{u_i}) = \frac{1}{\lambda} \sum_{\mathrm{j}=1}^{\mathrm{N}} \mathrm{A_{ij}} \mathrm{EVC}(\mathrm{u_j})$$

Let x be the vector $\left(\mathrm{EVC}(\mathrm{u_1}), \mathrm{EVC}(\mathrm{u_2}), \dots, \mathrm{EVC}(\mathrm{u_N})\right)$, we have

$$x = \frac{1}{\lambda} Ax \Rightarrow \lambda x = Ax$$

hence, x is an eigenvector of the matrix A.

3 Evaluation

In this section, we present the results of our simulation. In Section 3.1, we show how we setup the data and run the simulations. Some expectations before the simulation execution are discussed in Section 3.2. The main content is the Section 3.3 and the Section 3.4. The results are divided into two parts: The first is related to the network topology properties and the second, also the main part of the thesis is the resilience measure. In the first part, we measure the properties of networks such as the degree distribution and the richclub connectivity. In the second part, we analyze the behavior of the networks against various types of node removal.

3.1 Evaluation Setups

We use the graph generators described in Section 2.1 to generate graphs. Some of these graph generators use some parameters in growth mechanism or a small start graph. In this section, we describe the used parameters in our simulation. We want to adjust these parameter to create graphs similar to the AS graph from CAIDA. However, we encountered a few difficulties, such as the maximum or the average degree of nodes. There are many sources to get a graph of Internet AS level. Our AS graph from CAIDA is different from others from the papers, where the network generator was described. Another thing is, the parameter is not easy to change to adapt to new network. In IG growth mechanism, there are always three new edges to add for each iteration. So the average degree is always six. In PFP growth mechanism, we have another strategy to growth, the new node is connected to one host and this host is connected to a peer. So the number of added edges is two or three. Hence, the average degree is between four and six. It depends on the probabilities p and q. For the best result, see [12], these probabilities are already determined. In [2], the parameter β is chosen by performing a linear regression. We want to use the same values as the author did. Hence, except Barabasi Albert model, all others are used with the default parameters from the authors.

We did not implement the PARG model successfully. Because of its incompleteness of algorithm description, our produced networks have the different values of network topology metric from the results from [8]. Thereby, we do not take this model to our simulation.

To compare these graph generators with the actual AS graph, we also take the CAIDA graph into account. Six AS graphs from 2007 to 2012 were used. For each generator we also create six sets of graphs, each consists of 10 graphs. In one set, all graphs have the same number of nodes as the AS level graph. We compare the results year by year, and if there are no differences among these years, we take the result from year 2012 to our analysis. First we calculate the degree distribution of the graphs and compare their results with the degree distribution of AS graph. We see then which graph generator creates better results according to distinct metrics in terms of approximation CAIDA graphs. Afterwards, the second part of simulations is performed. Using each node sorting we remove nodes from graphs. After each removing, we recalculate the component size, the diameter, the shortest path lengths,... and see how these graphs behave with each type of removal.

The calculation of all pairs shortest path takes a very long time, especially when we must recalculate after each removing for a graph with thousands of nodes. But we recognize that after a certain times of removing the value is nearly unchanged or is not meaningful any more. For example the shortest paths based metrics, such as diameter, say nothing if the graph is divided into many small components. With some trials we see that after 10% the graph is nearly partitioned, so we decide to make a modification of our dynamical method: for the first 10% of nodes we remove one node at each step, for the last 90% of nodes we remove nodes percentagewise. That means, we divide the last into 100 groups of nodes and delete each at one step. So we have $\frac{|V|}{10} + 100$ steps for each calculation. In case with this modification the runtime is still too long, we perform the removing percentage already at the beginning. With this modification we save a lot of time for the simulation.

3.1.1 Contructing The Networks

The following is how we create the networks, how we set the parameters of each network generator.

Barabasi Albert Model

The AS graph from CAIDA have the average degree about 4, so we decide to use $m = 2$.

The Interactive Growth Model

As in [10], it starts with a random graph of 10 nodes and 10 edges. The probability p is choosen to be 0.4.

PFP Model

To procedure the best result, $\beta = 0.048$ is used [12]. The probabilities p and q are 0.3 and 0.1.

GLP Model

After their calculation, Tian Bu *et al.* [2] use $m = 1.13$ and $\beta = 0.6447$. In the model m must be an integer but we can recognize it using a random variable. With the probability 0.87, $m = 1$ is used and with the probability 0.13, $m = 2$ instead.

3.1.2 Calculating The Metrics

Properties of The Generated Topologies

The degree distribution, the richclub connectivity the transitivity and the clustering coefficient are calculated for each graphs.

Resilience Properties of The Generated Topologies

The largest component, the largest biconnected component, the diameter, the average shortest path length and the effective diameter are calculated dynamically, each with all of six node removing modes: random, degree, eigenvector centrality, closeness centrality, betweenness centrality and effective eccentricity centrality node removals.

3.2 Expectations

Based on the results of previous papers, some results are expected to be similar to the previous one from these papers. Here we discuss some predictions before performing the simulation.

3.2.1 Properties of The Generated Topologies

In the network models we use, the BA is given first, followed by GLP, IG and PFP. And it also pointed out that, the newer network models recreate more properties of the Internet. In [1], it was shown that the BA model create networks, which follows a power law distribution. The GLP model recreate the characteristic path length and the cluster coefficient properties of AS network [2]. The IG model in [10] resembles not only the degree distribution of the Internet, but also the rich-club phenomenon. Later on the PFP model in [12] reproduces many properties of the Internet, the degree distribution, the rich-club phenomenon, the disassortative mixing and so on. Our first part of simulation verifies these properties of the networks models.

Degree Distribution

Because all the networks are designed to follow the scale-free degree distribution of AS network, we expect that it is correct with our simulation too. With preferential attachment IG, PFP, GLP would have a higher maximal degree. The PFP has a strong preferential attachment, therefore we think that its maximal degree is at most. The PFP and IG should have a higher average degree because of its own mechanism. With an additional growth mechanism with only two new edges, the average degree of PFP should be a litte bit lower than IG but still not higher than the others.

Clustering Coefficient and Transitivity

First of all the BA network model should generate networks with small value of clustering coefficient and transitivity. The edges are added randomly using a weak preference attachment, no new edges are added among the old nodes when graph growths, so the graph is not well clustered. BA has no node with degree of one, because of our chosen parameters. These nodes which have degree one, has local clustering coefficient one. In other models, the edges are added with a stronger preference attachment. Especially GLP has a growth mechanism, in which no new node is added, only new edges appear among the old nodes, thereby more triangles occur. Hence the transitivity of the others is also bigger than BA. The new growth mechanism of PFP, that is not included in IG, would be a reason to make the transitivity of PFP lower than IG. Because with this, a fewer number of new edges is added among the old nodes while graph growths.

Richclub Connectivity

Except BA, the others three network generators are designed to follow the richclub connectivity of AS graph. We expect that our simulation show the same results too. And we can see how good the three generators capture the richclub phenomenon. With the same reason for the transitivity, we expect that GLP's richclub members are connected well together. PFP and IG should have the same result. We see later how the CAIDA behave compare to the others.

Diameter - The Average Shortest Path Length

PFP should have high average degree, a high maximal degree and a high clustering coefficient, we expect that PFP should has a small value of the diameter and the average shortest path length. The shortest path lengths of PFP should have lower values than the others. In contrast, the shortest path length range of BA should be higher than the other but its diameter should not be much different from the others. GLP should have a good approximation to CAIDA in the shortest path length distribution.

