

Small world and scale-free urban transportation networks: analysis and protection strategies in case of threats.

Fabio Lamanna. (fabio@timenetwork.org)
Dept. of Civil and Environmental Engineering.
University of Trieste.
Piazzale Europa 1. 34127. Trieste, Italy.
Tel. +390405583573/76.
Fax +390405583580.

1. Background.

During the last ten years the study of *complex networks*, has rapidly improved. Complex networks have been defined as 'systems whose structure is irregular, complex and dynamically evolving in time'. This activity, triggered by two seminal papers (Watts and Strogatz, 1998 and Barabási and Albert, 1999), aroused the interest of the scientific community in analysing the huge number of real networks under another point of view. These networks include social relationships, phone calls, Internet and the World Wide Web, actors' collaboration in movie databases, citations and neural or genetic networks. The massive and comparative analysis of these widely different networks led to the identification of a series of unifying statistical properties and principles which are common to most of the real networks considered; the main properties are the *small world* phenomenon and the *scale-free* behaviour.

Some real networks are characterised by correlation in their node degree, by having relatively short paths between two nodes and by the presence of a large number of interconnected groups of nodes, called *clusters*. This is the *small world* phenomenon, which is firstly related to social sciences, and then extended to other kinds of real systems in order to perform their closeness and therefore their stability in communication between nodes. It is hypothesised by some researchers (Barabási 2002) that the prevalence of *small world* networks behaviour in several systems may reflect an evolutionary advantage of such an architecture.

In many cases networks are characterised by some nodes, called 'hubs', with much more connections than others. This feature is one of the main reasons why these structures' topological behaviour differs from that of other structures. These networks are known as *Scale-free* because its dynamics is independent of the system's size, i.e. of the number of nodes the system has. In other words, a network that is *scale-free* will have the same properties no matter what the number of its nodes is.

After reviewing the classical and complex networks approaches on robustness and after introducing these study's purposes, the attention will turn to different complex network measures and to the robustness of *small world* and *scale-free* models. Therefore the description of the tool developed by the author will lead the study to the analysis on different *small world* and *scale-free* networks of Berlin consisting of three fixed networks (U bahn, S bahn and Regional Bahn) under different hypothesis of failures and attacks.

2. Overview.

Robustness and reliability of transportation networks have been studied so far in different ways, from a game theory approach (Bell, 2000) to social behaviour of users (Jenelius, 2006), from road elements weakness concepts (Schreuder et al. 2007) to risk-

related analysis (Ezell et al. 2000). All classical approaches never took into account different kinds of 'attack' strategies which may affect the network and therefore its protection against threats.

Complex networks physical theories (Latora and Marchiori, 2004) helped in considering the functionality of all the elements in the system related with the response of the whole network to their deactivation. Literature about complex network theories and related applications to real systems is very large. Gallos et al. (2005) introduced the concept of defence strategy of many information and computer networks against external attacks. Moreover, both Holme and Kim (2002) and Albert et al. (2000) started to study the behaviour of different real networks under failures or attacks.

The extension of complex networks theories to transportation cases needs almost all the mathematical instruments of the typical graph theory, along with some natural changes adopted in order to analyse several constraints as travel times and frequencies.

Some studies (Von Ferber et al. 2007) tried to model and analyse metropolis public transport networks but they didn't take into account many engineering time-related features of the system, as travel times and frequencies of services. A very useful measure has been recently developed (Nagurney and Qiang, 2007) and it takes into account flow, behaviour and cost in a vulnerability analysis on different elements of a network. It considers the equilibrium flow of a transportation network but it has been applied only to a theoretical simple network, without taking into consideration different transport modes and time-related communications. Another only topological analysis on urban road network vulnerability can be found in Zhang et al. (2007).

3. Methods and Purposes of the Study.

For the above mentioned reasons, the method leading this research to its objectives needs an accurate analysis on how the network is related to its dynamical components as travel times, frequencies of the services and to its topology. A more general graph theory will be adopted (weighted graphs) in order to take into account these basic and typical aspects of transportation engineering in a correct analysis of the robustness of the system.

