A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

1

A Complex System's View of Critical Infrastructures

V. Rosato[1,*], I. Simonsen[2,3,†], S. Meloni[1], L. Issacharoff[1], K. Peters[2], N. von Festenberg[2] and D. Helbing[2]

[1] ENEA, Casaccia Research Center, Computing and Modelling Unit, Roma, Italy
[2] Institute of Transport and Economics, Dresden University of Technology, D-01086 Dresden, Germany
[3] Department of Physics, The Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

1.1 Introduction and Motivation

Our contemporary societies are examples of highly complex systems with many interacting constituents that are organized in ways that often are hard to grasp. Their organizational systems and infrastructures are time-dependent and highly interconnected. Thus, what may appear as different parts of our societies, do indeed depend on and influence each other.

Large Complex Critical Infrastructures (LCCIs) are national — or international — technological systems whose correct functioning has a high social impact. A current definition of a "Critical Infrastructure" is a large scale infrastructure which if degraded, disrupted or destroyed, would have a serious impact on health, safety, security or well-beings of citizens or the effective functioning of governments and/or economy [1]. This definition therefore allows to label many infrastructures that we are well-familiar with from our daily lives, as being "critical". Among them are, for instance, the networks for the transmission and the distribution of electrical power, those allowing communication to occur (in all its forms, from telephones to the Internet), transportation networks like roads, railways and sky-routes up to pipelines for drinking water, gas and oil, *etc.* LCCIs are thus strategic (in the wider sense of the term); as such, an enormous care should be taken to keep them operational and efficient, preventing their failure due to accidents or intentional attacks.

A further major issue comes from the high level of *interdependency*, *i.e.* the fact that each LCCI interacts (in a more or less explicit way) with one another. This may have the implication that a disturbance in one of them might affect

[*] Also with Ylichron S.r.l., Roma, Italy.
[†] Current address: Saint-Gobain Recherche, Aubervilliers, France.

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

the functionality of others. This renders the task of preventing failures and, in general, the same operational control, an extremely complex task. It is indeed desirable to have the best possible control on *single* infrastructures in order to prevent faults. However, optimizing and securing individual infrastructures independent of the presence of others, is often *not* sufficient to securing such interconnected system.

LCCIs are also intriguing technological objects. They are "complex", according to the current definition of complexity, as their behavior cannot be simply predicted on the basis of the behavior of their single components. Complexity triggers the *emergence* of new phenomena which cannot be predicted by usual means but only through a complete description of all its components altogether. Emergence of new phenomena occurs, *a fortiori*, when many LCCIs are functionally coupled together: also in the case of a weak connection, there is the seed for the emergence of further unpredictable behavior.

All this conceptual entanglement has attracted the interest of the Complexity Science (CS) community. This work intends to introduce some basic statements, show the CS methods and tools and some recent results of their application in the field of LCCIs. In this chapter, we intend to make a first recognition of some basic problems which can be tackled by making use of mathematical models and numerical methods, with the aim of producing results useful for the understanding of some fundamental questions related to their structure.

1.2 Why LCCIs Become/Behave More and More Complex

Historically, in Europe (at least), the LCCIs were often national monopolies typically owned and/or controlled by the national governments. Over the last decades, this situation has changed to a large extent; many LCCI sectors have been *deregulated* and thus the monopolistic state removed. This opened up for new market players that together with the former monopolists (of a given region) could compete. Notice that this situation was not (in principle) restricted to a geographical (national) region, but also international competition was encouraged by the market liberalization. For instance, one prominent example of the latter is the European power market. The formal basis of the deregulation of the European electricity market was laid out in the 1996 EU Directive 96/92. However, about three years later, on 19 February 1999, the electricity market in thirteen countries in the European Union (EU) and the European Economic Area (EEA) began to open up on an international basis. A competitive European Power market was born!

With the deregulation of the European LCCI sector, new challenges were created. Now (big) consumers could, say, buy their electrical power from any market participant. This implied that the backbone of the European transmission grid had to be fully interconnected, and that it should be able to handle

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

rising loads. However, the European transmission grids were not designed for this purpose (and volume) in mind. Connections to neighboring states were typically built up for backup reasons, and to handling short term import-export scenarios. Hence, the new business model that was put in place (due to the liberalization) prompted some technically minded people to question the robustness of the ever more complex power transmission grid. This concern became strengthened by the increased terrorism threat as well as the recent large-scale Italian September 2003 blackout, and the similar previous cases from London, North America, Sweden and Denmark. For instance, the cause of the London blackout was traced back to a badly-installed fuse at a power station; indeed all the others happened for similar reasons. Furthermore, it was realized, by a careful analysis of the cause of events, that problems typically start at one place and propagate over large geographical distances, like a domino effect. For instance, the great 2003-blackout in New York initially was triggered by an event in the mid-west (Ohio) [2, 3].