3.2.2 Resilience Properties of The Generated Topologies

In the second part of simulation, we analyze the resilience with regard to various metrics. With each metric, nodes are eliminated using a specific node sorting. We see then which network is better than the others in terms of each metric and which node sorting can find the best vital nodes.
In [3] it was shown that the Internet is resistant to a random nodes removing even if nearly 100% nodes are removed. But the Internet is highly sensitive to an intentional nodes removing [4]. The first choice is removing node using the degree centrality. The other node removals should have high impact comparing to random like degree node sorting. The BA is designed only to follow the degree distribution of the Internet AS level, hence it should be only sensitive with the degree node removal. Other networks approximate the other Internet topology properties better, such as the clustering coefficient and the shortest path lengths. We expect that these networks are more sensitive against the other node removals than BA.

The Largest Component

As stated in [3] and [4], the scale-free networks are stable again random attack but very sensitive again intentional attack. It was shown that the largest component is partitioned fast when we delete the high degree node first, but this component's size is still big when the nodes are removed randomly. We expect that our results validate this conclusion and we see then how the networks react again the other type of intentional attacks.

The Largest Biconnected Component

In [6], it was shown that this metric has same properties as the largest component. Its size would decrease as nodes are removed. It is stable again a random attack and is not partitioned quickly by random node removing. But the size of the largest biconnected component should be smaller than the largest component. And its decreasing speed should be faster than the largest component.

Diameter - The Average Shortest Path Length - The Effective Diameter

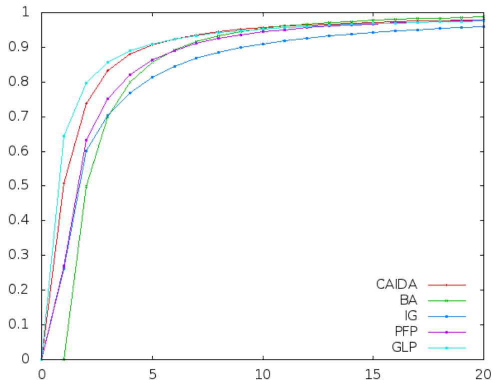
These three metrics are somehow similar. The diameter is the longest values of path lengths, the average shortest path length is their arithmetic means. The third, the effective diameter is a modification of diameter, we hope that it behaves not much more different than the diameter. At the beginning, after each removing, the graphs should have bigger diameter, ASLP and perhaps effective diameter too, because node deletion affects the shortest paths of the network. Pairs of nodes must find other paths with longer distance when a node in their shortest path is deleted. Hence the paths become longer. But this happens only until a point where the network is already broken. At this point, the network is split into many disjunct components. From this point, the nodes that we have selected at the beginning are no longer vital to the network. Node removing has no effect anymore and the metric value goes to zero as the number of nodes decreases.

3.3 Properties of the generated Topologies

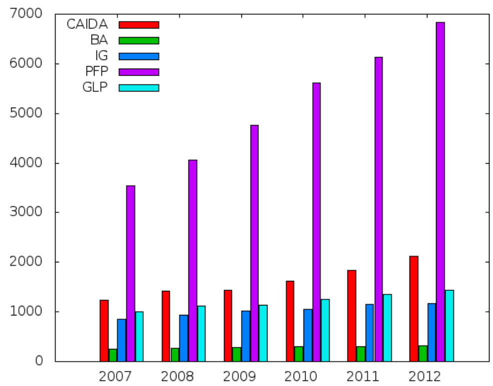
We discuss now the results of network topology. It includes the degree distribution, the richclub connectivity, the clustering coefficient and the transitivity. The simulation results of the network generators are compared with the CAIDA's results to see what is the best approximated network generator.

3.3.1 Degree Distribution

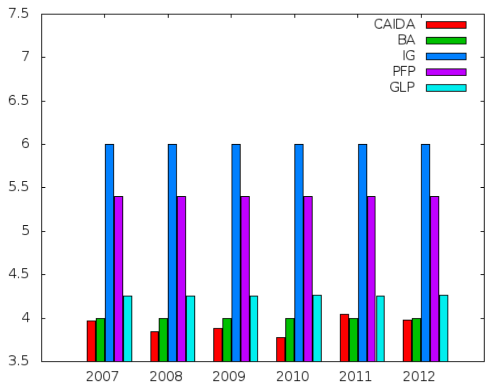
In Figure 7 we can see that the maximal degree of PFP is larger than the others. But it does not mean that the PFP generates poor quality networks in term of approximating the properties of CAIDA. The reason is that the average degree of PFP is also too much higher than CAIDA, BA and GLP. Later we see that PFP recreates good approximations to other CAIDA's topology properties. A strange result is that IG has more average degree than PFP but less maximal degree than PFP. In [12], it's stated that because of the fact that the AS graph has more maximal degree than IG and BA, the PFP was designed to make the maximal degree higher. However from our simulation's result we can see that the maximal degree of PFP is too high compared with CAIDA. The problem is, our AS level graph is different from the AS graph in [12]. The average degree and maximal degree of AS graph in [12] are 5.4 and 2839, but our CAIDA 's results are much smaller. This is why the PFP degree distribution is considered not so good to approximate CAIDA graph. But our results confirm once again the result from [12], the maximal degree of PFP is about a quarter of the number of nodes.
The average degree of BA is nearly equal to CAIDA's value, but the maximal degree is much smaller. It is easy to see it, because new edges in BA are selected randomly and BA has no low degree nodes. We have no reference attachment, so we don't have the phenomenon *rich get richer*. Overall it can be indicated that in our simulation the GLP and PFP generate good approximate networks to CAIDA than the other. In Figure 7, the degree distribution of GLP is nearly the degree distribution of CAIDA.



(a) Degree distribution



(b) Max degree



(c) Average degree

Figure 7: Degree distribution

3.3.2 Richclub Connectivity

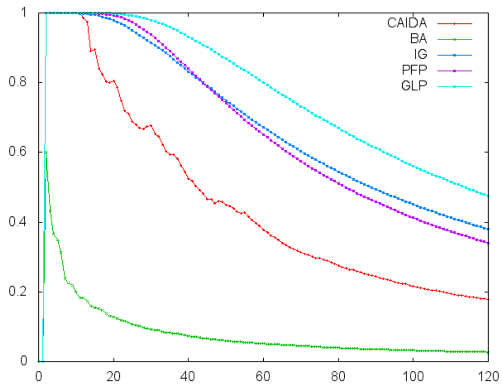
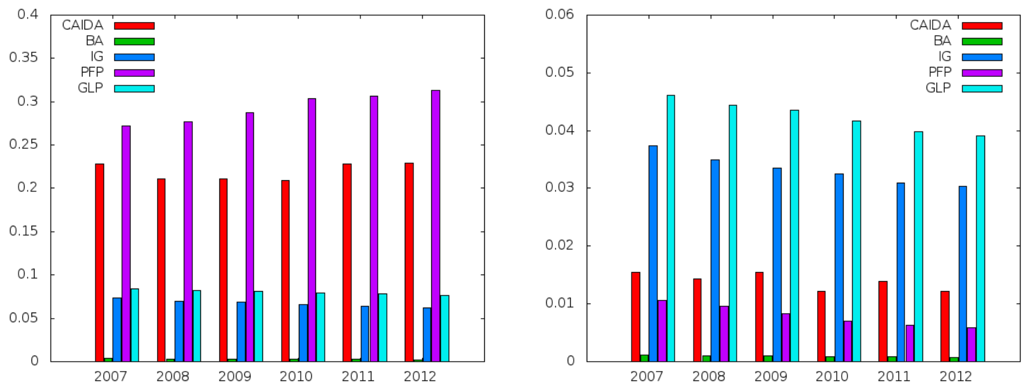


Figure 8: Richclub connectivity

As expected, BA graph does not have the richclub phenomenon. The rich nodes of the BA graph are not well connected together. The first 10 high degree nodes are connected together by just only 20% the number of possible edges between them. Whereas the first 10 high degree nodes of the others make a full-mesh subgraph. The others three network generators have higher richclub connectivity but also higher than CAIDA's value. With a richclub of more than 12 highest degree node members, the club of CAIDA is not well connected together as these clubs from IG, PFP and GLP. Compared with CAIDA, IG and PFP give the better results. GLP has the highest richclub connectivity. The reason is the growth mechanism of GLP. With the probability p, $m \leq m_0$ edges is added using the defined preference, see Section 2.2.4. So, the richclub has more edges than the other network generators. This phenomenon describes an important property of the Interet AS level network. The network are constructed hierarchically. First it is based on a small network called backbone. The member of the backbone is very well connected together. They are the vital part of the overall Internet that makes the Internet operate well. Each of these nodes are connected to a large number of nodes making a subnetwork of the Internet. When a node wants to send a packet to another nodes not living in a same subnetwork, it sends the packet to the backbone network. Because the backbone is nearly a full-mesh graph, it does not take a long time to send a packet to one member of backbone that connected to the target. It's why packets are efficiently transmitted in the Internet. The richclub member in our analysis is considered as the backbone of the network.