The aim of this thesis is the development of a tool which can extend complex networks theories, *small world* and *scale-free* models to transportation systems under a new time-related approach. In particular it should provide an instrument to estimate which are the most critical components of the network in case of random failures or deliberate attacks and how the system can be protected against threats.

4. The Structure of Complex Networks.

A graph of N nodes can be represented by means of a $N \times N$ square matrix $\mathbf{A}(a_{ij})$ called *adjacency matrix*: its entries a_{ij} take the value 1 if an edge connects vertices i and j and 0 otherwise. The diagonal of this matrix contains only zeros. The *degree* or connectivity of a node i is defined as the number of edges incident with the node in terms of its adjacency matrix A .

Shortest path plays a fundamental role in transport and communication within a network. A measure of the typical separation between two nodes is the so called average or *Characteristic Path Length* L , defined as the mean of shortest path length over all couples of nodes:

$$L = \frac{1}{N(N-1)} \sum_{i, j \in N, i \neq j} d_{ij}. \quad (4.1)$$

As the reader will notice, this widely adopted measure shows a problem inside its definition; in case of disconnected components of a graph, the shortest distance between non-connected nodes goes to an infinite value, and therefore L diverges. To avoid this problem, an extended definition of distance between nodes has been introduced (Latora and Marchiori, 2001) with the aim of considering how efficiently information are exchanged through the graph.

The description of the extended approach needs both the elements of the basic graph theory elements and some changes. A graph G needs two matrices to be fully described: the adjacency matrix $\mathbf{A}(a_{ij})$ and the matrix related to the weights associated with the edges, called $T(t_{ij})$ because it can be related with travel times between nodes. In general $\{t_{ij}\}$ is supposed to be known even if nodes i and j are not connected (e.g. physical distances).

In this case $\{d_{ij}\}$ is calculated using information contained both in A and in T matrices, and represents the smallest sum of the travel times (in this particular case) between nodes i and j . When taking into account the event of disconnected nodes (i.e. there is no path between them) the efficiency ϵ_{ij} in the communication between vertices can be defined as the inverse of the shortest distance. When there is no path between two nodes d_{ij} goes to ∞ and therefore ϵ_{ij} goes to 0.

$$E(G) = \frac{\sum_{i \neq j \in G} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i, j \in N, i \neq j} \frac{1}{d_{ij}}, \quad 0 \leq E(G) \leq 1 \quad (4.2)$$

In general E is the global efficiency and it is normalised by taking into account the ideal case in which the graph G has all the $N(N-1)/2$ possible edges. Although the equality $E = 1$ is valid when there is an edge between each couple of vertices, real networks can reach a high value of efficiency.

A final useful measure is the clustering coefficient: it is a typical property of connected networks and it measures how two nodes connected to a third one are also connected to each other. It is defined as the ratio between the number of connected neighbour pairs of node i and the maximum number of edges in **the subgraph of node i** and, by definition, its value ranges from 0 to 1 in case of a fully interconnected graph.

5. Information Centrality.

Various measures of the so called *centrality* of a vertex within a graph have been introduced. This concept is very useful in network analysis because it takes into account the relative importance of a node (or of other elements of the graph) within the graph. The importance is related both to the topology and to the behaviour in terms of efficient communication between vertices. The most important centrality measure related with this work is the so called *Information centrality* (IC) defined as follows:

$$Cl_i = \frac{\Delta E}{E} = \frac{E[G] - E[G'_i]}{E[G]}, \quad (5.1)$$

where G is the complete graph, i is the removed element, G'_i represent the graph with N nodes and $K - k_i$ edges obtained by removing from G the edges incident in point i . Information centrality has been introduced in order to evaluate the relative importance of an element in the network by its ability to communicate between other elements. In other words this measure shows the relative drop of performance in the network efficiency

caused by the removal of an element of the graph. This approach can be found in Latora and Marchiori (2005). It aims at the analysis of the vulnerability of information networks in case of failure of links or nodes.

6. The Robustness of Networks.

As briefly reported in the first paragraph, small world networks are recognisable by having small characteristic path length L and high clustering coefficient C . Differently, scale-free networks have the degree distribution $P(k)$, defined as the probability that a node chosen uniformly at random has degree k , that significantly deviates from the Poisson distribution to a power law (*scale-free*) tail with an exponent having value between 2 and 3. These models have been called respectively WS (small world, Watts and Strogatz) and BA (scale-free, Barabási and Albert).