Analogous problems must be faced in telecommunication (TLC) systems, where a large number of stakeholders crowd common infrastructures and compete for bandwidth and customers. TLC routes are constantly stressed by a constantly increasing traffic level.

Most LCCIs have grown in an unsupervised regime (there is not a general controller of worldwide Internet network) and needs to face a dramatic increase of their usage by adopting an "intrinsic" ability to adapt themselves according to changing external conditions. This seems to be a key point in this matter: Are technological networks able to autonomously react to external input in a way to adapt their functioning to constantly guarantee a reasonable efficiency? If so, which are the agencies that allow adaptive behavior to occur? What can LCCI managers reasonably do to let adaptation mechanisms run faster and more efficiently, and to better respond to mutated external conditions?

Complexity Science tries to answer these questions also by identifying common scenarios which subtend rather "universal" behaviors which take place in complex systems. This approach has allowed a flow of data and methods from diverse scientific fields and triggered the customization of ideas and methods, typical in one domain, to other domains. Living objects, for instance, as bacterial colonies, swarms and bird flocks do display a number of intriguing control strategies which, if properly understood, could be mutuated and used to analyse and control technological systems (*bio-mimetic* strategies).

This scenario has prompted calls for improved coordination between basic and applied research on the evaluation and the design of new tools for the analysis and the control of LCCIs at a multi-national level. For instance, many EU funded projects have been launched within this domain in the sixth Framework Program (FP6), and similar projects have received public funding in the US. Dedicated programs within this area are forecasted also for FP7.

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

In the remaining part of this chapter, we will introduce and discuss LCCIs from a complex network perspective. In particular, the *graphs*, which are the network's basic model and will be the object of the present study, will be introduced. They are mathematical objects onto which CS can deploy its methods whose results can provide measures of their ability of providing useful information on the networks. Various analyzing methods suitable for LCCIs will be proposed and discussed. Examples of results that can be obtained from such analysis will be given for some example LCCIs.

1.3 What is A Network?

The term network is used in every-day life, so most of us have an impression of what is meant by it. As we will attempt to reply to basic questions on these networks, also the metaphors which will be used to describe networks will be at a high level of abstraction. However, here we will put a specific meaning to that word. By a *network* we will mean a set of N objects, referred to as *nodes* or *vertices*, that are connected through what is typically known as *links*, *arcs* or *edges* (L). A network G will thus be indicated as a collection of objects $G = (N, L)$ (in these terms, the network can be also represented as a mathematical *Graph* which is indicated by explicitating the same entities). Some simple networks are sketched in Fig. 1.1. In this figure, the nodes are indicated by red (or grey) filled circles, while the links are black lines between the nodes. For instance, for an electrical power network, nodes correspond to power generators or distribution (or transmission) stations, while links represent the power transmission lines connecting the nodes.

It should be noticed that links do not have to be *physical* connections; they might also represent *logical* connections between nodes, such as in the case of a so-called social network. Here nodes are persons, and a link exists between two persons if they are considered (in some way) to be friends.

Networks represent the natural starting point for modelling infrastructures. As mentioned above, the investigation of networks has received an increasing attention in the last decade in the CS community. Genuinely, the field was embodied in mathematics as *graph theory*, which, due to progress in computer power and the growing consciousness of the relevance of network's structure in several fields, has prompted this topics to a wider scientific audience. The field experienced a strong push forward when the famous "small-world" paper by Watts and Strogatz was published in 1998 [5] by delivering examples of networks where seemingly distant nodes actually are surprisingly close to each other due to the peculiar network structure; this property has been called "small-world" as it perfectly reproduces the situation of a globalized world where local events can have a long-range impact. This property has also been fixed in the common language by the phrase "six degrees of separa-

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

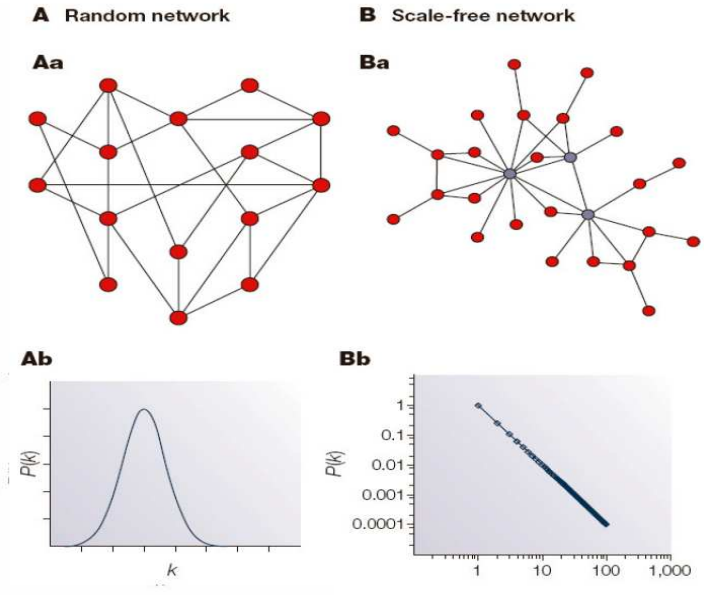

Fig. 1.1. Examples of complex networks. Fig. 1.1(A) depicts a *random network*, while a scale-free network is shown in Fig. 1.1(B). The typical degree distribution, $P(k)$ of each class of network is shown in the lower part of the figure, *i.e.* the distribution of the number of links, k that is associated with each node. Notice in particular the marked difference in topology that results from the change in the degree distribution. (After Ref. [4])