3.3.3 Clustering Coefficient and Transitivity



(a) Clustering coefficient **(b)** Transitivity

Figure 9: Clustering coefficient and transitivity

It is easy to see that BA has much lower clustering coefficient, see Figure 9a, than expected. The reason is the random edges addition with a weak preference attachment and no edges are added among the old nodes while networks grow. The others have a recognizably more value of clustering coefficient. IG and GLP has nearly equal clustering coefficient, besides PFP and CAIDA has a higher value of clustering coefficient. The reason is that, the PFP has a high average degree. With more edges, the nodes are better connected together. The IG has a high average degree too, but its max degree is lower than PFP's value and the PFP has more low degree nodes than IG, these nodes have a higher value of local clustering coefficient.

The transitivity of GLP is at most,see Figure 9b, because of the edges addition mechanism of GLP. More edges are added between old nodes, so that there are more triangles in the network. The IG has more transitivity than PFP because that, while graphs growth, IG has more added edges to old nodes than in PFP. The meaning of the transitivity is that, a network with high transitivity follows strongly the phenomenon: *if A is connected with B, B is connected with C then A and C are also connected*. This phenomenon usually appears in a friendship network such as social networks. Here, we can see that the transitivity values of all networks are very small, under 0.05.

With the purpose to approximate the Internet AS level network, PFP has here the best results. Its clustering coefficient and transitivity are nearly the same values of CAIDA.

3.3.4 Shortest Paths - Diameter

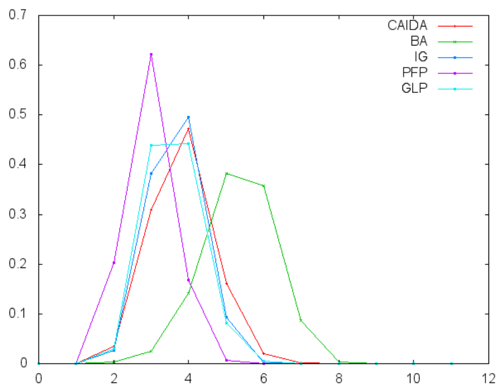


Figure 10: Shortest path lengths distribution

Because of a higher average degree, PFP has shorter shortest path lengths than the others. Its shortest path length lies just between 2 and 4 while this value of CAIDA, IG and GLP is between 3 and 5. Having no richclub phenomenon makes the shortest path lengths of BA longer. The values are between 4 and 7. The results also show the small-world phenomenon, see [9]. It explains how and why packets are transmitted efficiently in our Internet network. The shortest paths lengths in the Internet AS level network is from 3 to 5 while the total number of nodes in our simulation is about 20000.

3.4 Resilience Properties of the generated Topologies

In this second part of simulation, the networks are partition by various types of node removal. The two types of them, random and degree node removals, are already well analyzed with the largest component and the largest biconnected component, see [3] [4] and [6]. But in these studies, the authors do not use all of these metrics, the node removals and the networks that we use. With the following results, we can know the affect of the other types of node removal such as how different the networks behave against them. An important aspect is that, the shortest paths calculation takes more time than the partition based metrics. So sometimes we must remove more than one nodes in each iteration to get an acceptable runtime for our simulation.

3.4.1 The Largest Component

Random Node Removal

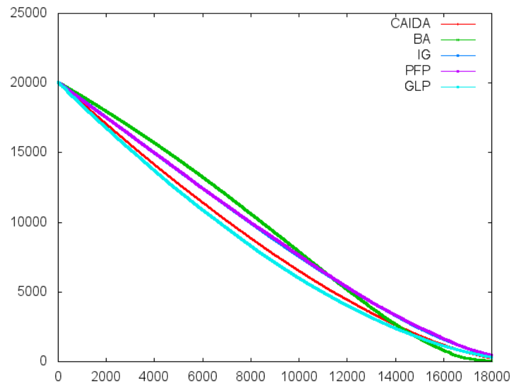
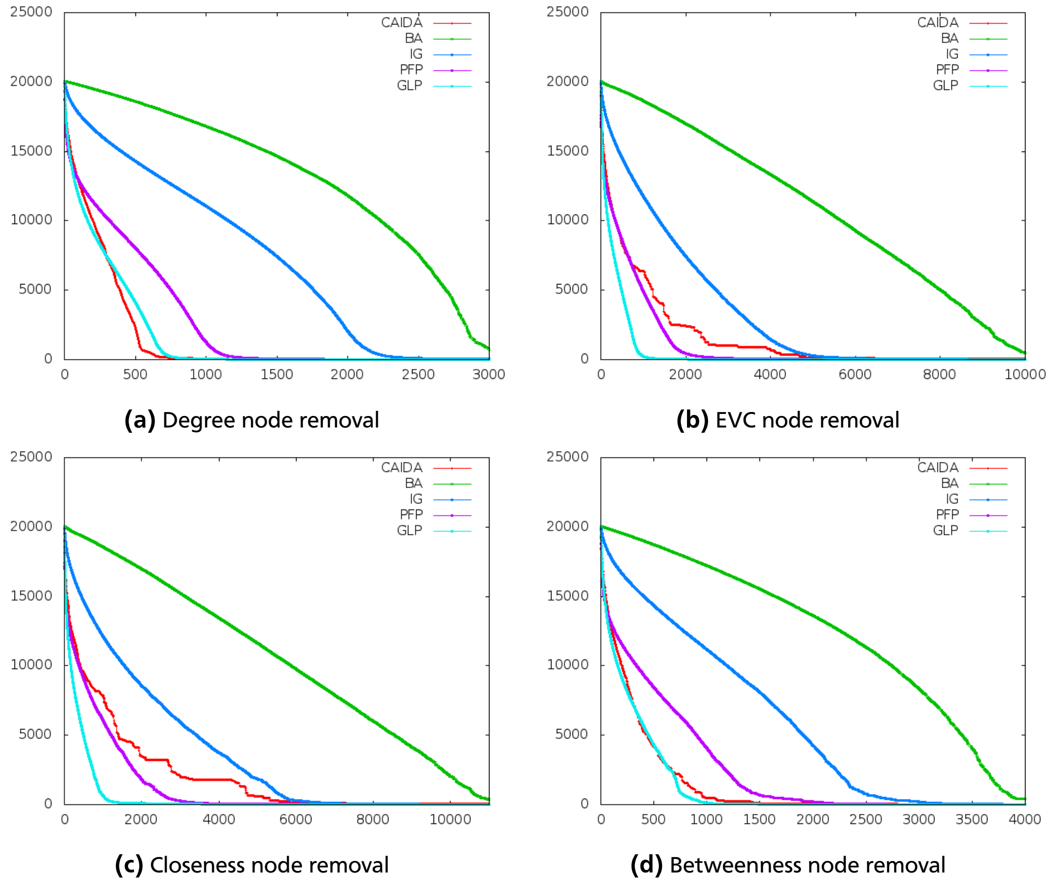


Figure 11: The largest component against random node removal

All of the networks are very stable against the random attack. The largest component's size decreases nearly linear, see Figure 11. As we can see in Figure 7a, the number of high degree nodes is very small. These nodes however play an important role in the connectivity of the network. When we delete nodes randomly, the probabilities that we get these nodes are very small. So the largest component is not much affected. It validates the result in [3]. The size of the largest component decreases just because the number of nodes decreases. At the beginning we have 20000 nodes, and after 12000 removals we have still 5000 nodes in the largest component.