Both *Small world* and *scale-free* networks are generally very robust to failures and attacks, despite their differences in structure and degree distributions. On one hand, the WS model does not exhibit the power law distribution, as many real networks do and as the BA model successfully did. On the other hand, the WS model has got high clustering, while BA model has a clustering coefficient that ranges towards 0 as N goes to infinity. In general the integrity of a network is destroyed after a critical percentage p_c of the system nodes has been removed; for scale-free networks it has been shown (Cohen et al. 2000) that $p_c = 1$. This means that in order to destroy the network almost all nodes must be removed.

Small world networks have almost never been involved in such studies, despite they are considered to be very stable and well-built in term of exchanging information within the network.

7. Threat Models.

When systems are considered as networks and studied for their robustness, threats must be translated into the framework of vertex and edge removal. In the scientific literature there are no existing models for specifically threats to infrastructures or transportation network services. However, many studies have been done on failures and attacks on different kinds of networks. Some of the threat models used in these studies can be applied to transportation networks as well.

- **Random Failures.**
A simple model for probabilistic threats is to considered them as entirely randomly, i.e., the damage can equally occur anywhere in the network. Each failure is considered independent on the others. In infrastructures this may best correspond to failures. It may also be a crude model for other probabilistic threats, such as other kinds of equipment failures or extreme weather. Instead of randomly removing the nodes (or, in general, the elements of the network) another model follows a probabilistic function of historical recording of failures in the system in order to predict if there are some elements which are more prone to critical behaviour in a certain period of time.
- **Attack Strategies.**
For the study of network robustness the selection procedure of the order in which vertices are removed is an open choice. The method maximises the destructive effect at any fixed number of removed elements. However, this plan requires the knowledge of the whole network structure and it is very expensive in terms of time.

A more tractable choice is to select the nodes to be removed according to their number of connections in descending order of degree. Classical attack models can be described (Holme and Kim 2002) related to *static robustness*, i.e. there is no redistribution of flows through the network after the removal of an element: the INITIAL DEGREE DISTRIBUTION (ID removal). Despite many attack strategies can be analysed, this one seems to fit better for this study's purposes because the degree of a node would be a more likely known measure if someone tried to make a serious threat to an important element of the system.

8. Time=net.work.

In the previous paragraphs several new elements and models related with network analysis have been presented. They help in defining both topological and structural characteristics of a network in order to find information to predict the robustness of a system under failures or deliberate external attacks. The general approach on real networks now has to be extended to transportation networks in order to capture typical engineering features related with travel time and frequencies of services.

The tool (*time=net.work*) developed by the author has been written using both MATLAB and Octave codes, due to their ability and functionality in matrix based computations. In the tool almost all complex network algorithms have been implemented, along with threat models and several graphical output formats.

The extended approach needs a weighted analysis in order to give each edge of the graph a measure that can be associated with travel times between two stations, or the frequencies of services in any part of the network. Moreover, the approach is so general that almost all kind of weights can be implemented in order to perform several analysis related with different transportation cases. In this paper the results obtained with a time-based approach will be presented.

This approach requires an extended definition of efficiency as well in order to take into account the importance of every element (e.g. Information Centrality) of the system by its influence on the scheduled services if subject to attack. The original formula (4.2) has been adapted to this study's topics as follows:

$$E_{TF}(G) = \frac{\sum_{i \neq j \in G} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i, j \in N, i \neq j} \frac{F_{ij}}{t_{ij}}, \quad (8.1)$$

where t is travel time between every pairs of nodes and F is the frequency of the services evaluated on the shortest path. By means of this approach, every element extends its role in the graph from a structural to a service point of view. In particular, the new efficiency measure E_{TF} (Time, Frequencies) maximises the frequencies of services on the shortest travel time path between every two elements of the graph. This measure can be easily associated to a Level of Service of the network.