tion"[4]. It expresses that in a small-wold class network, two arbitrary chosen nodes can be connected in, on average, only six steps. For social networks, this effects (as well as the number of six) had already been known empirically since the 1960's due to some cleverly designed experiments conducted by the social psychologist Stanley Milgram [6]; with a small chain of friends and friends-of-a-friend, each of us can reach whatsoever other person in the world in (on average) six steps.

A major outcome of this new branch of theoretical disciplines is the recognition that diverse networks (from sociology, technology and biology) display a peculiar structure with clear small-world characteristics. This seems to be a property *emerging* from complexity and, as such, probably brings some added-value to the network's property. Much work has already been performed in order to show which are exactly the benefits engendered by such a topological structure.

[4] The phrase was made well-known outside scientific circles by John Guare's popular theater play of the same name (and later movie).

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

Recently, several comprehensive reviews on network research (graph theory) have appeared in the literature [7–10] displaying the current state of its application to real world networks. Most of the work, so far, has focused on static properties and behaviour of networks, *e.g.* the question of network robustness [11].

The main property of a network stems from its classification as belonging to a specific topological class. These are related to the specific form displayed by the distribution of the node's *degree*, k, of the network, $P(k)$ (degree distribution). The degree, k, of a node is defined as the number of nodes to which it is *directly* (physically or logically) connected. The most relevant topological classes are:

- Random networks
- Scale-Free networks

In the first case (see Fig. 1.1A), $P(k)$ has a Poissonian shape; the network is thus composed of *almost* equivalent nodes, with an average degree $\langle k \rangle$ and a given standard deviation. In the second case (see Fig. 1.1B), the situation is more complex, as $P(k)$ follows a power-law, *i.e.*

$$P(k) \sim k^{-\gamma}, \tag{1.1}$$

where γ is a real positive constant which has been found to take values typically in the range $2 < \gamma < 3$ [9]. This situation occurs when nodes are highly non-equivalent. Such networks have been named Scale-Free (SF hereafter) because a power-law has the property of having the same functional form at all scales. In fact, power-laws are the only functional forms $f(x)$ that remain unchanged, apart from multiplicative factors, under a rescaling of the independent variable x. They are the only solutions to the equation $f(\alpha x) = \beta f(x)$. SF-networks, having a highly inhomogeneous degree distribution, result in the simultaneous presence of a few nodes (the *hubs*) linked to many other nodes, and a large number of poorly connected elements (the *leaves*). Each of these network-classes occurs in specific cases; there are, however, other topological classes which will be referred to, in the following, when they will be eventually mentioned. Up to the eighties, the current opinion was that practically all networks representing real world structures (from social to technological networks) could be ascribed to the class of *Random* networks. After all, they were thought of as resulting from unsupervised growth processes and, as such, believed to be produced by a growth mechanism where new nodes stuck randomly to existing nodes (random-growth mechanisms). Relevant studies, at the end of the last decade, have shown the inadequacy of this scenario to represent the topological features of real networks: they have demonstrated that, although resulting from unsupervised growth processes, a large number of networks grow under the action of some *effective selective pressure* whose resulting effect is the realization of a structure more appropriately ascribed to the SF class [7] .

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

From the knowledge of the network's graph, many different topological properties can be deduced which further specify the network's properties and characteristics. These data allow to design specific *growth mechanisms* able to design networks with desired topological properties. For comprehensive reviews on the proposed growth mechanisms to reproduce networks with different topological structures, the reader is referred to Refs. [7, 9].

1.4 Critical Infrastructures as Networks

In this work, we will attempt to analyze available data of several CIs by using the methods and the ideas of topology analysis. According to the definition of CIs given previously, the following technological infrastructures may be certainly ascribed to the CI set:

- Public power supply networks
- Telecommunication networks
- The Internet

In the following sections, we will apply the methods of graph analysis to the graphs resulting from the available data of the technological networks of the above mentioned CIs.

1.4.1 Public Power Supply Networks

The Power Grid

The public power supply network transmits power from generation to loads thereby providing the link between producers and consumers. The network connects large numbers of generators and loads together thus *(i)* improving the reliability of the power supply, *(ii)* reducing needs for reserve, peak, control and storage capacity, *(iii)* enabling more efficient and economic power production, and *(iv)* providing a necessary platform for the electricity market. The strengthening of the cross border transmission capacity has made the public power supply network increasingly international and spatially very extended. The power supply network is an essential, but often very international part of the national critical infrastructure.