Degree, EVC, Closeness and Betweenness Node Removals

As expected, the networks are quickly partitioned. After about 700 removals, CAIDA has no giant component more and for BA, we must delete only about 3500 nodes, see Figure 12a. We must take into account that the size of the network are about 20000 nodes. There are not many differences between the networks when nodes are randomly removed because they have similar average degree and the random sorting can not choose the vital nodes of a network. But now, we can see that the BA is partitioned more slowly than the others. The reason is that BA has no richclub phenomenon and the clustering coefficient of BA is small, see Figure 8 and Figure 9a. In BA, nodes do not tend to cluster together, it also has no such high degree nodes as the others. On others network, the high degree nodes play an important role on the network. When we delete a high degree nodes, these node are the members of the richclub, we bring more damage to these networks. In this type of attack, CAIDA behaves like GLP. PFP also give us a good result but IG does not.
The shapes of four graph generators have no changes for all types of removal except effective eccentricity node removal, see Figure 12. The difference is the critical points, the point from which the network has no more components. For EVC node removal, it is 10000; for closeness node removal, 11000; betweenness 4500. The way that the CAIDA behaves to these types of attack is very interesting. CAIDA behaves against EVC and Closeness node removal at the beginning as the PFP but then it comes close to IG. In EVC and closeness node removal, CAIDA and PFP critical point is about 1000 and GLP's value is about 500. But against Betweenness and degree node removals, CAIDA and GLP behaviors are nearly the



(a) Degree node removal

(b) EVC node removal

(c) Closeness node removal

(d) Betweenness node removal

Figure 12: The largest component against degree, EVC, closeness and betweenness node removals

same. In this type of removal, CAIDA and PFP critical point is about 400 and PFP's value is 600. We can see that the degree and betweenness centrality are related more to the degree and the shortest paths of the network. In the degree and betweenness removals, the GLP is better than PFP and IG in approximating CAIDA because of it well approximated degree distribution, see Figure 7a, and shortest paths, see Figure 10. In contrast, PFP and IG is much better regarding the richclub connectivity, clustering coefficient and transitivity. This explains why the CAIDA behaves like PFP in the other types of removals. We see later that with the largest biconnected component, CAIDA behaves similarly.

Effective Eccentricity Node Removal

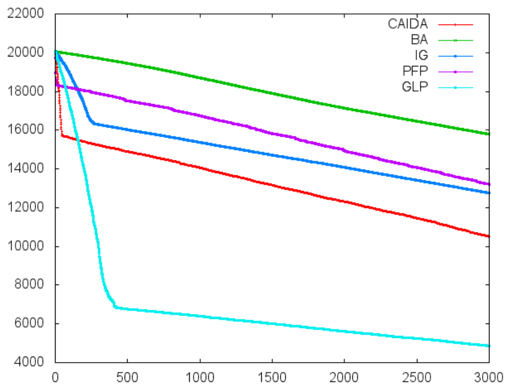


Figure 13: The largest component against effective eccentricity node removal

We have now an interesting phenomenon. At a specific point the damage of node removal suddenly change, for example, until 300 to 400 removals, GLP and IG largest component decrease fast but after that too slow, see Figure 13. An explanation is that at this point the effective eccentricity is not more efficient, the important nodes of the graph can not be found. Perhaps because at this point, the number of fragmentations is too big, so the effective eccentricity is not meaningful anymore overall the network. Another reason is that the index range of the effective eccentricity is too small. At this point, there are many nodes that have the same index. A node is chosen randomly among these nodes. It is not sure whether or not a vital node is selected.

3.4.2 The Largest Biconnected Component

To analyze the connectivity of a network without nodes removals, the largest biconnected component has more meaning than the largest component. Of course all the communication networks are general connected, so its largest component is itself, the whole graph. But we can see from Figure 14 that already at the beginning, the largest biconnected component sizes of the networks are different. Following this, we discuss how these metrics behave during the attack scenarios.

Random Node Removal

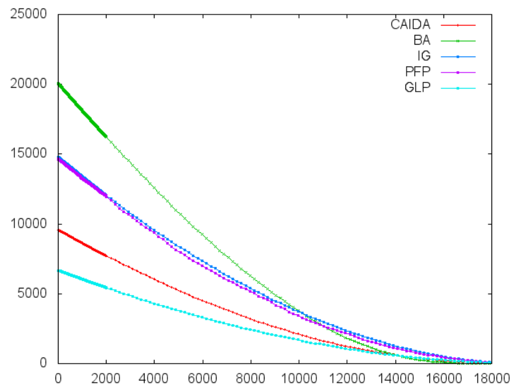


Figure 14: The largest biconnected component against random node removal

At first, we can see that the behaviors of networks are the same as in the largest component, see Figure 14 and Figure 11. The network has about 20000 nodes and after about 17000 nodes was removed, the network is partitioned, see 14. They look different only because at the beginning, the size of their largest biconnected component are different. We must take into account that their critical points are the same and just a little bit lower than the critical point of the largest component. It is what we have expected, as in [6]. At the beginning, BA has a bigger biconnected component than the others. BA is not well clustered, see Figure 9a. This means that the network has a small number of bridge. The new edges in the model are randomly selected with a weak preference attachment, so not all edges tend to be connected to the high degree nodes, therefore it can have more distinct path between two nodes. The CAIDA and GLP have smaller biconnected component than PFP and IG because of the lower average degree. In this simulation, GLP makes a better result in approximating CAIDA.

Degree, EVC, Closeness, Betweenness Node Removals

As in the largest component, against the removals the plotted shapes look like the same, see Figure 15, except the critical points. But the degree and the betweenness node removal are more efficient. The degree node removal partitions the BA graph at about 2800, the betweenness node removal at about 3500. The high degree node has more links connected to other nodes, when we delete it, more paths are affected. Therefore it is the best way to attack the largest biconnected component. A high betweenness centrality node lies on more paths between others nodes, so it is a good choice for that, too. Similar to the result from the largest component, against the degree and betweenness centrality node removals CAIDA behaves like GLP. Against EVC and closeness centrality node removals the PFP is better. This was explained in Section 3.4.1.

