University of Twente

Faculty of Electrical Engineering, Mathematics and Informatics

Bachelor of Science thesis

Error tolerance analysis of the Telefónica de España optical fibre network

Pieter van Wijngaarden

Supervisors:
Telefónica Investigación y Desarrollo, Madrid, Spain
Mary Luz Mouronte Lopez
Victor Feliu Edo

University of Twente, Enschede, The Netherlands
Jan Schut

Enschede
September 2008

BACHELOR'S THESIS

Error tolerance analysis of the Telefónica de España optical fibre network

Pieter van Wijngaarden
0065536

September 2008

University of Twente
Jan Schut

Telefónica Investigación y Desarrollo
Mary Luz Mouronte Lopez
Victor Feliu Edo

Preface

This publication is a combination of two academical papers, discussing the network characteristics of the Telefónica optical fibre network. The first paper discusses basic network characteristics and general network theory, the second one discusses the robustness of the network (from a topological point of view) regarding node and link failures, taking several parameters into consideration that are important to find out if the network is able to cope with for example directed attacks on hardware, what happens on the average if a link fails etc. This information can help identifying weak points in the network. They are both written with support (in knowledge and in resources) from Telefónica Investigación y Desarrollo (Research & Development) in Madrid, Spain. Although the process has taken a long time, I hope the result is satisfactory and via this way I would like to thank Telefónica I+D for the opportunities, and especially my two supervisors; Mary Luz Mouronte Lopez and Victor Feliu Edo for their help and support during my Bachelor’s project.

Analysis of the Telefónica España optical fibre network complexity

Mary Luz Mouronte-López
Victor Feliu Edo
Pieter van Wijngaarden

June, 2007
Telefónica Investigación y Desarrollo, Emilio Vargas 6, 28043 Madrid

1. Abstract
Complex systems and networks are growing in importance from a scientific point of view. We analyzed the Telefónica de España optical fibre network in order to find out whether it behaves as a complex network, so that this model can be used to identify weak points in the network to prevent service disruption. We found out that the optical fibre network shows some scale-free network properties, but also has its practical limits and geographical constraints, and therefore applying the scale-free model to this network can be done but its peculiarities must be taken into consideration.

2. Introduction
Last few decades, the understanding of complex systems has been described as the key to further scientific advancement in many different fields. Expectations are high, because a lot of real-world phenomena indeed can be described using these complex systems. Examples are an ant nest, national economies or computer networks. The notion that, in a complex system, the whole is more than the sum of the individual parts is fascinating and its underlying concepts are not yet totally understood.

This article presents the key concepts of network modelling and network complexity and discusses and analyses the complexity of the Telefónica España optical fibre network in order to provide better insight on how this network behaves and is structured topologically. This information will aid us in preventing service disruption, improving the robustness and error-tolerance of the network, and obtaining knowledge about in what ways the network behaviour can be used to our advantage.

The remainder of this introduction will explain some basic principles about complexity, network modelling and network complexity. In Section 3 we will discuss the Telefónica España optical fibre network, present results of our analysis of this network on a few different scales and compare these results to the scale-free network model to see if we can describe our network as such. Section 4 contains our conclusions obtained from this analysis.

2.1. Complex systems
Definitions of a complex system still differ from source to source, since there is no generally accepted one. The best fitting definition is

A complex system is a system or network of relatively simple components with no central control, in which emergent complex behaviour is exhibited [1]

This means that out of a complex system can *emerge* unpredictable behaviour, despite being built up of really simple parts. A good example of this is a small toy rotator (Figure 1) or an ant's nest which is capable of total self-organization without any guidance or centralized control [2].

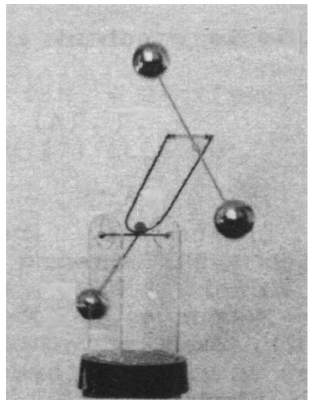

Figure 1: A toy rotator with totally unpredictable behaviour.

2.2. Networks

A fair deal of the real-world phenomena which exhibit complex behaviour can also be explained in terms of networks. Examples are gene regulatory networks, the human brain and the World-Wide-Web. For the first time in history these complex systems and networks can be accurately modelled and analysed with modern technologies. The question we want to answer in this paper is if the Telefónica España optical fibre network behaves as a complex network as well. But in order to understand how complex systems are modelled in network terms, we must first discuss some basics from graph theory and network analysis.

A network can be simply described as two mathematical sets: a set of vertices (points, nodes) and edges (links), defining which vertices are connected to each other and which ones are not. These links can be directed or undirected (the direction indicates in which way the two connected vertices interact, or in which way information can flow) and weighted or unweighted. Figure 2 is an example of a network.

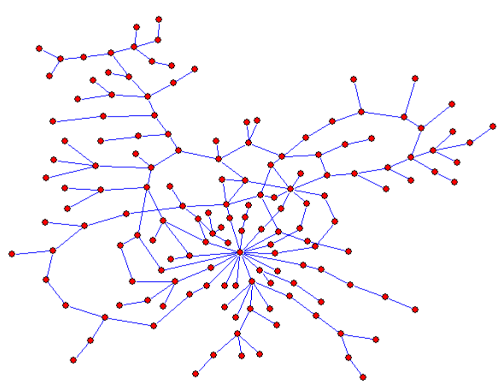

Figure 2: Example of a network. This is a graph representation of the optical fibre network from the Spanish province Teruel (however not a geographical representation).

Although in the example all nodes are connected to each other, a graph (the mathematical representation of a network) doesn't always have to be connected. Also link direction and weight can mean different things depending on what real-world phenomenon the network models; usually link weights indicate the importance of a link and have something to do with the amount of traffic or information flow through that link. An example of an unweighted network is a gene-regulatory network, where a node is a gene, which is controlled or activated by a certain protein (the link or edge). The brain is a giant network of neurons linked by synapses. A weighted network can be a network of highways between cities where the weights indicate the amount of traffic per day. There are quite a lot of characteristics that can be obtained from a network through network analysis; we will describe the most important ones and the relevant terminology here.

If there is a link between two nodes, they are *neighbours*. The *degree* of a node is the number of neighbours the node has. The *degree distribution* of a network defines the relation between degree $\boldsymbol{k}$ and the probability P($\boldsymbol{k}$) that a random node in the network has degree $\boldsymbol{k}$. The *shortest path* or *geodesic* between a node pair (a, b) is the shortest route that can be found in the network from a to b. Analogue to the degree distribution a *shortest path distribution* can be defined as well, this is the relationship between a distance $\boldsymbol{k}$ and the probability that a random node pair has this shortest distance. The *diameter* of a network is the length of the shortest path between the two most distant nodes (i.e. the longest shortest path).

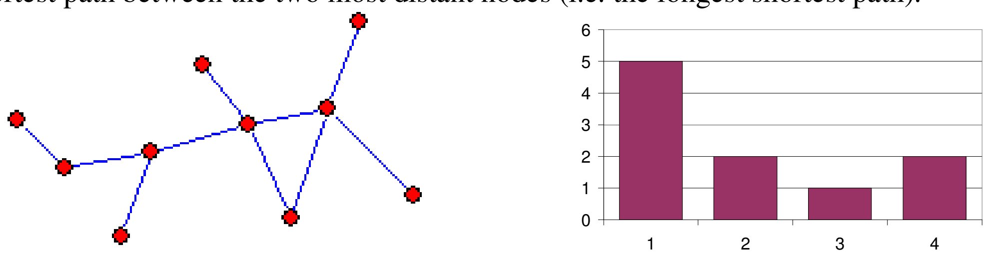

Figure 3: A very small network with its degree distribution (in absolute numbers instead of probabilities). This network has a diameter of 5.

Another important network characteristic is the *clustering coefficient distribution*. The clustering of a node can be viewed as the level of connectedness of the node's neighbours (the number of links between the node's neighbours divided by the number of possible links). If every neighbour of node A is, apart from being connected to A, also connected all the other neighbours of A, then the *clustering coefficient* of A is 1. If the clustering in a network is high, a node is more likely to be connected to its neighbours´ neighbour than to another random node in the network. Clustering is a phenomenon that occurs in a wide variety of real-world networks, for example in almost every social network there are many groups of people who know everybody in that group, but (much) less people outside of that group. This is a typical example of clustering.

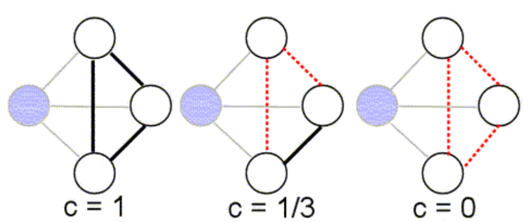

Figure 4: Three different situations with their clustering coefficients (of the blue node) [3]

It can be useful to classify a network to help model it in a way which is closest to the real world. There are a few different network models, all of these exhibit different (above described) characteristics, and thus, network behaviour.