The power supply network is hierarchically organized to transmission and distribution networks. Transmission networks cover very wide geographical areas, and have typically very high voltage levels and large power flows. Distribution networks, on the other hand, connect the loads and distributed generation with the transmission network. The distances are traditionally shorter and the voltage level lower than in the transmission network. Distribution networks are normally organized in a radial way, although redundancy is provided by a meshed network topology. Low voltage customers are connected to the distribution network via low voltage distribution networks.

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

About one third of the cost of power supply comes from the distribution of power. The power distribution network has also a much higher impact on the reliability and power quality than the power transmission network. Failures on the transmission network are relatively rare, but their impact spreads over much wider areas than those occurring on distribution networks.

A power network is characterized by the fact that it has very little buffering storage capacity and the physical balance of supply and demand must be maintained, otherwise the power transmission system will collapse. The deregulated electricity market is an important tool for finding a cost efficient initial solution for this balancing problem. Operation of the power network is highly and increasingly dependent on protection, automation, information and communication systems.

Distributed generation, intermittent generation from renewable energy sources (*e.g.* wind energy), pressures to cost reduction and power quality improvements, ever bigger generation and transmission units are expected challenges to the power network. This calls for significant changes in the power distribution systems and their automation and operation, but the power distribution systems have much inertia. The required lifetime of the power network related investments is very long. Thus rapid fundamental changes are seldomly possible.

Topological Analysis

From the topological point of view, the graphs representing electrical transmission networks cannot be properly ascribed to *random* nor to SF networks. In fact, as it happens for networks whose structure is constrained (*i.e.* by geographical reasons) or that cannot present an arbitrarily large node's degree (as for roads, for instance, where there is a very low maximum degree), electrical networks have a Gaussian shape, with a heavy exponential tail that drops the values of the highest degrees to smaller numbers (for electrical transmission lines the maximum degree of a node is usually of the order of 10) [12–14].

The electrical network which has been widely studied in recent years, and which will also be the object of the present analysis, is the Italian high-voltage (380 kV) electrical transmission network (HVIET hereafter). A graph of HVIET, as deduced from publicly available data, is depicted in Fig. 1.2, and it consists of $N = 310$ nodes and $L = 361$ links (transmission lines). In fact there are different node types; generators (117), loads (139), and junctions (54), but a distinction between them will not be made in our analysis. Moreover, 14 (of the total 361) links are *double* (transmission) lines.

Several topological analysis have been performed on the HVIET network. One of the relevant properties of the network, allowing to classify the topology of the network, is constituted by the distribution of the node's *degree* k (the degree is the number of links connecting each node to its nearest neighbors). The distribution of node's *degree* of HVIET is reported in Fig. 1.3 which

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

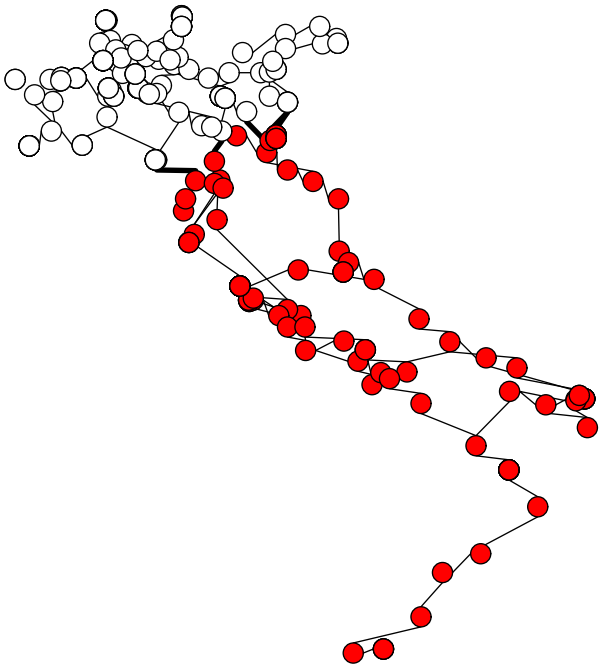

Fig. 1.2. Graph of the Italian high-voltage (380 kV) transmission network, where nodes are located at approximately correct geographical location. The 6 links located in the central region of Italy and represented by thick solid links correspond to the first critical section (min-cut selection) that divides the network into two almost equal-in-size parts (as indicted by the dark and white node symbols). Different node types (generators, loads and junctions) are not distinguished.

confirms that HVIET does not show neither a clearcut SF nor a random character. The network has a limited number of hubs, whose maximum degree is $k_{max} = 11$. Another property which has been measured on the HVIET network is the average *clustering* coefficient C[5], which measures the propensity of nodes to form small-scale communities (*c.f.* Refs. [7, 8] for additional details and formal definition). The clustering coefficient C is large when nodes, neighbors of a common node, are also neighbors of each other, *i.e.*, if node 1 is connected to node 2, and 2 to 3, then, if C is large, there is a relatively high probability that node 1 is also connected to node 3. Hence, we see that C measures in some sense the (relative) number of *connected* triangles in the network. In the HVIET network, the tendency to form connected triangles is rather small, and the clustering coefficient is as small as $C = 2.06 \times 10^{-2}$ (we

[5] Notice that some authors refer to this same effect as network transitivity [8].