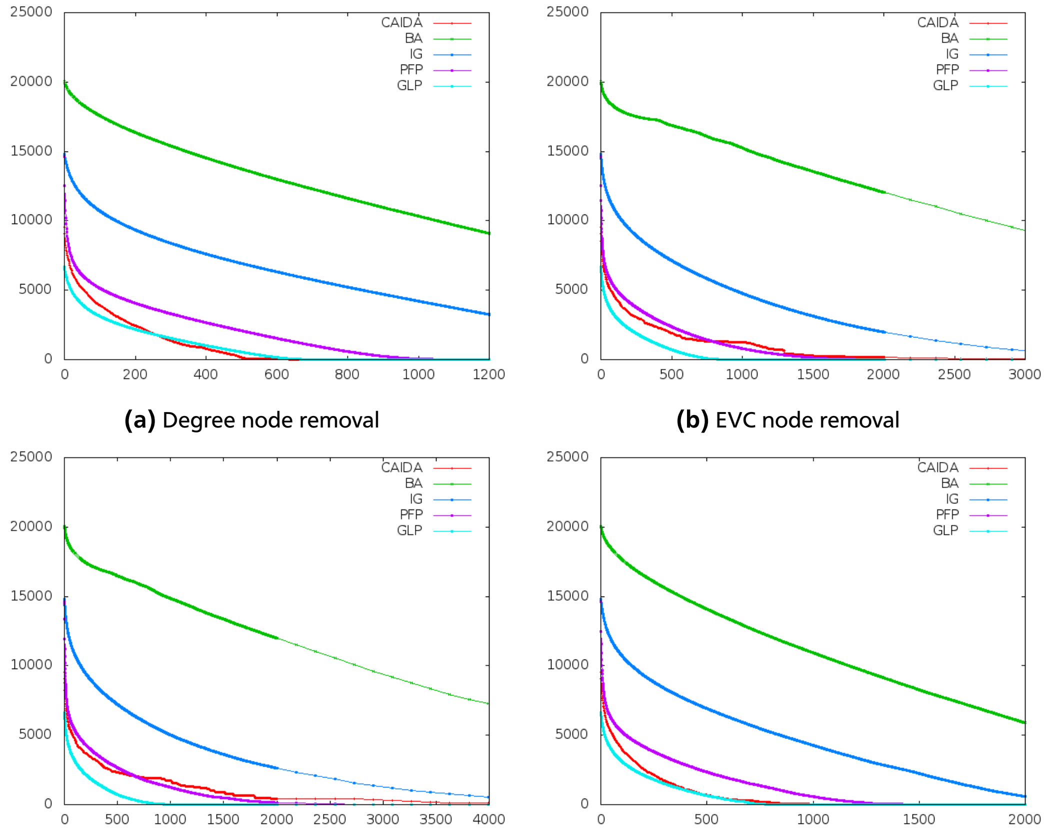
Effective Eccentricity Node Sorting

The plotted shapes looks like what happens when we delete nodes using effective eccentricity in Section 3.4.1, see Figure 16 and Figure 13. The behaviors of the largest biconnected component looks nearly like the largest component. This once again confirms that the biconnected component behaves against the node removals like the largest component.

3.4.3 Diameter - Average Shortest Path Length - Effective Diameter

Random Node Removal

Although we have seen in Figure 11 that the largest component size of all networks are nearly the same, the BA network has here a different shape compared with the others. And we think that the ASPL, AISPL, diameter and the



(a) Degree node removal

(b) EVC node removal

(c) Closeness node removal

(d) Betweenness node removal

Figure 15: The largest biconnected component against degree, EVC, closeness and betweenness node removals

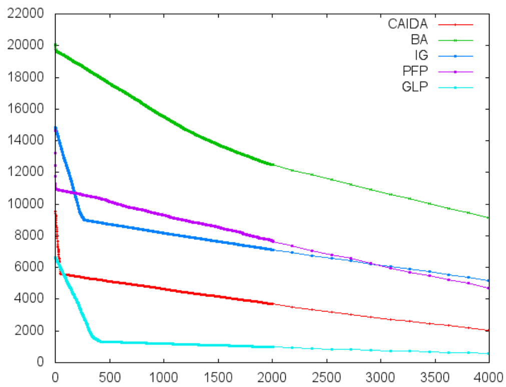
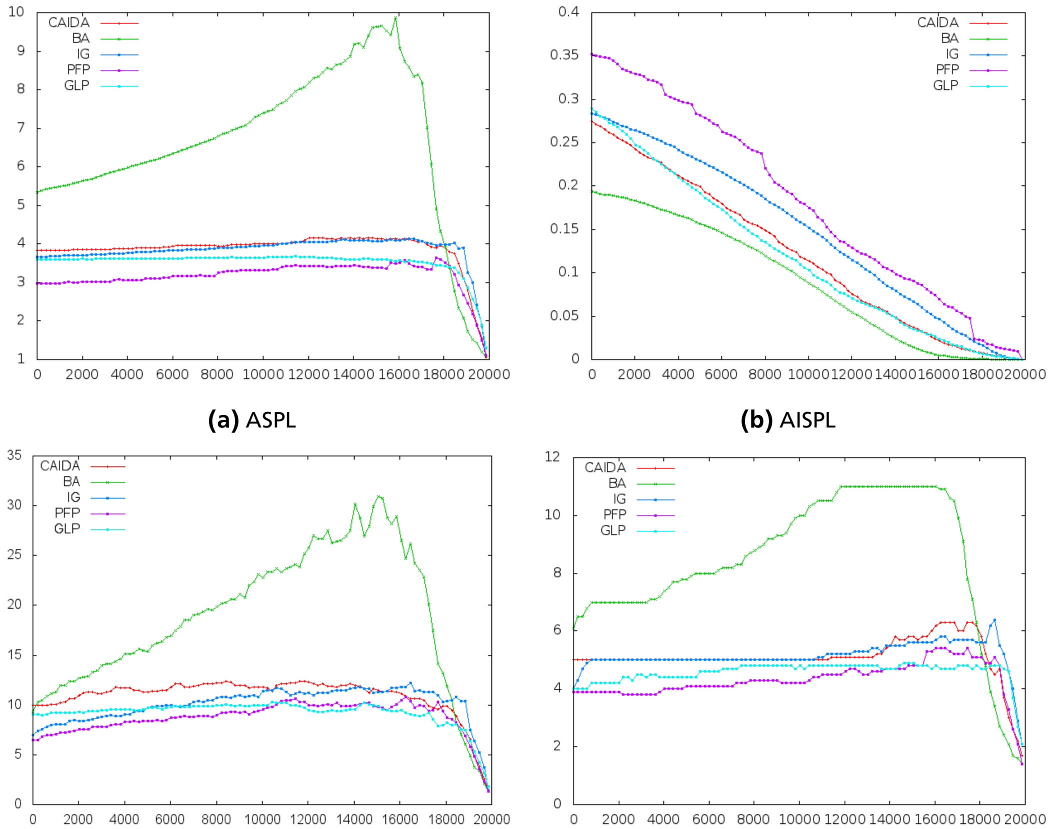


Figure 16: The largest biconnected component against effective eccentricity node removal

effective diameter depend strongly on the largest component. But we must remember that BA does not recreate the richclub phenomenon as the others. The high nodes of the others should be deleted with a small probability, so their largest components maintain the richclub phenomenon. It is different with BA. The richclub phenomenon maintains the shortest paths of the network except BA. Thereby, the shortest paths of BA increase while nodes are removed. This means that, although random removal has a very small impact on all of graphs, this type of attack can be dangerous with some types of network, such as BA.
From Figure 17, it also shows that PFP has always lower shortest path length because of its high average degree. The good approximate of CAIDA in this simulation are IG and GLP.

Degree, Eigenvector, Closeness, Betweenness Node Removals

For the intentional types of attack, the networks behave as expected, see Figure 18 and Figure 19. Because of the similarity of the metrics behaviors in each of the nodes removal degree, EVC, closeness and betweenness, here we



(a) ASPL **(b)** AISPL

(c) Diameter **(d)** Effective diameter

Figure 17: ASPL, AISPL, diameter and effective diameter against random node removal