2.2.1 Random

This kind of network model (also called the Erdös-Rényi model) was introduced by Paul Erdös and Alfréd Rényi in 1959. This model creates a network G(n, p) with n nodes by taking all node pairs and randomly draw a link between all node pairs with probability p. This kind of network structure appears to be quite different from most real-world networks; however they are interesting to study from a mathematical point of view. The degree distribution of these kind of networks is Gaussian (normal) distributed. The average clustering coefficient of this kind of network is p, which is also much smaller than clustering coefficients in real-world networks.

2.2.2 Small-world

Another network model is the small-world model, introduced in 1998 by Duncan Watts and Steven Strogatz. The key characteristic of a small-world network is that the average shortest-path length is incredibly small compared to the number of nodes. In a regular network the average shortest path length grows along with the number of nodes, however this is not the case in small-world networks. A number of real world networks exhibit this characteristic. The popular term 'six degrees of separation' refers to the idea that everybody in this world is connected to everybody else with an average of six persons in between. According to small-world network theory, the average path lengte in a small-world network is around *log(n)*, where *n* is the number of nodes in the network. The model that Watts and Strogatz used is as follows: start with a network with arbitrary n nodes in a ring-like structure where every node is connected to a few of its nearest neighbours. Then, for every link, rewire one end of the link (with a small probability) to a randomly selected other node in the network. If the probability p is rightly chosen, this results in a network where the majority of nodes have most of their links to their direct neighbours (which indicates clustering), but a few nodes in every cluster have links to other clusters in the network. This means that the shortest path between two randomly chosen nodes is probably very small, smaller than in a random network. However, the small-world model still shows significant differences with the most real-world networks, the degree distribution usually doesn't match.

2.2.3 Scale-free

The third and last well-known network model is the scale-free model, Albert-László Barabási and Réka Albert developed this model in 1999 as another attempt to describe the characteristics of real-world networks in a more precise way. The model is called scale-free because there is a relatively large number of nodes with a fairly large degree, much more than you would expect in a random network. Hence, the degree distribution of this kind of network follows a *power-law* distribution, which means that P(k), the chance of finding a node in the network with degree k, can be roughly described with $k^{-\gamma}$, where γ is between 2 and 3. Scale-free networks also show the small-world property, i.e. the average shortest path length is a lot smaller than expected in a random network with the same number of nodes.

The scale-free network is the model closest to many real-world networks. Barabási and Albert used a web crawler to investigate the structure of the World Wide Web (in this structure websites are nodes and hyperlinks to other websites are the links), and during this study they came up with the scale-free model.
One characteristic quite common for real-world scale-free networks is the so-called *hierarchical clustering*. This means that clustering increases in the 'lower' parts of the network, and hubs (the nodes in the network with a high degree) in general have a lower

clustering coefficient than the nodes with a low degree. Compared to random networks, the clustering of a scale-free network is approximately 5 times higher, although it is still decreasing (following a power-law function) as the number of nodes increases, which is different from the small-world model where the clustering coefficient is independent from the number of nodes in the network.

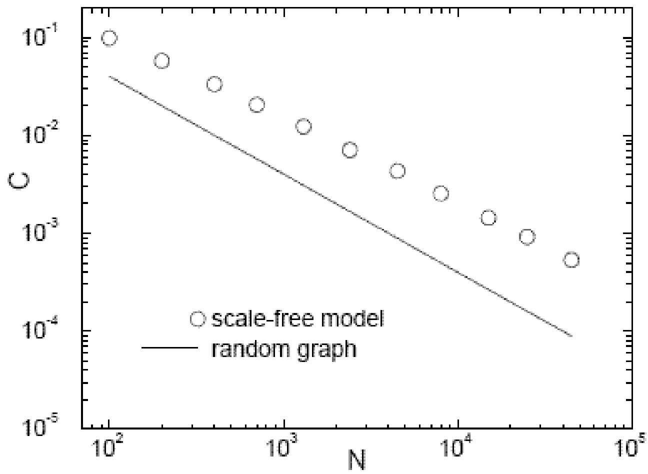

Figure 5: The average clustering coefficient (clustering C versus number of nodes N) from a scale-free network (with average degree k = 4) compared to a random graph. The clustering is approximately 5 times higher, and this factor increases slightly as the number of nodes increases. (Figure 24 from [4])

Because of the typical topology that the majority of the nodes are connected to another node with a high degree (these nodes are called *hubs*), a scale-free network is quite robust and very resistant to random, irregular node failure. However, this comes at the price of having increased vulnerability against directed attacks towards the nodes with the highest degree, because if those nodes fail the entire network falls apart very quickly. [4]

The 'generative model' for this kind of network is based on two principles, namely *growth* and *preferential attachment.* This means that while the network grows, a new node is more likely to become connected to an already highly connected node, instead of randomly being connected to any node. Formally described, the growth model is as follows: first, start with a small network of initial nodes. Then for every step, add a new node to the network and connect it to *m* other nodes (*m* is predefined). The probability that the new node becomes connected to another node already present in the network is proportional to the degree of that node. Hence, the nodes with the highest degree obtain the most new links. It can be mathematically shown that this growth model results in a power-law degree distribution (see Figure 6). [5] [6]

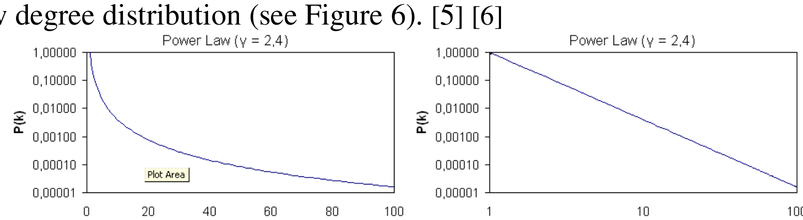

Figure 6: Two graphs of a power-law function (left: vertical scale logarithmic, right: both scales logarithmic)

3. The Telefónica España optical fibre network

As was said before, the aim of this research was to find out whether the Spanish optical fibre network from Telefónica also behaves as a scale-free network, and if not, which other mathematical model is best suited for it.

The Telefónica España optical fibre network is a fairly large network, with approximately 90.000 nodes and more than a 100.000 links. The network undergoes daily changes, so tomorrow the number of nodes may be a bit larger (although not much). The network can be divided into provinces, which resemble the 50 provinces of Spain, including a few links to other countries that are abstracted as another province. These provinces put a large geographical constraint on the network, since more than 60% of the nodes are located in the 10 largest provinces (with Madrid, Barcelona and Sevilla being the largest ones). The questions we are trying to answer are:

-Do the geographical constraints have influence on the network characteristics?
-What mathematical network model fits the best to our network?

Our hypothesis is that the network behaves like a scale-free network, such as the World Wide Web, the Internet and other similar networks. Therefore our focus will lie on the comparison to the scale-free model. We also investigate in which way the structure of the entire network is compared to that of the individual provinces (the biggest ones as well as the smaller ones).

3.1 Telefónica de España network analysis

Through extensive analysis of our optical fibre network, with the help of a network analysis program called Pajek (for the geodesic distributions, [7]) and with self-written analysis software in Java, we obtained the relevant information about the network. We first started analyzing just a few provinces, from small to big, because we also wanted to know in what way the behaviour of the separate provinces relates to the behaviour of the entire network.

3.1.1 Separate provinces

The separate provinces we analyzed are Teruel, Zaragoza, Madrid and Barcelona. Teruel is one of the smallest provinces of Spain (considering the number of nodes in that province, not geographically), while Zaragoza, Madrid & Barcelona belong to the 10 biggest provinces. Remarkably, all 4 provinces, however greatly different in size, show more or less the same kind of behaviour.

To give an idea of the scale and size of these networks, we give the following two pictures; they are drawings from the Teruel and Zaragoza provinces. Since Barcelona and Madrid are even 5 to 10 times bigger than Zaragoza, it is impossible to make a drawing of these networks on a single page. As can be seen, it is also very hard (if not impossible) to show the Zaragoza network without any crossing links.

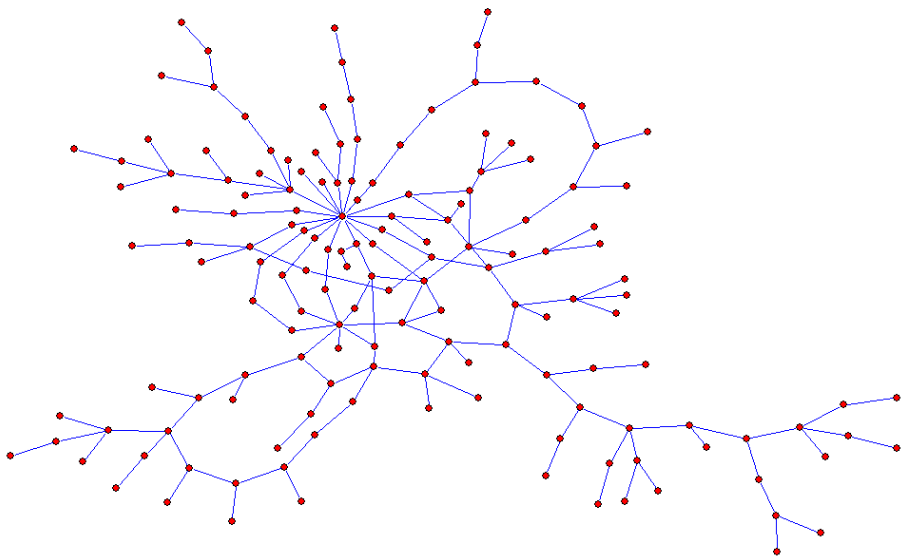

Figure 7: The Teruel province optical fibre network (figure obtained with Pajek).