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

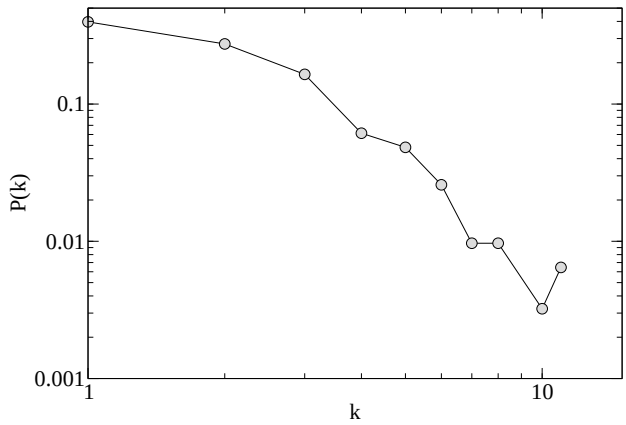

Fig. 1.3. The degree distribution, $P(k)$ *vs.* node degree, k, (in log-log scale) for the HVIET network depicted in Fig. 1.2.

will later see in Sec. 1.4.2 that, for instance, the clustering in the backbone of the Internet can be orders of magnitude higher).

An interesting result on HVIET has been evaluated by using the min-cut theorem associated with the spectral analysis of the so-called *Laplacian* $\mathcal{L}$. In order to define this matrix, let us start by introducing the adjacency matrix, $\mathbf{A}$, who's matrix elements, A_{ij}, take the value 1 if node i and j are connected, and 0 otherwise [8]. Then, in terms of $\mathbf{A}$ the (symmetric) Laplacian matrix is defined according to

$$\mathcal{L}_{ij} = \begin{cases} \sum_{k=1}^{N} A_{ik}, & \text{if } i = j \\ -A_{ij}, & \text{if } i \neq j \end{cases} \tag{1.2}$$

An interesting result that can be obtained from the spectral analysis of $\mathcal{L}$ can be stated as follows: The signs of the components of the eigenvector associated with the first non-vanishing eigenvalue of the Laplacian allow to optimally bisecate the network. As $\mathcal{L}$ is symmetric, the first eigenvalue is always vanishing. The n components of the eigenvector $\mathbf{v}_2^{\mathcal{L}} = (v_1, v_2, ..., v_n)$ associated with the second eigenvalue, solve the one-dimensional *quadratic placement* problem of minimizing the function

$$z = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (v_i - v_j)^2 A_{ij}. \tag{1.3}$$

The vector is subjected to the constraint $|\mathbf{v}| = (\mathbf{v}^T \mathbf{v})^{\frac{1}{2}} = 1$ [15]. The process allows to partition the graph $G = (N, L)$ into disjoint subsets G_1

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

and G_2 such that $L_{12}/(N_1 \cdot N_2)$ is minimized, where L_{12} is the number of links to be removed and N_1 and N_2 are the number of nodes in the two resulting subnetworks. It comes clear that this procedure allows for the "optimal" bisectioning of the graph, *i.e.* it forms the closest possible subnetworks G_1 and G_2 with the minimum amount of broken links L_{12}. If one applies the min-cut procedure, one gets the following bisection: the HVIET network is divided into two, connected, parts HVIET_1 and HVIET_2. The first is formed by $N(\mathrm{HVIET}_1) = 195$, and the second by $N(\mathrm{HVIET}_2) = 115$ nodes. The two parts are separated by *only* six links; The removal of these links allows to totally bisecate the network, which would separate it into two, non-communicating, parts (see Fig. 1.2). The ideal line, joining the location of the removed links, has been called *first critical section*. Indeed, there are several other lines of cut of the network, grouping a set of links whose removal produces a bisection of the network. These sets, although being composed (in some cases) by a lower number of links, do not minimize the function of Eq. (1.3) and, ultimately, are less efficient in separating the graph into two (almost) equal parts. This is a major outcome of the spectral analysis; this provides a way to locate the critical vulnerability lines of the network. The procedure can be iterated on the different components of the graphs, by creating *critical sections* of higher orders. If on a simple graph, the min-cut procedure can be almost done by visual inspection, for larger graphs it cannot be simply performed by any other mean.