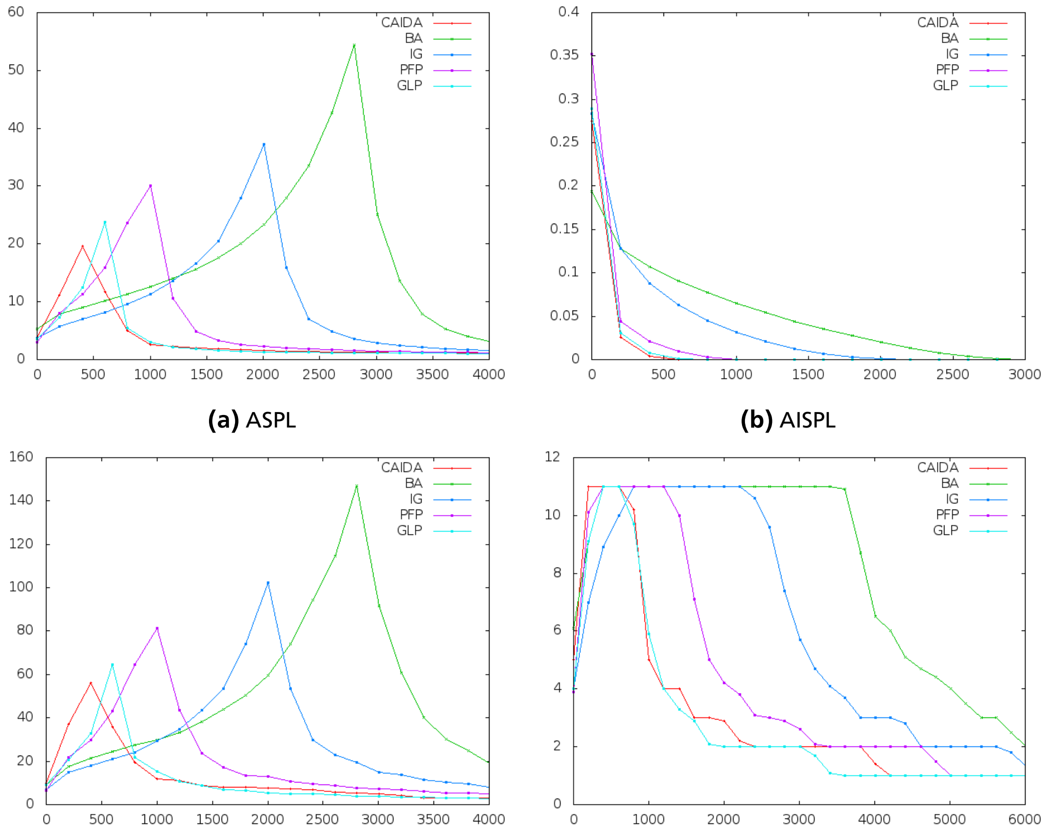
present only the full result of the degree node removal and for others we take only the result from the ASPL. The shortest path lengths increase suddenly at the beginning of node removal. The BA is the most resistant. It takes nearly a double of removals as IG to get the maximal diameter and ASPL. The GLP makes the best approximation to CAIDA, so does the PFP, yet the PFP has larger maximal metric's value than CAIDA and GLP.

The behaviors of the networks against degree node removal are slightly different from the others. The shortest paths get the max value and immediately decrease its value. In contrast, against the other node removals, the shortest paths maintain the max value shortly after they get it. This is also an evidence to show that the degree node removal is the most efficient among the analyzed methods.

Effective Eccentricity Node Removal

In this type of removal, GLP has a higher value of diameter and ASPL, see Figure 20. The reason is, the effective eccentricity find the richclub members of a network very well. The GLP is the network having the best richclub connectivity. So when we delete the high effective eccentricity node, the GLP has more impact. The BA has no richclub phenomenon, it behaves very differently from the others having this phenomenon. Again, PFP is the best resistant network against a node removal in terms of a metric based on the shortest paths. This is clearly understood, because the shortest paths of PFP is lower and PFP has more edges, see Figure 10 and Figure 7c. We can see in Figure 20 that PFP's diameter and ASPL are nearly unchanged during the simulation. But approximating the Internet AS level network PFP does not succeed here and neither do the other networks. Perhaps the reason is the disadvantage of the effective eccentricity ranking, in which we must choose a node from many nodes having the same index.

From the Figure 13, it shows that the networks are partition after about 3000 to 6000 of node removals, at this point the largest component is only a half from the original network. And from the Figure 20a and Figure 20c, we can see that from 0 to about 500, the diameter increases fast. But from 500 to about 6000 the increasing speed is clearly lower. It means that if we combine two types of metrics: partition based metric and path lengths based metric, we have a better view on the network. For example, here it can be indicated that after 6000 removals, the network is broken but during this, the network behaves differently or in other word, the node removal does not always affect the network in a same way, which in fact depends on how many nodes are already removed.



(a) ASPL **(b)** AISPL

(c) Diameter **(d)** Effective diameter

Figure 18: ASPL, AISPL, diameter and effective diameter against degree node removal

3.5 Node Removals Comparision

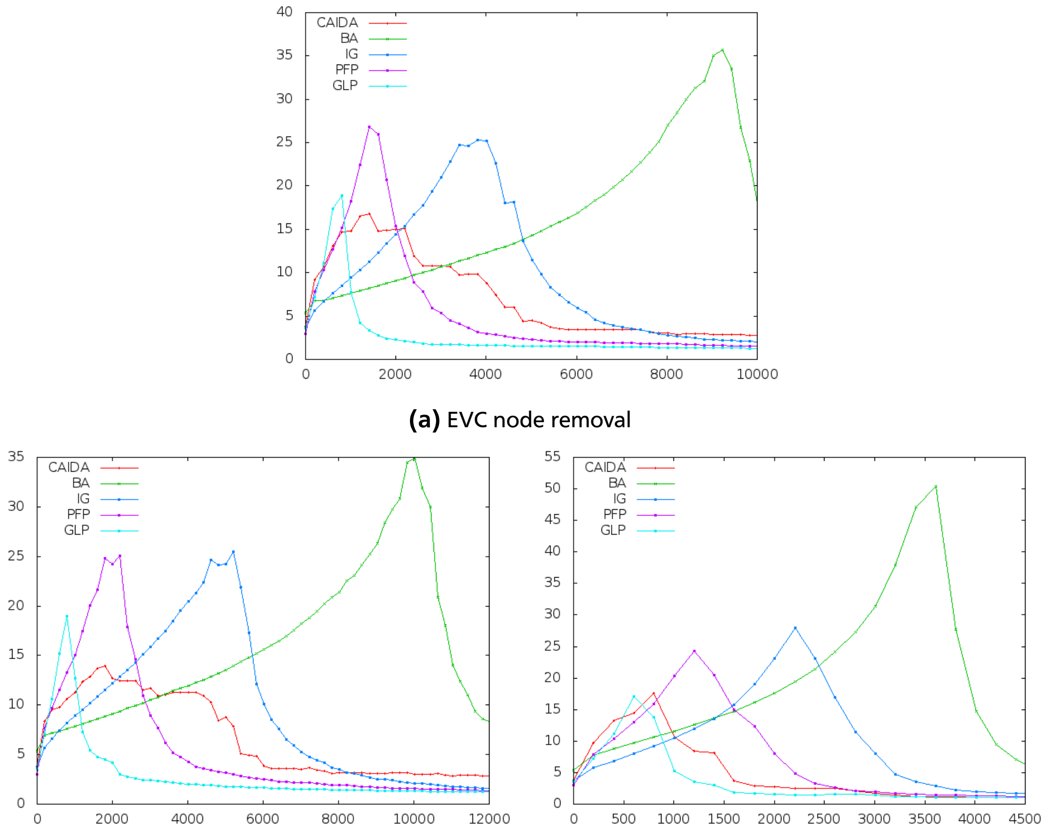
A big problem in network analysis is how we can find vital nodes in a network. In our study, we suppose that a node is important if when it is deleted, the network suffers from a big damage. Now, we discuss and compare the node removals that we have already used in our simulation.

The Largest Component and The Largest Biconnected Component

The first phenomenon is that the degree node removing is always the best choice to partition the network. In fact, we choose the deleting order only at the beginning of the simulation. During the simulation, the graph changes. After some deletions, a vital node can become a normal node and vice versa. But with degree node removing this does not have so much effect, the high degree node remains a high degree node with a large probability. That is the reason why degree node removing is the most efficient removal strategy.
As expected, the random node removing is a poor choice to attack a network. But we can still recognize that with BA network random method is closer to the other methods. The fact is that, BA has no preference attachment. The edges are added randomly while the network grows. The vital nodes in BA are not as much important as in other network models. In both cases, the largest component or the largest biconnected component, with BA the EVC, Closeness and Effective Eccentricity are not much better than random node removing. Only degree and betweenness centrality node removals are always much better than random node removing.
The other phenomenon we can see is that the effective eccentricity centrality node removing is soon no more efficient. After about hundreds step, this method's efficiency is nearly equal to random method.
With CAIDA and IG, we have two groups of efficient removals. The first is degree and betweenness, the second is EVC and Closeness. With PFP and GLP, these two groups however did not so much differently perform.