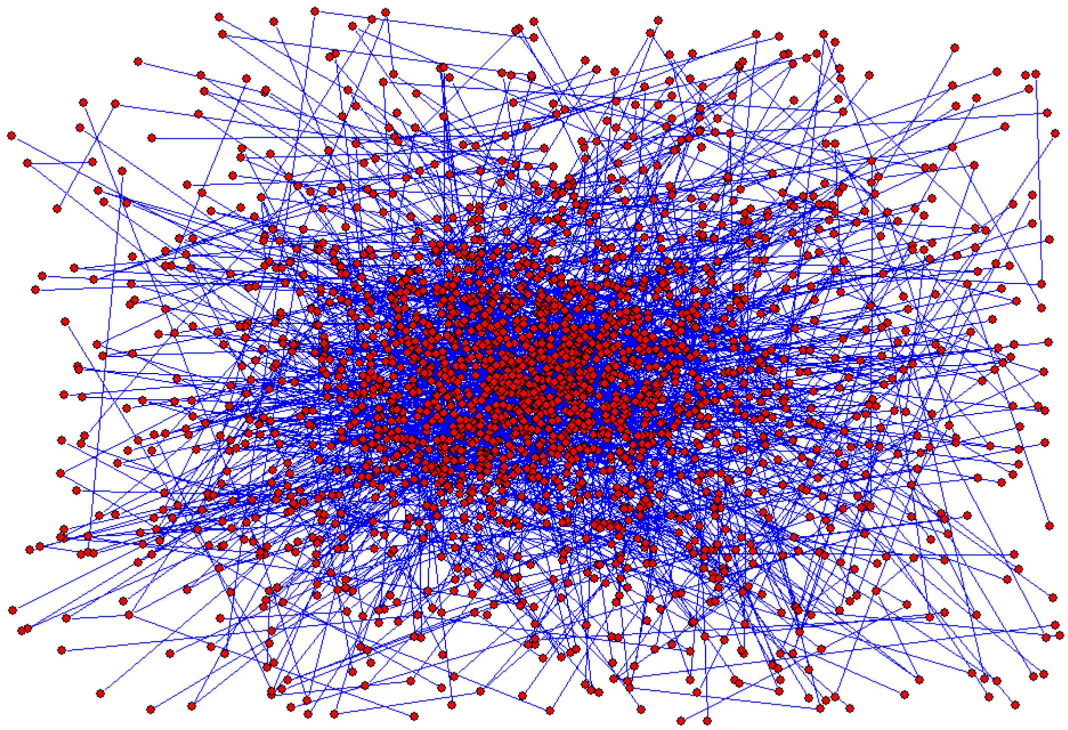

Figure 8: The Zaragoza province optical fibre network (figure obtained with Pajek).

Now, a little table with a few global characteristics is presented. *C* is the average clustering coefficient, *k* the average degree, and *l* the average shortest path length of the network.

Name	# nodes	*C*	*k*	diameter	*l*
Teruel	159	0,0079665	2,1383	19	7.73154
Zaragoza	2287	0.0066712	2.0017	75	29.33266
Barcelona	9057	0.0102759	2.1252	31	11.04942
Madrid	20846	0.0247895	2.1721	53	14.85322

As can be clearly seen, the provinces are quite alike, Teruel, Barcelona and Madrid all show similar properties (with respect to their number of nodes). The only exception is Zaragoza, this province shows a relatively small degree (almost 2), which implies also a smaller clustering coefficient (even smaller than Teruel) and a quite large diameter. Zaragoza is cannot really be called a scale-free (sub)network. Perhaps this can be explained by suggesting that Zaragoza is a sort of long-stretched traffic relay province between Madrid and Barcelona, part of the traffic between Madrid and Barcelona goes through Zaragoza. However we only have topological network information at our disposal, and no traffic statistics, so we cannot confirm this with hard evidence. However the network topology suggests that it would only be a part of the total amount of traffic; analysis of the number of links between the various provinces showed that there are 4477 links between nodes from the province Barcelona and Madrid, while between Madrid and Zaragoza there are 631 and between Barcelona and Zaragoza there are 494 links.

Below here, we present figures from the analysis of the 4 different provinces.

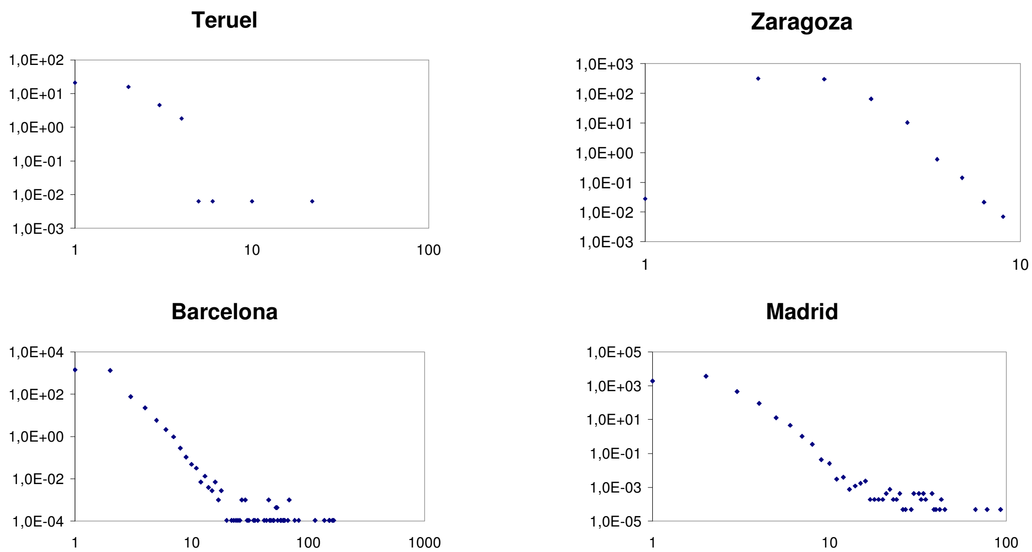

*Figure 9: Degree distributions from Teruel, Zaragoza, Barcelona and Madrid. On the horizontal axis, the degree k is plotted, on the vertical axis, N * P(k) (the number of nodes with degree k multiplied by the chance that a random node has degree k).Especially Barcelona and Madrid clearly show a scale-free type degree distribution.*

As can be seen, all provinces (except from Teruel which is relatively small) show a degree distribution similar to a scale-free network. It is a bit strange that Zaragoza, being almost 15 times bigger than Teruel, has a maximum degree which is smaller than the maximum degree of Teruel. The same with Barcelona and Madrid: Madrid is twice as big as Barcelona but has a smaller maximum degree. These subnetworks seem to show scale-free characteristics, but have their curiosities.

The following 4 graphs show the clustering coefficient distributions for the four provinces.

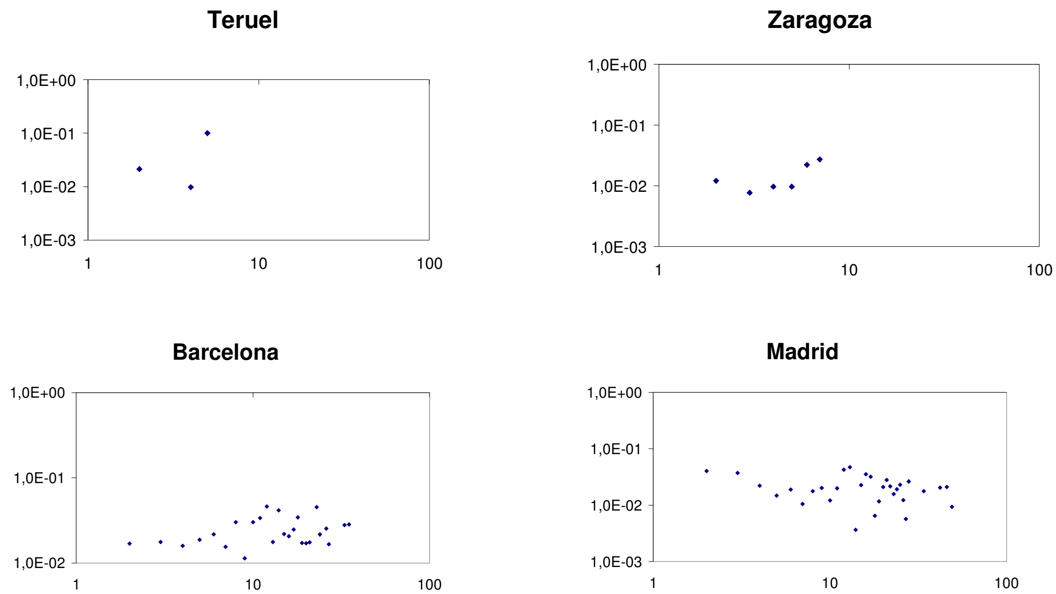

Figure 10: Clustering coefficient distributions from the four provinces. Horizontal axis: the degree, vertical axis: the average clustering coefficient of all nodes with that degree.

The clustering coefficients of Teruel can be kept out of the discussion; no useful information can be extracted from it because the network is too small. The other provinces are more interesting: it seems that with increasing degree, the average clustering coefficient stays around the same average, but with a quite a lot of fluctuations. We cannot speak about hierarchical clustering in these separate provinces, however the average clustering coefficients are higher than what you would expect in a random network.