9. Berlin Urban Network: layers and spaces.

The German capital urban rail network consists of 9 U-Bahn (U), 16 S-Bahn (S) and 14 Regional Bahn (B) lines operating throughout the city. These elements have been defined as *layers*, because they belong to an infrastructural level. All lines have been implemented in the tool as an input data (nodes = stations, edges = connections) along with their scheduled travel times and frequencies available from common timetables. The tool built

the necessary adjacent and weighted matrix A and T . Therefore all the networks have been characterised both topologically and related with services.

Analysis have been performed on different kinds of network structures (or 'spaces') related with the services; according to Kurant and Thiran (2006) different network models represent different levels of analysis:

- Space of Stations (SpS).
This space is strongly related with the physical structure of the network and, as the following results will show, can be modelled by means of a *scale-free* network, in which stations are connected by a number of edges corresponding to the amount of different services.
- Space of Stops (SpST).
This space is related with the physical structure of the network but it captures only the services which stop in two different stations, i.e. every edge of the graph is represented by a single link. This structure can be modelled by means of a *scale-free* network.
- Space of Changes (SpC).
This last space considers the amount of transport mode changes that a traveller has to do in order to reach its destination. Every node (station) is directly connected to all vertices which are reachable without changing transport mode. This network follows the *small world* behaviour.

The first step of the analysis consists in finding the basic structural properties of the networks in all the *layers* and *spaces* previously described. Different complex network models show different reactions to threats. For this reason it is necessary to associate a particular network model (*small world* or *scale-free*) with every system (layer + space) as shown in Table 9.1.

Layer	Space of Stations (SpS)			Space of Stops (SpST)			Space of Changes (SpC)		
	N	L	C	N	L	C	N	L	C
U	170	14.68	0.006	170	14.47	0.006	170	2.05	0.948
S	164	24.29	0.017	164	13.47	0.017	164	1.96	0.838
U+S	306	18.27	0.005	306	12.57	0.005	306	2.13	0.863
U+S+B	306	11.34	0.017	306	8.35	0.017	306	2.12	0.857
	Title:/Users/fabiolamanr Creator:MATLAB, The M CreationDate:01/30/200 LanguageLevel:2			Title:/Users/fabiolamanr Creator:MATLAB, The M CreationDate:01/30/200 LanguageLevel:2			Title:/Users/fabiolamanr Creator:MATLAB, The M CreationDate:01/30/200 LanguageLevel:2		

Table 9.1. Complex networks measures on different layers and spaces. This table shows the number of nodes of the layers (N) and the corresponding values of the characteristic path length (L) and of the clustering coefficient (C). Figures above each space show the probability $P(k)$ that a node chosen at random has degree k in semi-log scale. All layers follow a power law shape on all spaces, and they can be defined as *scale-free*. The values of L and C show the *small world* phenomenon only on SpC case for all layers.

All layers can be considered as scale-free on all spaces, but the small world phenomenon appears only on layers related to Space of Changes, i.e. to the network which represents the point of view of the majority of the users. The considerations about *small world* and *scale-free* network robustness in case of failures and attacks (par. 6), lead to the analysis of systems which should be very robust in case of random failures, and more vulnerable in case of targeted attacks. The following results will prove these general statements which have been found in many real networks of Berlin transportation system.

The analysis has to take into account the drop of efficiency related to travel times and frequencies according to the fraction of the nodes removed, for each infrastructural layer and space of service. Time=network considered the extended definition of efficiency introduced in the previous paragraph (Formula 8.1) and the Information Centrality measure. As explained, the latter quantity evaluates the relative drop of performance in the network efficiency caused by the removal of an element of the graph. Information Centrality allows to list the ranking of the nodes whose removal causes the maximum drop of performance (efficiency) in the network, for all *layers* in all *spaces* (Table 9.2).

Layer	Space of Stations (SpS)		Space of Stops (SpST)		Space of Changes (SpC)	
	<i>Critical Station</i>	<i>Drop</i> E_{TF}	<i>Critical Station</i>	<i>Drop</i> E_{TF}	<i>Critical Station</i>	<i>Drop</i> E_{TF}
U	Alexanderplatz	18%	Alexanderplatz	18%	Alexanderplatz	9%
S	Westkreuz	56%	Westkreuz	56%	Westkreuz	15%
U+S	Lichtenberg	11%	Lichtenberg	11%	Lichtenberg	5%
U+S+B	Südkreuz	5%	Südkreuz	5%	Südkreuz	4%

Table 9.2. Most critical nodes in the global performance of the system. SpS and SpST follow the scale-free attack tolerance which is very low under targeted attacks at its most critical nodes. Space of Changes represents a small world structure which performance loss is more tolerant of criticisms on attacked stations.