Diameter - ASPL - Effective Diameter

For these metrics, the degree and betweenness centrality node removals are the best choice too. The differences between these two types of node removal are not very clear when they are applied to PFP and GLP. Besides, on these two networks, the efficiency of closeness centrality and EVC node removals are only a little bit lower than them. As a



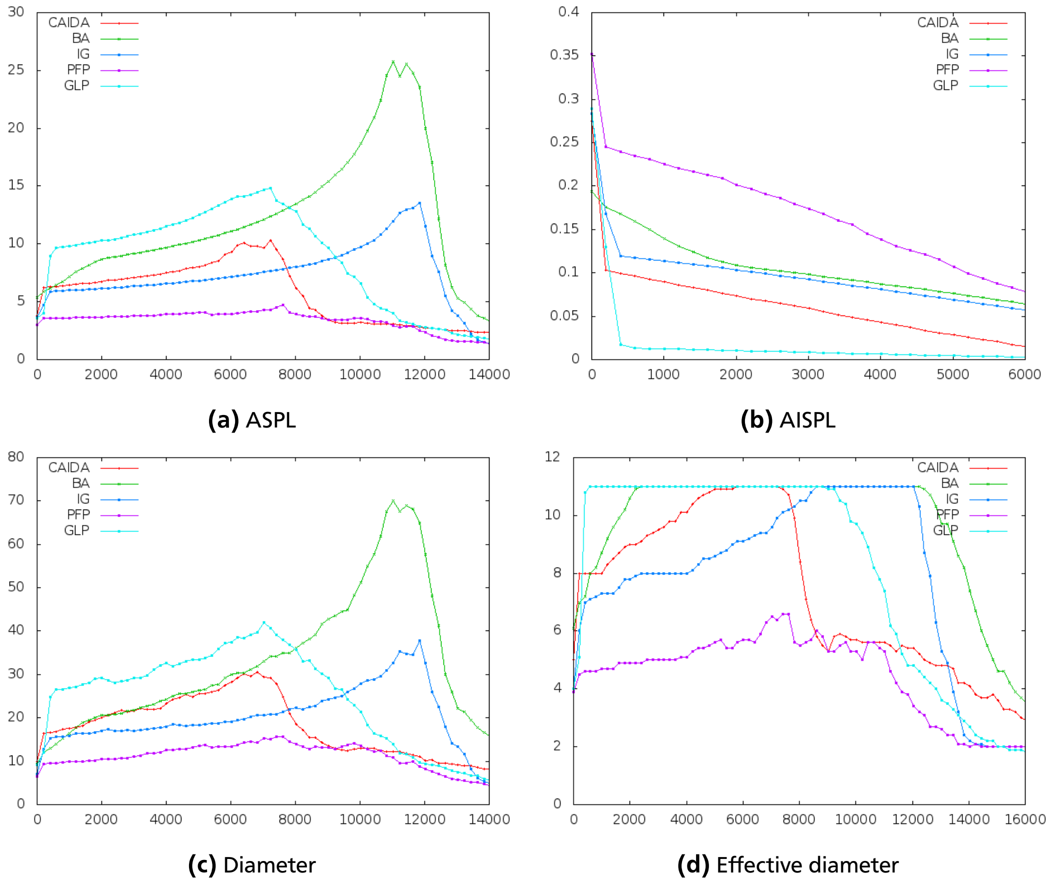
(a) EVC node removal

(b) Closeness node removal

(c) Betweenness node removal

Figure 19: ASPL against EVC, closeness and betweenness node removals

result, the largest component and the largest biconnected component, random and effective eccentricity node removal can not find vital nodes of the network. In some cases, effective eccentricity can do that but only a small number of nodes. The reason is that, the values range of effective eccentricity is small, ranging from one to eleven. The program must randomly be one from many nodes that have the same ranking. For example, to choose one from thousand nodes that have effective eccentricity of five, it can not be guaranteed whether or not it chooses a good one.



(a) ASPL **(b)** AISPL **(c)** Diameter **(d)** Effective diameter

Figure 20: ASPL, AISPL, diameter and effective diameter against effective eccentricity node removal

3.6 Discussion

Both the partition based metric and the path lengths based metric, the two types of resilience metrics, are good for the network resilience analysis. But we should use both of them, not only one. The partition based metric shows us the point where a network is broken and the path lengths based metric shows how differently a metric behaves during that time, from the beginning to the point that the network is broken, see Section 3.4.3

Random node removal does not affect the network resistance, especially with networks that depend on a small subset of vital nodes. It is the reason why the random node removal is not efficient to partition a network. However, for a network that does not follow the richclub phenomenon, random method has a little bit higher probability to choose a vital node. We can see from the Figure 21 that on the BA graph, the differences between random node removal and the others are not as much as on the other graphs. In contrast, the degree and betweenness node removals are very efficient. Removing a node with high degree or high betweenness centrality means that this node on many shortest paths in the network affect the connectivity and the shortest path length of the network, so it's not surprising that they are the most effective node removals. Eigenvector centrality index is assigned to all nodes based on the concept that connections to high-scoring nodes contribute more to the score of the node than connections to low-scoring nodes. So a low degree node can be in top rank of closeness centrality if it is connected to some high degree nodes. Effective Eccentricity has problem with rank index range (too small). At the beginning, removing a node using effective eccentricity has a nearly large impact as the degree and betweenness centrality node removal, but after a small number of removals, it does not work well anymore. It can be explained that its index range is too small. If we want to use this type of node sorting to remove nodes in a network, we should recompute the effective eccentricities of nodes at the beginning of each iteration. Or even more, if we can find a criterion to choose a node from many nodes that have the same effective eccentricity index, for example, we take the node with a higher degree, then the effective eccentricity node removal should works better.

BA's properties are not good in approximation with CAIDA. Despite the accuracy in average degree, BA has too small values of maximal degree, clustering coefficient and transitivity. The reason is BA was designed only to generate random scale-free networks. Its edges adding mechanism already has a reference attachment but this attachment is not strong as in IG, PFP and GLP. Besides, while a network grows, not only edges between the new node and the old network but also