Third and last, the geodesic (or shortest distance) distributions are shown below.

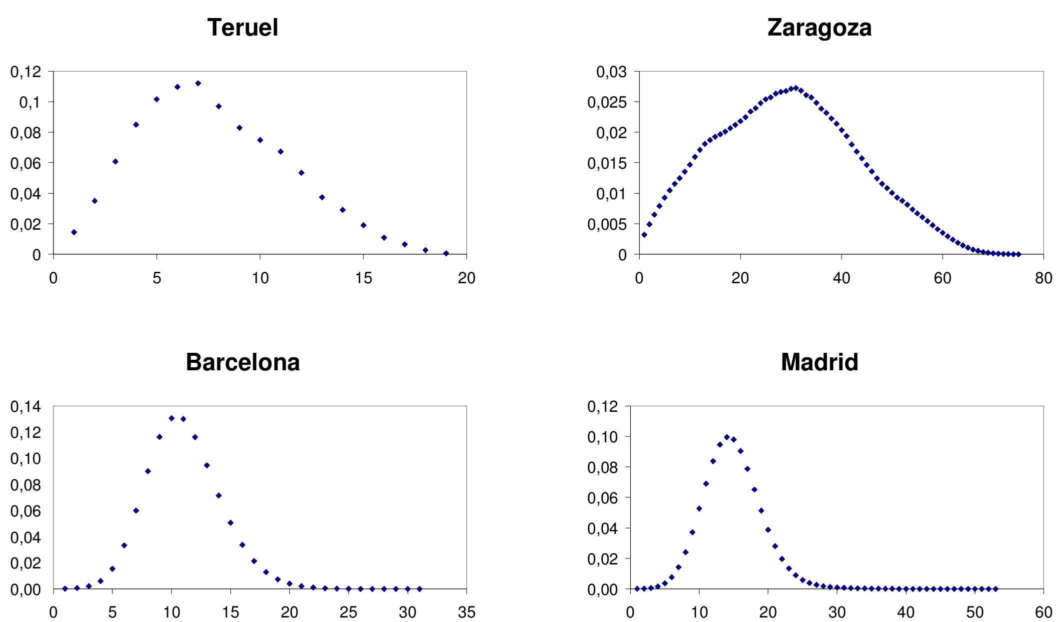

Figure 11: Geodesic distributions from the four provinces. On the horizontal axis the distance l, on the vertical axis P(l), the chance that a randomly selected node pair has shortest distance l. With increasing number of nodes, the distribution looks more and more like a Gaussian distribution.

For the biggest two provinces, the distribution is almost like a Gaussian (or normal) distribution; however it has a slight right tail. In the Teruel and Zaragoza networks, this tail is not exactly a tail but quite a large portion of the distribution. This might well be related to the smaller size of the network. It is also shown quite clearly that none of the 4 provinces can be considered really small-world, although Madrid comes closest to having this property.

Out of this we can draw a few conclusions. The provinces all look like a scale-free network, but none of them shows all the characteristics and all have their own peculiarities. The geographical properties of the networks (Madrid & Barcelona being networks mainly concentrated in a big city, and Zaragoza being a long-stretched province with a large diameter) are definitely reflected in the characteristics they show. None of the provinces show hierarchical clustering, but this is not strange since the provinces can logically be considered as the clusters of the entire optical fibre network. There is however a fairly large fluctuation in the average clustering coefficients, which points to more or less arbitrary clustering.

3.1.2 Entire network of Spain

Now, we will present the graphs of the complete network, Figure 12 depicts the degree distribution, clustering coefficients distribution and the geodesic distribution. Degree and clustering coefficients are obtained from our self-written Java software, the geodesic distribution is obtained using Pajek [7].

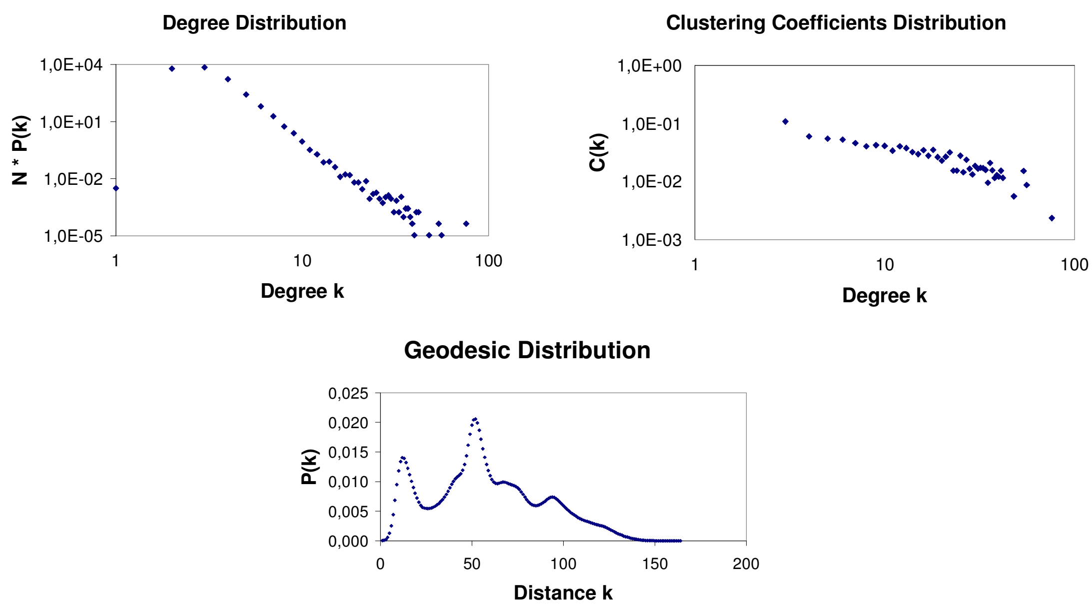

Figure 12: The degree distribution (top left), clustering coefficient distribution (top right) and geodesic distribution (middle below) of the Telefónica España optical fibre network.

The table in Figure 13 has a list of general characteristics of the network.

3.2 Comparison to the scale-free model

First, as a little recollection of the scale-free model: a scale-free network is a network with remarkable features which are closest to most real-world networks. Typically a scale-free network contains a fair number of nodes with extremely high degrees, relative to the degrees you would expect in a random network. Because of that, the network is called scale-free. It has a power-law degree distribution ($P(k) = k^{-\gamma}$) and nodes with lower degrees have usually a higher clustering coefficient.

In order to make good comparisons with the mathematical scale-free model, we let Pajek generate a random scale-free model for us; the number of nodes, links and the average degree of Telefónica network are given as parameters. Extra parameters are necessary for the generative model as described in the Network chapter (growth & preferential attachment); as an initial graph we take an ER-graph of 10 vertices, an initial link probability of 0.8 and α as 0,25 (for information where α is used for [8, p.18]).

The following table gives us the general characteristics of our optical fibre network, compared to two randomly generated networks using Pajek.

	TdE	Scale-Free	Random
Number of nodes	89738	89316	89738
Number of links	101499	101497	102716
Average degree (*k*)	2.27123	2.27276	2.28924
Average clustering coefficient (*C*)	0.05439757	8.3893E-6	7.8270E-6
Average shortest distance	57.72209	10.67226	13.41544
Diameter (*l*)	164	24	35

Figure 13: Table with statistics from two randomly by Pajek generated networks, one scale-free and one Erdös-Rényi random, compared to the statistics of our optical fibre network.

3.2.1 Degree- & Clustering Coefficients Distribution

Concerning the degree distribution, our network can absolutely be called scale-free. It roughly follows a power-law function with $\gamma = 2.8$, although it is a little higher for the nodes with lower degrees. This distribution certainly is not like a random graph, and although you would expect a limited maximum degree of a node (because the network is, in fact, real with practical limits, and is not virtual such as the World Wide Web), the degree distribution still appears to be a power-law function.

The average clustering coefficient of the network is quite a bit higher than expected, around 0.05. Although the clustering coefficient varies quite a bit in real-world networks, another real network of similar background (the Internet on a router-level, see Figure 15) has a clustering of around 0.18. Compared to the scale-free model, the clustering coefficient is very high, a mathematical scale-free network with k = 4 has an average clustering coefficient of around 0.0003 (see Figure 4), and our own generated scale-free network has a CC which is even a factor 30 lower. Looking at the two tables on [4, p. 8], the suggestion arises that the typical clustering coefficient found in the mathematical model is too low for most real scale-free networks.

The clustering coefficient distribution does show a slight hierarchical clustering, the nodes with higher degrees have a lower average clustering than the nodes with lower degrees. This is not typical for the mathematical model, although very common in similar real-world networks.

3.2.2 Average shortest path length & distribution

The diameter of the network is, as said above, 164, which means that the two most distant nodes have to travel over 164 links to reach each other. This is quite high, and unlike a scale-free network, because scale-free networks show the small-world property, although it's a bit less than the typical small-world networks. Our Pajek-generated scale-free network has a diameter of 24. The average shortest path length of our real network is 57.72, compared to 10.67 in the generated network. That means that an important feature of a scale-free network, the extremely short path lengths, is missing in our optical fibre network. This is very peculiar since the degree distribution tempts us to think otherwise. Perhaps the limit to the maximum degree in the network is lower than that of the mathematical model, or the number of hubs is smaller.