It is interesting to notice that the most critical stations are not the central ones, both in

Title:/Users/fabiolamanna/Documents/Do
 Creator:MATLAB, The Mathworks, Inc. Vers
 CreationDate:01/30/2008 22:51:14
 LanguageLevel:2

Figure 9.1. Attack and Failures on Layers in the Space of Stations.

Title:/Users/fabiolamanna/Documents/Do
Creator:MATLAB, The Mathworks, Inc. Vers
CreationDate:01/30/2008 23:20:54
LanguageLevel:2

Figure 9.2. Attack and Failures on Layers in the Space of Stops.

terms of geographic and strategic position within the city. In this analysis based on travel times and frequencies of services, the stations positioned on the 'ring' of Berlin are the most important for the global behaviour of the whole system. Critical stations are the same for all the analysed spaces. This means that a failure or an attack at these nodes is critical both for the infrastructural connections (service providers' point of view) and for the most efficient paths of travel (users' point of view). A service provider can take an advantage of knowing which nodes need to be better protected than others in order to minimise the impact of an accident or of a failure.

Different percentages of dropping efficiency for different layers show that by removing

Title:/Users/fabiolamanna/Documents/Do
Creator:MATLAB, The Mathworks, Inc. Vers
CreationDate:01/31/2008 08:42:13
LanguageLevel:2

Figure 9.3. Attack and Failures on Layers in the Space of Changes.

the most influential node of the S-Bahn network almost the half value of the global performance of the whole system is cut. Besides, the complete layer consisting of all the three rail networks is the most robust even in case of failures or attacks on the most important node. The *small world* behaviour of the Space of Changes scenarios takes an advantage of the global performance of the system in case of criticisms in the main nodes. Other space scenarios demonstrate the *scale-free* tolerance of targeted attacks because of their relative higher loss of efficiency in comparison to the other structures.

Figures 9.1, 9.2 and 9.3 show the drop of E_{TF} for each layer in the three different spaces of services in case of random failures or attacks to the nodes with higher connections. In the analysis the two threat models described in par. 7 have been applied. Random deletion of nodes simulates a failure in the system, while descending degree order deletion has been implemented in order to simulate targeted attacks on the most connected nodes. Threats to all layers play a different role in different spaces: the analysis done on SpS and SpST shows the high tolerance of random failures and the lower of targeted attacks, i.e. the efficiency drops only after the first deletion of the most connected node. Further deletions have only slightly consequences on the remaining infrastructural layout, in term of capacity in communication between nodes.

Besides, the analysis of the system under the users' point of view (SpC) has shown that the *small world* architecture takes an advantage both in terms of stability under random failures and under targeted attacks. All layers have almost the same grade of tolerance under accidental failures but, in case of a deliberate attack, the efficiency of the system doesn't drop so quickly as in the previous spaces. After the first intentional deletion, the system is able to perform a relative high Level of Service even when the second and the third most connected nodes are deleted.

According to the analysis performed to find the most critical stations of the network (Table 9.2), the S-Bahn layer is the most critical both on a topological and a service level. As shown in Figure 9.3 this layer's tolerance becomes critical only after the first attack on the most connected node. In case of an intentional offensive this network would be the first one to drop its ability in maintaining a minimum or desired Level of Service.

10. Conclusions and Further Research.

In this paper a new approach on transportation network robustness has been proposed. The complex network approach previously introduced helped in obtaining several results describing the behaviour of a modern urban network in case of threats by means of a tool developed by the author (Time=net.work). Results are related to travel times and frequencies of services scheduled on the rail network of Berlin and allow to define:

- A Protection Strategy, i.e. the most critical stations leading to a system loss of the higher Level of Service in case of failure or attack, both under the topological (infrastructural) and the users' point of view. Results show that the most important nodes are not the most central ones in a global view of the city network but the ones which are most important in order to allow a fast communication within the network.
- The reaction of all the network layers in case of random failures or deliberate attack against the most connected nodes. This allows to define the tolerance of each layer of different attacks. *Scale-free* and *small world* network models helped in associating a particular level of tolerance against threats with every layer and space of service.