Relevant information on the *robustness* of a network can be gained by simulations. Starting from the graph structure, for instance, one can evaluate what is the probability of *physically* disconnecting one (or more) nodes by disconnecting one (or more) lines. This will produce a qualitative evaluation of the *structural* robustness of the network. The knowledge of further technical details on the network (*i.e.* the electrical characteristics of lines) allows the formulation of a *dynamical* model for the power transport on the network. A recent work [16] has attempted to reproduce the flow on the HVIET produced by a given gauge of injected/extracted electrical power by/from the different nodes and by the real electrical admittance of the different electrical lines. The availability of such information opens the way to evaluate the so-called *flow* vulnerability. If one eliminates a given number of lines, the dynamical model allows to evaluate the new flux distribution; in case of overtaking given threshold of maximum flux on the lines, the flow equations are re-evaluated by starting from a different gauge of injected/extracted electrical power. When relevant lines are missed, the network must undergo a severe reduction of the injected power in order to be able to correctly sustain the power flux. If one associates the amount of power reduction to re-establish the flux to the specific removed line, one can classify the different lines as a function of the damage that their absence produce to the whole network. If applied to HVIET, this procedure allows to obtain a classification of the lines as a function of the damage which their absence is able to produce (which can be as large as 1.5GW, see Fig. 1.4).

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

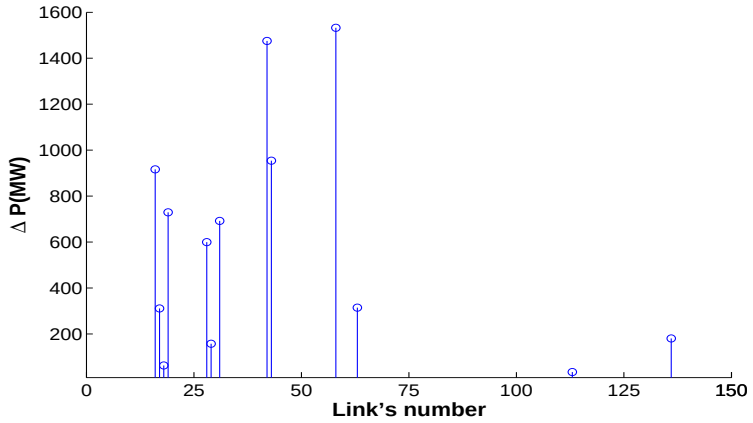

Fig. 1.4. Lines of HVIET whose removal is associated to the largest injected power reduction: The illustration shows on the abscissa the number of the power line, on the ordinate the amount of injected power (in MW) to be reduced to re-establish a correct power flux in the network.

1.4.2 The Internet

Organizational Issues

The Internet should be known (and appreciated) by all of us, and therefore probably does not need any further introduction. In the following, by the term "the Internet" we will be referring to the network formed by the so-called *Autonomous System* (AS) router level [17]. An AS is a collection of IP networks and routers under the control of one entity (or sometimes more) that presents a common routing policy to the Internet. Therefore any sub-network appears as an AS, and the important difference between *Intra-AS* routing and *Inter-AS* routing must be introduced. The entity that controls an AS can choose the routing protocol to be used inside it, so in general AS can use different routing protocols. But in order to make interconnectivity between AS possible, each AS must employ one or more routers to interface with the "outer world", in order to informing it of the AS presence and topology. Usually there are specifically designated routers dedicated to accomplish this task — the so called *Border Routers*. Clearly these routers must adhere to the Internet rules and protocol set (explained further on). Thus, the AS-level routers form the backbone of the Internet which speaks the same language (*i.e.* adhere to the same protocol).

Since the first Internet connection was made on June 6, 1969, its size and complexity has grown dramatically. A recent paper examined the growth rate for nodes (g_n) and links (g_l) of such a network during the end of last decade [12], and it was found that $g_n \sim 140$ nodes/month and $g_l \sim 300$

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

Data set	N	L	γ	C	k_{max}	d	$\langle d_{ij} \rangle$	$\langle d_{ij}^{rand} \rangle$
DIMES	14154	38928	2.41	0.41	1932	9	3.343	5.606
ROUTEVIEWS	11461	32730	2.35	0.35	2432	9	3.565	5.712

Table 1.1. Relevant topological properties of the DIMES and the RouteViews Internet network data. N denotes the number of nodes of the network, L the number of links, γ the exponent of the degree distribution (see Eq. (1.1)), C the clustering coefficient, k_{max} the maximum degree of a node (the largest hub of the network), $\langle k \rangle$ the average degree, d the diameter of the network (the largest inter-node distance) and $\langle d \rangle$ the average node's separation, $\langle d_{ij}^{rand} \rangle$ the average node's separation of a *random* network of equal N and $\langle k \rangle$.

links/month. It is worth recalling that a new AS-level router corresponds to the introduction of a new subnetwork which can also contain thousands of (internal) nodes.

Topological Data

To get *accurate* data on the topology of the Internet is difficult. In fact, the Internet should be measured "from its inside", since no one has the complete, up-to-date map of it. This need has prompted a number of large-scale projects aimed at "mapping" the Internet in the most accurate way. Examples of such projects are the DIMES [18] and the RouteViews Projects [19].