The shortest path length distribution is an entirely different story. It does not scale with the provinces as the degree distributions do, in fact it has 'mysterious' peaks and holes which were not present in any way in the shortest path distributions of the separate provinces. Our Pajek-generated network has the following geodesic distribution:

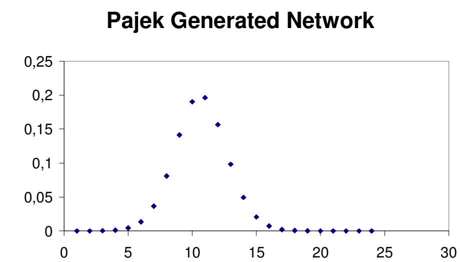

Figure 14: The geodesic distribution of our randomly generated scale-free network.

This distribution is clearly a Gaussian (or normal) distribution around the average geodesic length, 10.67. As we have seen, our individual provinces have distributions that do look quite like this one, but our own complete network is entirely different. Our suggestion is that the large differences in cluster sizes might cause these peaks. Our network has slight hierarchical clustering, but the 10 largest clusters or provinces, out of 50 in total, contain 66% of all the nodes, so these nodes also constitute the major part of all the possible node pairs. The fact that the shortest path length distribution is very different than the separate provinces while the degree distribution does scale very good (with increasing size) to the whole network, indicates that the separate provinces are very restricted by geographical constraints in their possibilities to connect to other provinces. The provinces are geographically divided which is probably the most important reason that this network is not totally scale-free. However the provinces on themselves are more or less scale-free. Such a long average path length & diameter also indicate that we can perhaps improve the network by putting extra fibres in the ground on the relevant places, however the above mentioned geographical constraints will be of great influence to this. Another question that arises is whether the network robustness (see Chapter 2.2.3, Scale-free networks) against random node failure is affected by these higher shortest-path lengths. These two subjects need further study.

3.2.3 Comparison table

For comparison and to illustrate that there are quite a few differences between various real networks, a table with a few typical real-world networks is presented below. On the latest line are the Telefónica España fibre network characteristics.

Network	Size*	k	*l*	l_{rand}	*C*	C_{rand}
WWW, site level, undirected	153.127	35,21	3,1	3,35	0,1078	0,00023
Internet routers, domain level (2 studies)	3015 6209	3,52 4,11	3,7 3,76	6,36 6,18	0,18 0,3	0,001 0,001
Movie actors	225.226	61	3,65	2,99	0,79	0,00027
Math co-authorship	70.975	3,9	9,5	8,2	0,59	5,4E-5
Neuroscience co-authorship	209.293	11,5	6	5,01	0,76	5,5E-5
Words, co-occurrence	460.902	70,13	2,67	3,03	0,437	0,0001
Power grid	4.941	2,67	18,7	12,4	0,08	0,005
Telefónica optical fibres	89.738	2,27	57,72	13,42	0,05439	7,8E-6

Figure 15: The general characteristics of several real networks. For each network is indicated: the number of nodes, the average degree k, average path length l and the clustering coefficient C. For comparison, average path length and clustering coefficient of a random graph with the same size and degree are added.
(Edited table, original including references to these studies can be found in [4, p. 8]*)*

From Figure 15 we see that there are quite some differences between various real-world networks concerning the network characteristics that we discussed in this paper, every network topology is greatly affected by the type of real-world information they model. The network that shows characteristics more or less similar to our network is the power grid, however the path length deviates quite a bit and this network has only 4941 nodes and we cannot say if these characteristics are scalable to the size of our network.

4. Conclusions

From our work we can conclude that the degree distribution of our network certainly fits the scale-free model in a satisfying way. The clustering coefficients distribution also looks very promising, and although the average clustering coefficient is another story, it is shown that the clustering coefficients in real networks differ quite a lot from the typical value in the mathematical model. Hence it is more plausible that the clustering of a scale-free model doesn't correctly fit our real-world networks, instead of the other way around. Our network does show slight hierarchical clustering, which is also typical for similar real-world scale-free networks. Compared to the separate provinces the degree distributions scale very good (with increasing size) to the complete network, but the clustering coefficients distribution changes for the whole network, and the shortest distance distribution changes even more. The strange peaks in the shortest distance distribution can probably be attributed to the very unequal distribution of nodes in the various network clusters.

The Telefónica España optical fibre network is a network with a scale-free degree distribution and clustering coefficient distribution, but lacks other important characteristics such as the small-world property. It has been demonstrated that a virtual Erdös-Rényi random network (with same number of nodes and same average degree) has a smaller average shortest path length and diameter than the ones in our network, this indicates that the network is in a large sense geographically divided and only locally scale-free. Whether these longer shortest-path lengths have influence on the robustness of the network has not yet been investigated. We can therefore use the scale-free network model as the basic model for further network investigation, but must keep into our minds that the small-world property is not present.

5. References

1 Melanie Mitchell. *Complex systems: network thinking*. 2006

http://www.santafe.edu/research/publications/workingpapers/06-10-036.pdf

2 John L. Casti. *Complexity – An introduction*

3 Source of picture: http://en.wikipedia.org/wiki/Clustering_coefficient

4 R. Albert & A. Barabási. Statistical mechanics of complex networks. 2001

5 A. Barabási, R. Albert, H. Jeong – Scale-free characteristics of random networks: the topology of the World Wide Web. Elsevier Physics A 281 (69-77). 2000

6 C. Christiansen & R. Albert – Using graph concepts to understand the organization of complex systems.

7 Vladimir Bagatelj and Andrej Mrvar. Pajek – Program for Large Network Analysis.
http://vlado.fmf.uni-lj.si/pub/networks/pajek/

8 Vladimir Bagatelj and Andrej Mrvar, Pajek manual.
http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf

Error tolerance analysis of the Telefónica España optical fibre network

Pieter van Wijngaarden

February, 2008

Telefónica Investigación y Desarrollo, Emilio Vargas 6, 28043 Madrid

1. Abstract

This paper elaborates further on the network complexity of the Telefónica España optical fibre network, by analyzing the network structure considering link failure, node failure and other error-tolerance related characteristics. We performed various kinds of error tolerance simulations, draw conclusions from them regarding the structure of the network and how this structure is affected by network failures. This gives us the opportunity to determine if the network has certain weak points, what the overall strength of the network is and whether or not it is needed to perform structural enhancements.

2. Introduction

Networks can sometimes behave as complex systems with unpredictable behaviour. Mathematically analyzing a complex system is a very difficult task. We concluded in the previous paper [1], the Telefónica network shows some scale-free network characteristics, but not all characteristics were decisive, and it appeared that the geographical constraints on the network have an impact on the network structure which cannot be ignored.

In a standard scale-free network, robustness against random failures is very high, because there are a few nodes (the hubs) which connect to each other and to a very large portion of the rest of the nodes. In this paper, we are going to research if the Telefónica network, despite its geographical constraints, shows similar robustness properties.

Thus, the key questions we want to answer here is the following:

What error tolerance characteristics does the Telefónica optical fibre network show, considering link and node failures?

In what way can we use this information to increase the network robustness or structurally enhance the network?

First, we describe the type of simulations that were performed and the information that we want to obtain by doing these simulations. Then, we present the results and analyze them in order to answer the questions formulated above and draw our conclusions.

3. Network overview

The Telefónica España network under survey is a large optical fibre transport network, with roughly 90000 nodes and 105000 links connecting them. It has an average degree of around 2,27, which means that most nodes are connected to 2 or 3 other nodes. Its deployment is still in an initial stage. We abstracted in a large sense from the underlying technologies, since they are irrelevant to the simulations.

Our network is shows scale-free network characteristics, but also lacks quite a few of them. The degree distribution follows a power-law function, and clustering is also fairly similar, but the average shortest path lengths between nodes are a lot higher. Even in a random Erdös-Rényi network with the same number of nodes, the average shortest path length is shorter than in our network. This indicates that the TdE network is a real transport network which intends to provide coverage to the entire country. Typical characteristics you see in virtual scale-free networks (like the WWW) are simply not possible, because we are dealing with a real network and there are physical upper bounds to the maximum degree of a node (for example). These geographical constraints could mean that link failures might have a larger impact on the network structure. [1]

4. Simulation descriptions

When performing a simulation, the main goal is always to imitate certain real-world events which might occur and study the impact and consequences of these events. In our case, we want to know how the structure of the network changes when it suffers from failures. We define a few different events, and for every event, we ask ourselves a few questions which we want to answer by doing the simulation:

-random single link failure

> Nowadays, optical fibres are very well equipped with various protection layers. It is thus not very probably that an optical fibre breaks, but in some occasions, it can happen. Examples are construction works or cable infrastructure maintenance, or a bit more extreme: natural disasters like floods and earthquakes.
>
> Questions:
>
> 1. What is the probability that a link failure will result in a disconnected (i.e., split into two parts) network?
> 2. If the network becomes disconnected, how many nodes cannot reach the main network anymore?
> 3. If the network is still connected, how about the traffic between the node pair that was previously connected? Is there a protective link? If there is no protective link, how many steps/hops does the new shortest path between the two nodes have?
> 4. Is there a relation between the capacity of the link and the impact of a possible failure? Do the large capacity links have more backup/protective links, and will the result of such a failure therefore be smaller, or will it be bigger because it is an important link connecting a large part of the network?