- In this test case, all networks are relatively more tolerant to failures than to targeted attacks under an infrastructural point of view. Besides, an attack on a station can slightly affect the communication and the Level of Service of users, who are still able to travel and to change transport mode throughout the network even after three of the most connected nodes have lost their functionality.

Further research is now directed to perform more analysis including road services as Bus and Tram. Moreover, more attack strategies have to be developed in order to predict the reaction of the system under threats. Finally, a more accurate weighted analysis can take into account also transfer time between transport modes in order to define the most critical elements under this point of view.

11. Bibliography.

- [1]. Albert R., Jeong H., Barabási A.-L. (2000) "Error and attack tolerance of complex networks", *Nature*, 406, pp. 378-382.
- [2]. Barabási A.-L. (2002) *Linked: The New Science Of Networks*, Perseus, Cambridge, MA, USA.
- [3]. Barabási A.-L., Albert R. (1999) "Emergence of Scaling in Random Networks", *Science*, 286, 509.
- [4]. Bell M.G.H. (2000) "A game theory approach to measuring the performance reliability of transport network". *Transportation Research Part B* 34, pp. 533–545.
- [5]. Cohen R., Erez K., Ben-Avraham D., Havlin S. (2000) "Resilience of the Internet to Random Breakdowns", *Phys. Rev. Lett.*, 85, pp. 4626-4628.
- [6]. Ezell B.C, Farr J.V., Wiese I. (2000) "Infrastructure risk analysis model". *Journal of Infrastructure Systems*, 6 (3), pp. 114–117.
- [7]. Gallos L.K., Cohen R., Argyrakis P., Bunde A., Havlin S. (2005) "Stability and topology of scale-free networks under attack and defense strategies", *Phys. Rev. Lett.*, 94, 188701.
- [8]. Holme P., Kim B.J. (2002) "Attack vulnerability of complex networks", *Phys. Rev. E*, 65, 056109.
- [9]. Jenelius E., Petersen T, Mattson L.-G. (2006) "Importance and exposure in road network vulnerability analysis". *Transportation Research Part A* 40, 7, pp. 537–560.
- [10]. Kurant M., Thiran P. (2006) "Trainspotting: extraction and analysis of traffic and topologies of transportation networks", *arXiv:physics/0510151 v2*.
- [11]. Latora V., Marchiori M. (2001) "Efficient Behaviour of Small-World Networks", *Phys. Rev. Lett.* 87, 19.
- [12]. Latora V., Marchiori M. (2004) "How the science of complex networks can help developing strategies against terrorism", *Chaos, Solitons and Fractals*, 20, pp. 69-75.
- [13]. Latora V., Marchiori M. (2005) "Vulnerability and protection of infrastructure networks", *Phys. Rev. E*, 71, 015103.
- [14]. Nagurney A., Qiang Q. (2007) "A transportation network efficiency measure that captures flows, behaviour and costs with applications to network component importance identification and vulnerability", In: *Proceedings of the POMS 18th annual conference, Dallas, USA, May 4th-7th 2007*.
- [15]. Schreuder M.A., Molenkamp L., Tamminga G.F., Kraan M.E. (2007) "Vulnerability of a national road network", In: *Proceedings of The Third International Symposium on Transportation Network Reliability, The Hague, The Netherlands, July 19-20 2007*.

- [16]. Von Ferber C., Holovatch T., Holovatch Y. (2007) "Attack Vulnerability of Public Transport Networks", *arXiv:physics/0709.3206* v1.
- [17]. Watts D.J., Strogatz S.H. (1998) "Collective dynamics of 'small world' networks", *Nature*, 393, pp. 440-442.
- [18]. Zhang Y., Xiang P., Yang X. (2007) "Error and attack tolerance topological analysis on urban road networks". In: *Proceedings of The Third International Symposium on Transportation Network Reliability, The Hague, The Netherlands, July 19-20 2007*.