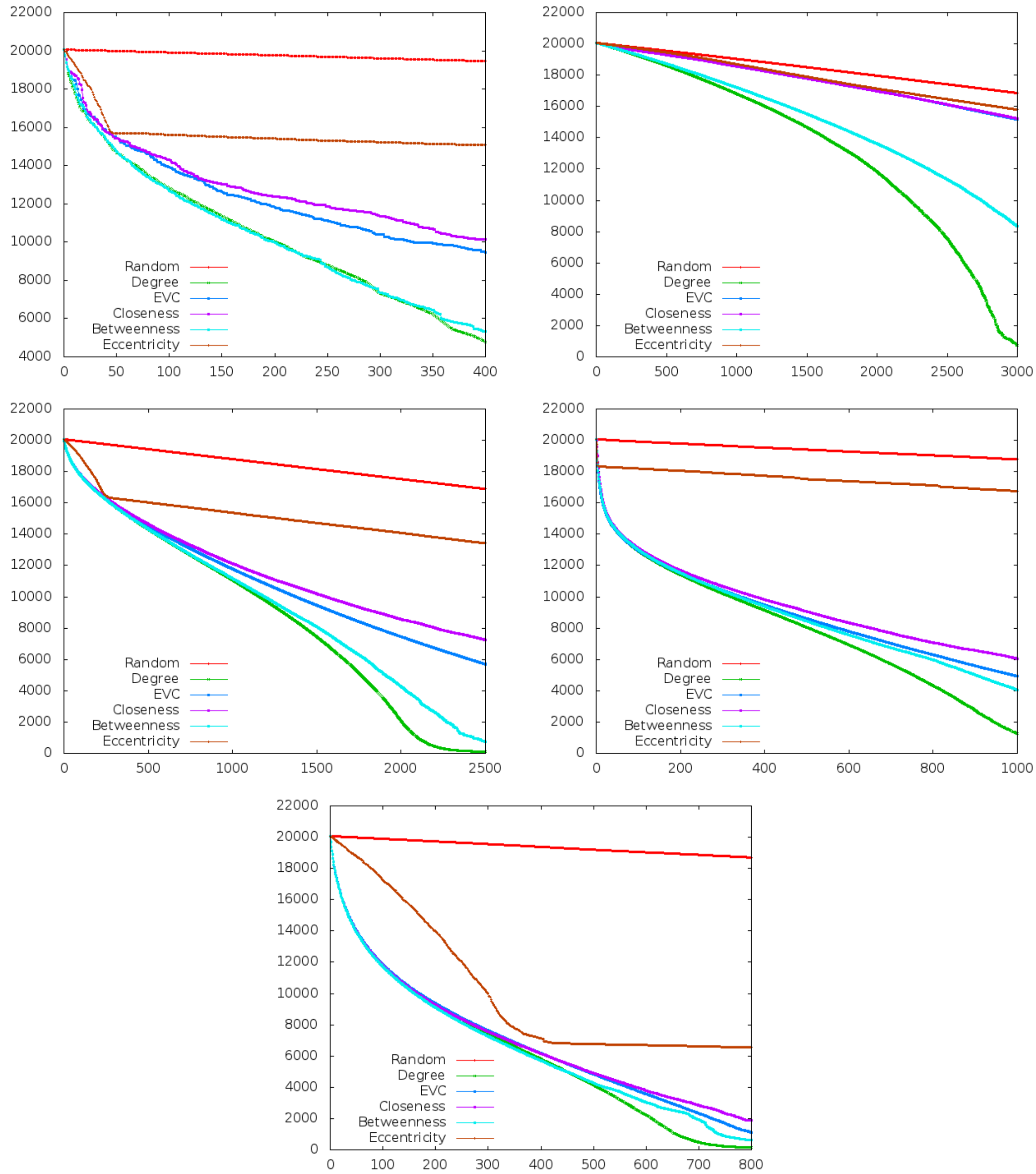


Figure 21: Node removals comparison for the largest component

new edges between the old nodes appear. This edges adding mechanism make the max degree, the richclub connectivity, the clustering coefficient and the transitivity values higher.
GLP approximate the degree distribution of CAIDA well, you can see in Figure 10 that the shortest path lengths of GLP are closer to CAIDA than PFP, and in 7 that the average degree and the maximal degree from GLP is close to CAIDA. But its clustering coefficient is too small and the transitivity is too big compared with CAIDA. Also the richclub connectivity of GLP is bigger than the others. Because of the edges adding of its growth mechanism, more triangles appear while the network grows. With these edges, the richclub members of GLP are also well connected together. In contrast, PFP recreates the clustering coefficient, the transitivity and the richclub connectivity are better in terms of approximating CAIDA.
From the results, we can see that BA is the most resistant of all. While nodes are being removed, BA maintains its largest component and largest biconnected component well. The reason is, BA has no richclub phenomenon. The maximal degree of BA is only 312 (from the BA2012 graph). The low degree nodes are well connected together. The network connectivity does not depend on a specific set of high degree nodes. The shortest paths do not necessarily include some high degree nodes. So when the nodes are deleted, the impact is not as much as on the other networks.
None of the network generators recreates an exact richclub connectivity property compared with CAIDA. The first 12 high degree nodes of all networks except BA make a full graph together. But after that the high degree nodes of CAIDA are not well connected as the others.

4 Summary - Conclusions - Outlook

The purpose of the current study is to determine the impacts of the various types of attack scenario on many types network generator and also on the Internet AS level by measuring the resilience metrics value after each node removal.

To analyze the connectivity of a network without nodes removing, the largest biconnected component has more meaning than the largest component. When we want to analyze the behavior of networks against some attacks, these two types of metric would be good candidates. We can see that, the results of the network generators are clearly different, which can be used to compare networks. The three metrics, Diameter, ASPL and and AISPL, give us nearly the same results. AISPL is just an inverse version of ASPL. For simulation purpose, we can use just one of this metrics. No additional information is added by the effective diameter. At the beginning, the metrics values of the different networks are not clearly distinct. The later differences do not give us much information because at these points, the network properties and topologies are changed or the network is already broken. If we take the runtime into account, the largest component and the largest biconnected component have advantages. For the calculation of ASPL, AISPL and the Diameter, we must compute all shortest paths in a network. Yet, choosing a metric apparently depends on the analysis purpose. But using both of these two types of metric is the best way to have a better view on a network.

The largest component and the largest biconnected component behaviors against the node removals are similar. Their critical points also have nearly the same values. The ASPL, AIPL and diameter are all based on the shortest paths of the network. Because of their very similar results, taking one of this is also enough for the network analysis.

Various types of node removals have different impacts on the metrics. They make the largest component and the largest biconnected component smaller. Besides they make the shortest paths longer as well as the ASPL, the diameter, the effective diameter. Most notably, the degree and the betweenness centrality removals bring more damages to the network. The closeness centrality and the EVC centrality node removals are less efficient but still better than effective eccentricity and random node removal. The random node removal has nearly no impact on the shortest paths and when it is applied to a network, the largest component decreases only because the number of nodes in network also decreases.

BA bahaves differently from CAIDA and the other networks. For all types of metrics and attacks, overall it is always the most resistant network. IG is already better than BA but its next generation PFP is closer to CAIDA. The best approximate network generators for CAIDA are PFP and GLP. The degree distribution, the clustering coefficient, the transitivity, the richclub connectivity and their behavior against the node removals are closer to CAIDA than the others two, BA and IG. GLP is better than PFP in degree distribution and shortest path lengths approximation, but the structure properties, such as the clustering coefficient, the transitivity and the richclub connectivity, of PFP is closer to CAIDA than GLP. The behavior of PFP and GLP against the node removals have also some differences in terms of comparing to CAIDA. CAIDA behaves halfway as PFP and halfway as GLP, it depends on which metrics and node removals we use.

In our simulation, we have run the node sorting only at the beginning of each node removal. In the future, we want to recalculate a sorting after each removal and we hope that it makes the node removal more efficient. To overcome the disadvantage of index range of the effective eccentricity, when choosing a node from many nodes with the same ranking index, we should use another strategy instead of random method. For example, from many nodes that have the same metric value, we takes the node with a higher degree. There are some new Internet topology generators, such as PARG, but we have not implemented. In the future, more network generators and possibly some more meaningful metrics are used in our study.

References

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[2] T. Bu and D. Towsley. On distinguishing between internet power law topology generators. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 638 – 647 vol.2, 2002.

[3] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, Nov 2000.

[4] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86:3682–3685, Apr 2001.

[5] P. Erdos and A. Rényi. On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61, 1960.

[6] M. E. J. Newman and G. Ghoshal. Bicomponents and the robustness of networks to failure. *Phys. Rev. Lett.*, 100:138701, Mar 2008.

[7] C. R. Palmer, G. Siganos, M. Faloutsos, and C. Faloutsos. The connectivity and fault-tolerance of the internet topology. In *Workshop on Network-Related Data Management (NRDM 2001), In cooperation with ACM SIGMOD/PODS*, 2001.

[8] M. Piraveenan, M. Prokopenko, and A. Y. Zomaya. Local assortativity and growth of Internet. *The European Physical Journal B - Condensed Matter and Complex Systems*, 70(2):275–285–285, July 2009.

[9] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, June 1998.

[10] S. Zhou. Towards modelling the internet topology - the interactive growth model. In *in Proceedings of the 18th International Teletraffic Congress (ITC18*, page 0303029, 2003.

[11] S. Zhou and R. Mondragon. The rich-club phenomenon in the internet topology. *Communications Letters, IEEE*, 8(3):180 – 182, march 2004.

[12] S. Zhou and R. J. Mondragón. Accurately modeling the internet topology. *Phys. Rev. E*, 70:066108, Dec 2004.