-random multiple link failure between one single node pair

> In quite a large fraction of the node pairs, multiple links exist between them. This is of course to increase robustness, various examples are known of optical ring networks which have a working link and for every working link, there is also a protection link present. This protection link is seamlessly put into operation as soon as the working link breaks, hereby keeping the ring intact. Little information is known about the paths of these protective links: do they follow exactly the same route as the working link? It is not unreasonable to believe that in some cases, they do so indeed. If they follow the same path, it is also fairly reasonable to believe that if one of them breaks because of external influences, the other one breaks too.

Therefore we also model the event of failure in all the optical fibre links between one random single node pair.

Questions:

We ask ourselves the same questions here as the ones mentioned above, with the only exception that we know that the protective link is not present anymore. Also, question 4 is largely irrelevant in this simulation, because it specifically considers the protective links that we do not take into consideration in this simulation.

-random node failure

Nodes are typically switchboards (can be optical or hybrid, but usually electrical) connecting multiple fibres. As we have shown in [1], the amount of fibres connected by a node is typically very small, 2 or 3, but there are nodes with a very high degree. Nodes can suffer from failure because of various reasons: hardware damage (the hardware life span is a lot smaller than the lifespan of the optical fibre cable itself), power loss etc. The event of a node failing temporarily is perhaps the most probable thing to happen, in case of failures in the network.

Questions:

1. What is the probability that a node failure will result in a disconnected (i.e., split into two or more parts) network?
2. If the network becomes disconnected, how many disconnected parts do we have, what sizes do they have and what is the total number of nodes that cannot reach the main network anymore?
3. Is there a relation between the degree of a node and the impact of the eventual failure of that node? We suspect that nodes with a large degree indeed have a more important function in the network, is this indeed the case?

We mention here that there are many different ways of analyzing the robustness, and there are many more parameters which can be taken into consideration. For example, the amount of traffic flowing through certain links is also very important when analyzing the network: will the network still be able to support the same data flow between certain node pairs, if a failure occurs? This of course depends on the capacity of the link and the amount of traffic it generally needs to transport. However, we had no traffic statistics at our disposal when performing the simulations, so this research focuses on the structural and topological behaviour of the network only. Incorporating the amount of traffic in these simulations is a recommendation for the future.

All the simulations were performed on the very same network that was also described in [1]. The network did probably undergo some minor changes in the meantime, but we will not take them into consideration.

5. Simulation results

Single link failure

With this simulation we investigated the failure of one single link, and what impact this has on the structure of the network. For a total of 30.000 times, we removed one link (on a totally random basis), analyzed the network structure and then put the link back in. From these 30.000 simulations we derive the statistics which are now given.

Question 1: What is the probability that a link failure will result in a disconnected (i.e., split into two parts) network?

If you consider only one link, you can at most get two subnetworks that are disconnected from each other when it fails. We present the following table:

# (sub)networks	*# simulations*	*probability*
1	14833	49,44
2	15167	50,56
total	30000	

Table 1: Disconnection probabilities

The first column contains the number of resulting (sub)networks, one subnetwork means the network is still totally connected. As we see, in a little more than 50% of all the simulations, the network got disconnected (i.e. split up in two parts). This seems like a rather high figure, but corresponds with the clustering coefficients we found (which were lower than average). A high average clustering coefficient means [1], for high clustering effectively also means that link removal has a very limited effect.

Question 2: If the network becomes disconnected, how many nodes cannot reach the main network anymore?

Very interesting to know is how big the damage is exactly, if a node failure resulted in a disconnected network. This information is given in the following table:

damage	*#simulations*	*%*	*cum. %*
0	14833	49,44	49,44
1	7770	25,90	75,34
2	3409	11,36	86,71
3	1115	3,72	90,42
4	608	2,03	92,45
5	435	1,45	93,90
6 - 10	1021	3,40	97,30
11 - 20	603	2,01	99,31
20+	196	0,65	99,97
50+	9	0,03	100,00
100+	1	0,00	100,00
	30000		

Table 2: structural damage with single link failures

In this table, the damage indicates the number of nodes which got disconnected from the main network, and next to it the number of simulations where this was the yielded damage. From this table, we find that in approximately 94% of the simulations, the damage was smaller or equal to 5, and in only 1/3000 (=0,033%) is the damage larger than 50. If a random link failure would occur, we can say that the probability of a large structural impact is very limited.

Question 3: If the network is still connected, how about the traffic between the node pair that was previously connected? Is there a protective link? If there is no protective link, how many steps/hops does the new shortest path between the two nodes have?

shortest distance	*# simulations*	*%*	*cum. %*
1	4487	30,25	30,25
2	2983	20,11	50,36
3	1939	13,07	63,43
4	1065	7,18	70,61
5	789	5,32	75,93
6 - 10	2153	14,51	90,45
11 - 20	1006	6,78	97,23
20+	336	2,27	99,49
50+	62	0,42	99,91
100+	13	0,09	100
	14833	100	

Table 3: New shortest distances

In this table, we can see that in approximately 30% of the simulations there is a protective link present. Although most shortest paths are quite reasonable, there is however a fair percentage (around 10 %) where the traffic has to be rerouted over more than 10 hops. It is fair to assume that in these cases a serious delay increase will be noticed. The average new shortest distance is 4,78 hops.

The last question we want to answer is an interesting one. In a typical optical fibre network, the high-capacity links are located in the center of the network (where the most traffic flows). What is actually the damage if those really important links are damaged or removed?

Question 4: Is there a relation between the capacity of the link and the impact of a possible failure? Do the large capacity links have more backup/protective links, and will the result of such a failure therefore be smaller, or will it be bigger because it is an important link connecting a large part of the network?

First, we put the weight (capacity) of a link against the probability that removal results in network disconnection:

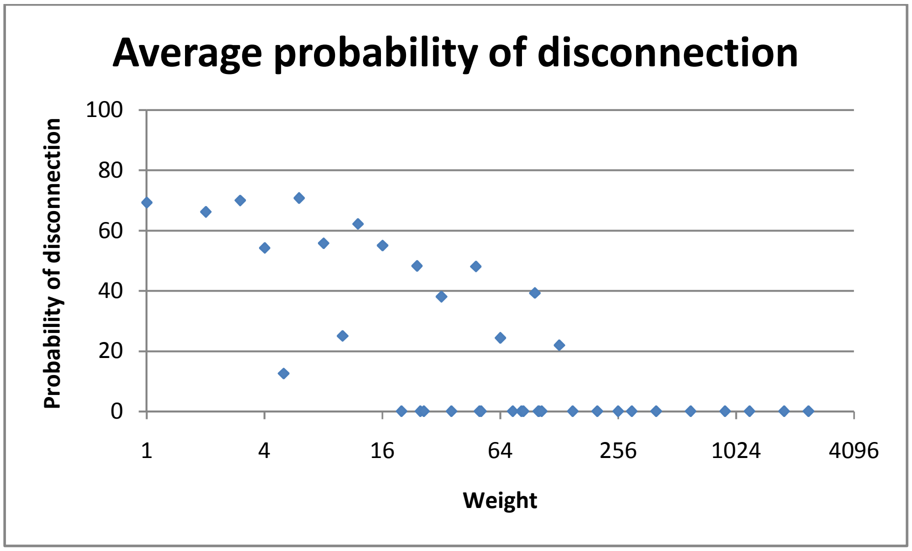

Figure 1: the average probability of disconnection. The larger the capacity of the link, the smaller the probability that the network gets disconnected, for the large weight, this probability even reduces to zero.

Here we can see that for higher capacity links, the probability that the network disconnects becomes smaller. This indicates that the higher-capacity links are more in the core part of the network (as can be expected).

Now if we also look at the relation between the weight of a link and the average size of the disconnected part (given that removal of the link indeed caused disconnection), we get the following graph:

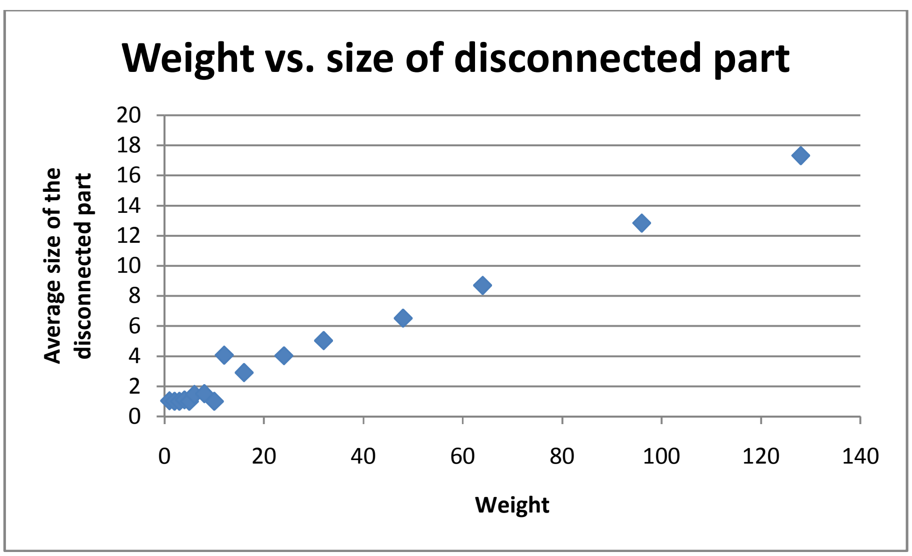

Figure 2: The relation between the weight (capacity) of a link and the size of the disconnected part (if removal causes disconnection)

Random multiple link failure between one single node pair

Now, we change the type of simulation a little bit and instead of removing one link between a node pair, we remove all of them (if there is more than 1 present). We now look at the network as if there were no protective links. We present this data and compare it with the previous simulation.