Data which will be referred to in the present work have been collected from the DIMES project funded by the EU. They refer to a snapshot of the map taken at a given date (July 2005). A repository of several snapshots, collected at different times, is also contained in the projects web site. These are useful in order to monitor the growth of the network (or, at least, its time variation); These data could be used to infer growth mechanisms underlying the time variation of the network's properties (size, degree, clustering *etc.*) [12].

Other less accurate data sets of the Internet, but covering a larger geographical region, can be found in *e.g.* Ref. [20].

Topological analysis, and Network Growth Models

Topological analysis to evaluate the major topological properties have been performed on the DIMES network and, for comparison, on similar data taken from the repository of a US-funded project (RouteViews). Data collected on the two sets of data (DIMES and RouteViews) are reported in Table 1.1.

Several points of Table 1.1 need to be highlighted. First of all the small characteristic path length $\langle d_{ij} \rangle$ (which should be compared to its predicted value for a *random* network of a similar dimension $\langle d_{ij}^{rand} \rangle$ ($\langle d_{ij}^{rand} \rangle = \log N / \log \langle k \rangle$). This is a quite controversial point in the literature. Standing on their analysis, some authors have claimed an Internet characteristic path length higher than that predicted for a random networks [21], others

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

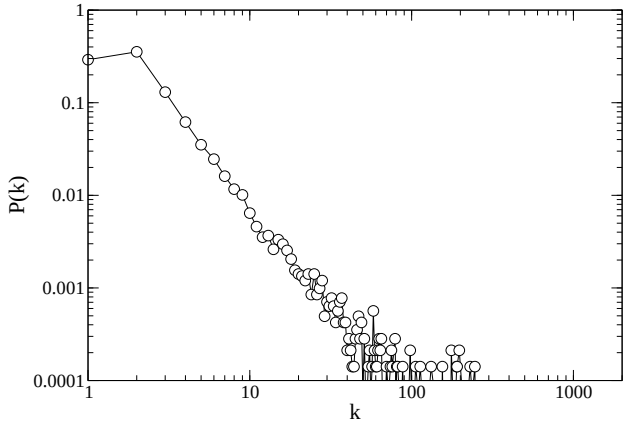

Fig. 1.5. The degree distribution, $P(k)$ *vs.* node degree, k, (in log-log scale) for the DIMES data.

have measured a slightly lower diameter [7]. Then the large *clustering coefficient*, C, of the network which measures the propensity of nodes to form small-scale communities (Refs. [7, 8]). SF networks *do* not necessarily have large C values. Thus is a peculiar feature of the Internet network and of many social networks [7]; other Critical Infrastructures (such as Power grids) do not share this feature.

DIMES shows, as expected, a distribution of node's degree which properly fits a power-law with exponent $\gamma = 2.41$ (Fig. 1.5).

In order to summarize the observations made concerning the node's degree distribution and of the large *clustering* coefficient, we have attempted to define an "empirical" growth mechanisms allowing to reproduce the Internet topology. We succeded in this task by using a suitable combination of the Preferential Attachment (PA) [7] and the Triad Formation (TF) [22] mechanisms, the first allowing a SF network to be produced, the second able to account for an arbitrarily high value of the *clustering* coefficient. If, in a growth mechanisms where, starting from an initial set of n nodes, we wish to add a new node, we define that $P_{(n+1)\to j}$ is the probability that the new node $(n+1)$ sticks on the node j belonging to the network, it will be as follows:

$$P_{(n+1)\to j} = (1-\beta)PA^{\alpha} + \beta TF. \tag{1.4}$$

It means that the new node will stick with a probability $(1-\beta)$ with a modified PA algorithm (indicated as PA^{α} or with probability β with a TF mechanism) [23]. The value of the parameters providing the best agreement with the DIMES data set are: $\alpha = 1.44$ and $\beta = 0.93$.

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

Our speculations follow a previous attempt made on the issue of modeling the Internet's large-scale topology [24]. The authors pointed on a modification of the PA mechanisms by introducing a further dependence on the *distance* among nodes: highly connected nodes are favoured if geographically close. With this assumption, links to far away nodes are discouraged while clustering is favored because node's proximity tends to enhance the establishment of links particularly among neighboring nodes.

The Random Walk Approach

In the previous sub-subsection, we saw that one could characterize the "clusteredness" of a network by, *e.g.*, the clustering coefficient C. However, given a network topology, how can one identify the nodes belonging to the same cluster? For large networks, like the Internet, this is a highly non-trivial (and often computationally daunting) task. Recently, several dedicated numerical algorithms have been proposed with this purpose in mind [7–9, 25–30]. Here we do not intent to present a full overview of such *clustering-algorithms*, but instead outline a particular approach based on diffusion or random walkers.