First, we consider the global disconnection.

# (sub)networks	# simulations	probability without protective links	probability with protective links
1	12534	41,78	49,44
2	17466	58,22	50,56
total	30000		

Table 4: Disconnection probabilities II – Network with vs. without protective links

As you can see, the probability that the network becomes disconnected raises significantly. This is a good indication of the importance of the protective links that are present in the network. If we take into account that roughly 15% of the connected nodes have an extra link, we can conclude that the protective links indeed have a positive effect on the network structure.

Second, we looked at the number of nodes that got disconnected (i.e. the size of the disconnected part). For the network without protective links, this graph is as follows:

damage	#simulations with protective links	# simulations without protective links
0	14833	12534
1	7770	9544
2	3409	3702
3	1115	1174
4	608	603
5	435	461
6 - 10	1021	1120
11 - 20	603	638
20+	196	211
50+	9	13
100+	1	0
total	30000	30000

Table 5: Size of the disconnected parts II – Network with vs. without protective links

As we can see, the differences are small. Protective links make sure that there is less disconnection, but if there is disconnection, the damage is, on the average, not much smaller than in the network without the protective links. Removing all the protective links thus only slightly decreases the inflicted damage in case of link failures. This is better illustrated in a graph:

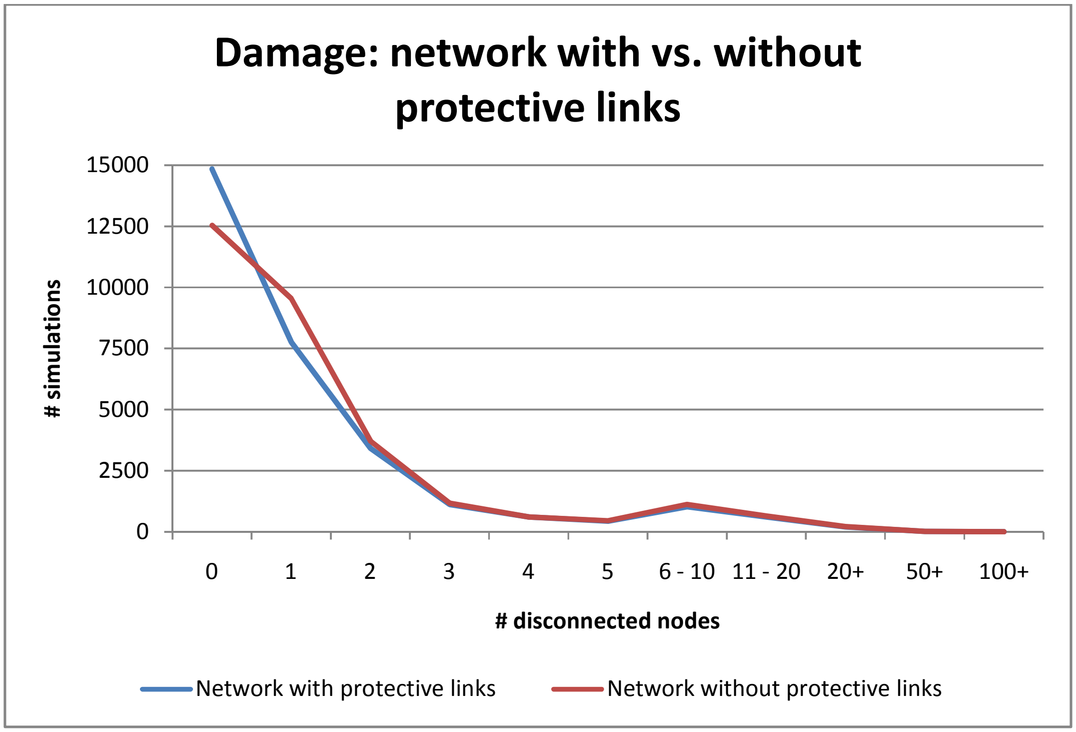

Figure 3: Damage on the network, comparing the network with and without protective links. The difference in caused damage is minimal.

Regarding the traffic between the nodes if there is no network damage, there should not be considerable differences: the new shortest distances can be a bit higher, because there will never be a protective link that can deliver the traffic in just one hop to the other node. However the shortest distance as it was if there were no protective link in the first place, will of course still be the same.

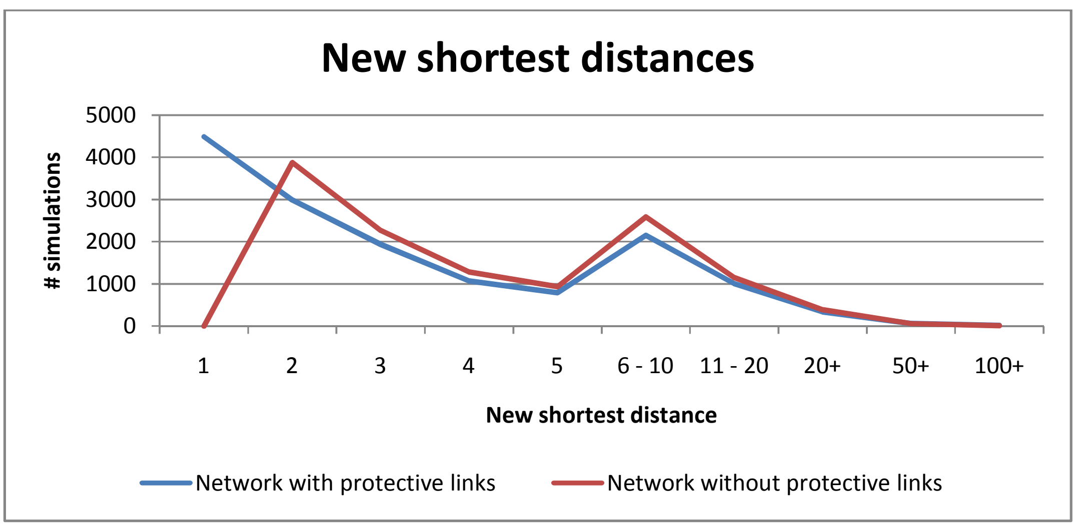

Figure 4: The shortest distances that 'emerged' after removing a link, comparing the network with and without protective links. Because the protective links are removed, the new shortest distance between two nodes will never be one.

We can see that the distribution of these new shortest distances is still relatively the same, the numbers are just a little bit different.

The comparison between these two simulations shows that the protective links definitely serve their purpose, especially for the more important links, but a random error or failure is more likely to occur in a link without protection, so the effect on the average is limited.

Random node failure

Another interesting event to simulate is the failure of a node. This node might be a switchboard or switching centre and might fail due to power loss, hardware damage etc.

We ask ourselves more or less the same questions as in the case of the random link failure. The set-up of the simulation is the same; i.e. we simulate node failure by randomly selecting one from the network, removing it and removing the links that are connected with it (in order to prevent 'dangling' links which do not have a node at both ends). Then we check the network structure and restore the network in the original state, before performing the next simulation run.

Question 1: What is the probability that a node failure will result in a disconnected (i.e., split into two or more parts) network?

# subnetwork parts	*# simulations*
1	16687
2	8961
3	2857
4	865
5	310
6 - 10	277
11 - 20	36
20+	7
	30000

Table 7: Global network disruption with random node failure simulation.

The table above shows the same information as table 1 and 4, but because one node can be responsible for connecting multiple subnetworks, the number of parts can be larger. In this simulation, a little more than 50 percent does not cause network disruption (every node can still reach every other node). The probability that the network is still intact is thus larger than with a link failure, but if there is damage, the impact is larger. How large this impact really is can be illustrated with the following question:

Question 2: If the network becomes disconnected, how many disconnected parts do we have, what sizes do they have and what is the total number of nodes that cannot reach the main network anymore?

Now we only look at the simulations where network disruption actually occurred. This happened in 30000 – 16687 = 13313 cases. We want to know how many nodes have become disconnected, this gives us an idea of the impact of the failure. We present this information in the following table:

# nodes disconnected	*#simulations*
1	5536
2	2229

3	1166
4	859
5	606
6 - 10	1583
11 - 20	938
21 - 50	330
50 - 100	52
100+	14
total	13313

Table 8: Number of disconnected nodes in random node failure simulation.

Here we can see that in most cases if there is structural damage to the network, the impact of it is very limited. This image fits very good with a general scale-free network [2]. If we present the same information in a pie graph:

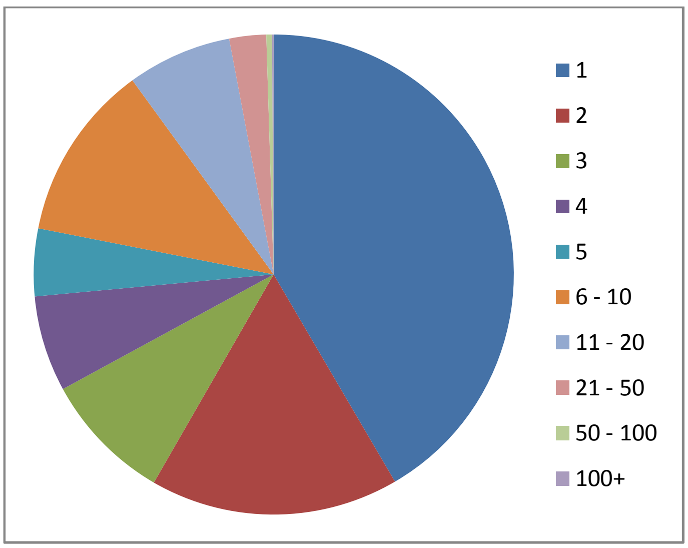

Figure 5: Number of disconnected nodes in a random failure simulation (graphical form).

Here we can see that if there is damage, in more than 75% of the cases this damage is limited to 5 or less nodes. Relative to the size of the whole network, this damage is very limited.

Question 3: Is there a relation between the degree of a node and the impact of the eventual failure of that node? We suspect that nodes with a large degree indeed have a more important function in the network, is this indeed the case?

This third question is also interesting, since it gives us an indication of the (structural) importance of a node to the whole network. If a node is more important than others, the damage is bigger if it is removed. We tested if there is a relation between the degree of a node and the impact of the node removal. We therefore looked at all the nodes with a specific degree, and checked how many percent of the nodes caused network damage (if removed). We present the following graph:

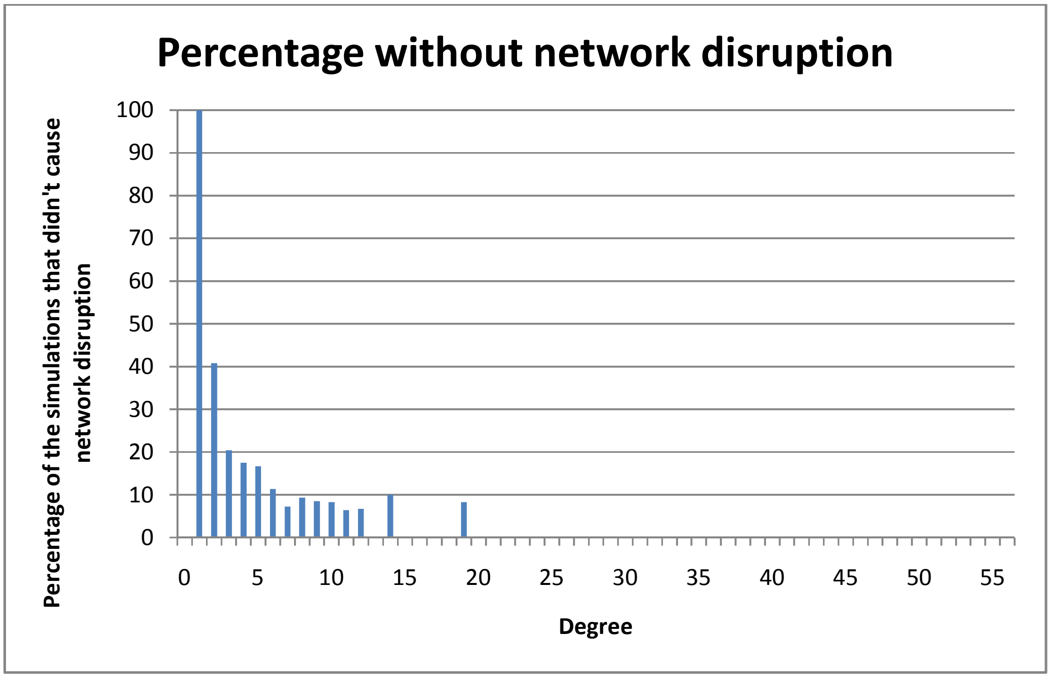

Figure 6: The relationship between the degree of a node and the percentage of simulations which did not cause network disruption.

This graph shows for a certain degree the percentage of simulations which did not cause network disruptions. If a node has degree 1, of course the network is still intact if it is removed, since it has only one neighbour. It is shown that for higher degrees, the average probability of network disruption becomes bigger. However, it must be noted that during simulation only 53 nodes (53/30000 = 0,177 %) of the tested nodes had a degree of 20 or more. This relates of course to the degree distribution of the network (see [1]). We can conclude that nodes with a higher degree indeed have an important role in the network (from a structural point of view).

Above we investigated the relation between the degree of a node and the number of disconnected parts or subnetworks that the removal of that node causes. The last important question is if the *number of disconnected nodes* is also bigger if the degree of a removed node is larger. We thus also investigated the relationship between the degree of a node and the number of nodes which got disconnected. This information is given in the following graph:

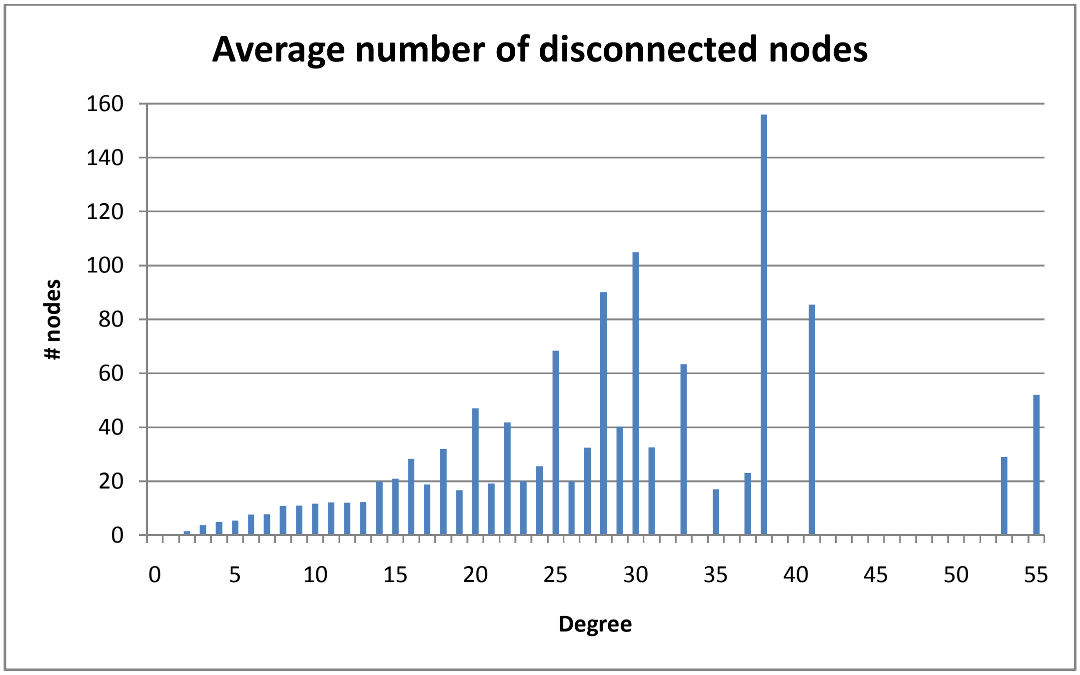

Figure 7: The relationship between the degree and the number of disconnected nodes.

This graph shows that for higher degrees, the number of nodes that got disconnected raises slightly. The 'gaps' in the graph at the higher degrees mean that there were simply no nodes among the tested ones that had that degree (i.e. the gaps do **not** indicate that for certain degrees, the resulting damage is suddenly zero).

6. Conclusions

The simulations that we performed provided us a few valuable supporting insights on the structure and behaviour of the network. We can see that the effects of link failures are very limited; in most cases there is some damage (i.e. one or a few nodes are disconnected from the main network) but the probability of a large structural impact is almost zero. If the network is still connected, the two nodes which 'lost' their link can still reach each other in less than 5 hops on the average.

We also see that links with a higher capacity have a smaller probability of causing a network disconnection, which indicates that these links at least have structural backups (i.e. if a large-capacity link fails, there is certainly an alternative path). What this means to network delays and if the alternative path has enough capacity, is not considered here. However, the weight (or capacity) of the link and the impact are still linearly related to each other, i.e. if the link removal causes structural damage, then the damage is larger with a high-capacity link.

If we remove all the protective links from the network, we see that the probabilities of structural damage raise slightly, however in these cases where there is extra damage, it often limited to 1 or 2 nodes that got disconnected (this can be seen in Figure 3). Also, the new shortest distances do not change significantly.

These conclusions also apply to the node failure simulations. The damage is limited, and it is clearly shown that removing a node with a higher degree has a larger impact.

We can thus conclude that the Telefónica optical fibre network shows error tolerance characteristics similar to typical scale-free networks, i.e. random failures usually have a small effect, but if a directed attack hits the nodes with the highest degrees or the links with the largest capacity, the impact is significantly larger[2]. However, further analysis might be necessary to determine if the QoS-demands can still be met if certain parts of the network fail.

7. References

1 Mary-Luz Mouronte-Lopez, Victor Feliu, Pieter van Wijngaarden. Analysis of the Telefónica España optical fibre network complexity. June, 2007

2 Réka Albert, Hawoong Jeong, Albert-László Barabási. The Internet's Achilles Heel: Error and attack tolerance of complex networks. 2000