To motivate this algorithm in simple terms, let us consider the following mental image; Assume the (very hypothetical) scenario that a car driver is located randomly somewhere in North-America, without the ability to gain information about direction from traffic signs, maps *etc.* Whenever he approaches a cross road, he randomly picks (with equal probability) one of the possible connecting roads. In this way, the driver randomly moves around on the road network without being assisted by any directional information that we all are so used to benefiting from. If the main aim of our "random driver" is to reach a given destination in, say, South America, you can probably easily guess, that the drivers strategy is far from being optimal. The driver will most probably find himself driving around in North America for a very long time, simply because there are relatively few roads "connecting" North and South America. In other words, the random driver will spend most of his time in the "northern" cluster where he started off. There is only a small probability that he will find his way through the bottleneck, here represented by Central America.

If there is not only one (random) driver, but instead a large number of them, one may ask for the relative fraction of drivers being at a particular node i at time t. This fraction, or density, is simply $\rho_i(t) = N_i(t)/N$ where $N_i(t)$ is the number of drivers at node i at time t, and N is the total number of drivers. If the system is evolving according to the random dynamics outlined above, one may suspect that the walker density in highly connected regions of the network, *i.e.* within a cluster (if any), will reach an almost constant value much faster than in not so highly connected regions of the network. It was this suspicious that, in the first place, lead us to consider it as a candidate for a clustering detection algorithm.

A Complex System's View of Critical Infrastructures
Buchreihen Understanding Complex Systems
Verlag Springer Berlin / Heidelberg
ISSN 1860-0832 (Print) 1860-0840 (Online)
Volume 2008
Buch Managing Complexity: Insights, Concepts, Applications
DOI 10.1007/978-3-540-75261-5
Copyright 2008
ISBN 978-3-540-75260-8
DOI 10.1007/978-3-540-75261-5_11
Seiten 241-260

Given the underlying network topology, the process of the random drivers (or walkers) can easily be formulated mathematically, and the suggestions made above can be confirmed within a solid framework. The process is mathematically described by the "diffusion-like" equation [28–30]:

$$\partial_t \boldsymbol{\rho}(t+1) = \mathbf{D}\boldsymbol{\rho}(t) \tag{1.5}$$

where $\boldsymbol{\rho}(t)$ is the density vector of walkers at time t, and $\mathbf{D}$ a matrix that can be called the *diffusion matrix* (operator) for the system. This matrix is related to the adjacency matrix A_{ij} in the following way $D_{ij} = A_{ij}/k_j - \delta_{ij}$, where k_j refers to the degree of node j, and δ_{ij} is the Kronecker delta function. Notice that $\mathbf{D}$ is non-symmetric, unlike the adjacency and Laplacian matrices (1.2). The solution to Eq. (1.5) should be readily obtained as the linear combination of $\mathbf{v}^{(\alpha)} exp\big(-\lambda^{(\alpha)} t\big)$ where $\mathbf{v}^{(\alpha)}$ and $\lambda^{(\alpha)}$ are corresponding pairs of eigenvectors and eigenvalues, respectively, of the diffusion matrix. The index α is used to label the ordered sequence of eigenvalues so that $\alpha = 1$ corresponds to the largest one (that can be shown to be exactly one), $\alpha = 2$ to the next-to-largest one, and so on. Hence one realizes that the terms corresponding to increasing α's (where $\lambda^{(\alpha)} > 0$) correspond to faster-and-faster decaying modes of the system. The interpretation of this observation is that the largest α's different from one (the slowly decaying modes), can be related to the large scale topological features of the network. This has been demonstrated in recent publications [28–30] by plotting *e.g.* the *current of walkers*, $c_i^{(\alpha)} = \rho_i^{(\alpha)}/k_i$, leaving node i for an increasing number of modes α. For a given (small) mode $\alpha \neq 1$, the signs of the corresponding currents, $c_i^{(\alpha)}$, indicate a partitioning (into two parts) that may, or may not, correspond to a well-defined module or cluster for the network. To determine if a given partitioning can be characterized as a module we have used the so-called *modularity* measure. It is defined, given a (predefined) partition, as essentially the total number of links falling within modules minus the expected number of links for an equivalent network where links are placed at random [31–34].

If the modularity for a given partitioning is large, one says that the partitioning represents a "good" modular structure, otherwise not. By repeating this process for higher and higher (diffusive) modes, α, a rather rich community structure can be identified (*cf.* Ref. [30] for additional details).

We will now analyze the topology of the Internet by this random walk current mapping technique. In the following we will consider an AS-data set obtained from Ref. [20]. It consists of about 6 500 nodes from various parts of the world. Fig. 1.6 shows the 2-dimensional current mapping of the network using the two slowest decaying diffusive modes, *i.e.* $\alpha = 2, 3$. All AS-systems have been labelled with black dots. Later all nodes from some selected nations have in addition been labeled differently for convenience. The star-like structure indicates that there is a hierarchy of vertices where those located the furthest away from the origin of the current plot are the most peripheral vertices of the network. Furthermore, each hierarchy corresponds roughly to the