PERGAMON

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chaos, Solitons and Fractals 20 (2004) 69–75

CHAOS
SOLITONS & FRACTALS

www.elsevier.com/locate/chaos

How the science of complex networks can help developing strategies against terrorism

Vito Latora [a,*], Massimo Marchiori [b,c]

[a] *Dipartimento di Fisica e Astronomia, Università di Catania, and INFN, Via S. Sofia 64, I-95123 Catania, Italy*
[b] *W3C and Lab. for Computer Science, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, MA 02139, USA*
[c] *Dipartimento di Informatica, Università di Venezia, Via Torino 155, 30172 Mestre-Venice, Italy*

Abstract

A new method, based on a recently defined centrality measure, allows to spot the *critical components* of a generic complex network. The identification and protection of the critical components of a given communication–transportation network should be the first concern in order to reduce the consequences of terrorist attacks. On the other hand, the critical components of a terrorist organization are the terrorists to target to disrupt the organization and reduce the possibility of terroristic attacks.

1. Introduction

The science of complex systems is a new multidisciplinary field aiming at understanding the complex real world around us. Neural networks, artificial intelligence, traffic patterns, social and economic systems and many other scientific areas can be considered to fall into the realm of complex systems and can be studied by using nonlinear mathematical models, statistical methods and computer modeling approaches.

Since the September 2001 terrorist attacks, scientists and the policy community have focused on the ways in which the science of complex systems might be applied towards reducing the risk or consequences of future attacks. In this paper we discuss how some ideas and methods developed in the study of *complex networks* [1–7] can be successfully exploited to elaborate good strategies against (1) terrorist attacks and (2) terrorist organizations. In particular, we show how the information on the topology of a network, when available, can be used to spot what we define as *the critical components* of the network, i.e. the most important components for the efficient functioning of the network. Such a result can be used in many different ways. Here we propose two different applications by discussing two concrete examples.

(1) A generic communication (Internet, the World Wide Web) or transportation (cities connected by highways, by trains or by airplanes) system can be modeled as a graph (network), i.e. a set of nodes and links between couples of nodes. For instance in an Internet network the nodes indicate the cities with Internet access presence, and the links are the cables between pairs of cities. The identification of the critical components (nodes) of the communication network can be very important because those are the nodes whose protection from terroristic attacks must be assumed as the first concern of a national policy.

(2) Also a terrorist organization can be represented as a network. The terrorist are the nodes of such a network, and the links describe the interaction–collaboration relationships between pairs of terrorists. In this case the critical nodes are the terrorists to target if we want to disrupt the terrorist organization.

[*] Corresponding author. Fax: +39-95-383023.
E-mail address: vito.latora@ct.infn.it (V. Latora).

0960-0779/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0960-0779(03)00429-6

The method we propose is simply based on the knowledge of the architecture (the connectivity) of a network, and does not take into account the dynamics of the system. In a way this approach is complementary to the agent-based modeling [8].

2. The efficiency of a network

Our method to identify the critical components of a generic network i.e. the nodes that are crucial for the perfect functioning of the network, is based on an ad hoc definition of network performance: the importance of a node is measured by the drop in the network performance caused by the removal of that node. The other centrality measures of a node previously proposed are only based on local information such as the number of ingoing or outgoing links [9–11].

The network *efficiency* E, is a measure introduced in Refs. [5,6] to quantify how efficiently the nodes of the network exchange information. To define E we need to give some mathematical formalism. All the details can be found in Ref. [6]. A given network, for instance a communication network, or a terroristic organization network can be modeled by a *weighted* graph $\mathbf{G}$ with N nodes and K links (edges). [1] Such a graph is described by two matrices:

(1) the *adjacency matrix* $\{a_{ij}\}$, containing the information about the existence or not existence of a link, and defined as a set of numbers, $a_{ij} = 1$ when there is an edge joining i to j, and $a_{ij} = 0$ otherwise;
(2) the *matrix of physical distances* $\{\ell_{ij}\}$. In the case of a communication network this matrix contains the Euclidean distances between cities i and j. Of course, when it is not possible to associate a length or a weight to the links (case of an unweighted graph) $\ell_{ij} = 1\ \forall i \neq j$ and the graph is simply described by the adjacency matrix.

To define the efficiency of $\mathbf{G}$ first we have to calculate the shortest path lengths $\{d_{ij}\}$ between two generic points i and j. In a weighted graph d_{ij} is defined as the smallest sum of the physical distances throughout all the possible paths in the graph from i to j, and is computed by using the information contained both in matrix $\{a_{ij}\}$ and in matrix $\{\ell_{ij}\}$. Again, when $\ell_{ij} = 1\ \forall i \neq j$, i.e. in the particular case of an unweighted graph, d_{ij} reduces to the minimum number of edges traversed to get from i to j.

Let us now suppose that every vertex sends information along the network, through its edges. We assume that the efficiency ϵ_{ij} in the communication between vertex i and j is inversely proportional to the shortest distance: $\epsilon_{ij} = 1/d_{ij}\ \forall i,j$. Note that the assumption that efficiency and distance are inversely proportional is a reasonable approximation for all the systems considered in this paper. Of course, sometimes other relationships might be used, especially when justified by a more specific knowledge about the system. By assuming $\epsilon_{ij} = 1/d_{ij}$, when there is no path in the graph between i and j we get $d_{ij} = +\infty$ and consistently $\epsilon_{ij} = 0$. Consequently the average *efficiency* of the graph $\mathbf{G}$ can be defined as [12]:

$$E(\mathbf{G}) = \frac{\sum_{i\neq j\in\mathbf{G}} \epsilon_{ij}}{N(N-1)} = \frac{1}{N(N-1)} \sum_{i\neq j\in\mathbf{G}} \frac{1}{d_{ij}}. \tag{1}$$

Such a formula (1) gives a value of E that can vary in the range $[0, \infty[$, while it would be more practical to normalize E in the interval $[0, 1]$. The most natural way to do so, is to consider the ideal case $\mathbf{G}^{\text{ideal}}$ in which the graph $\mathbf{G}$ has all the $N(N-1)/2$ possible edges. In such a case the information is propagated in the most efficient way since $d_{ij} = \ell_{ij}\ \forall i,j$, and E assumes its maximum value $E(\mathbf{G}^{\text{ideal}}) = \frac{1}{N(N-1)} \sum_{i\neq j\in\mathbf{G}} \frac{1}{\ell_{ij}}$. The efficiency $E(\mathbf{G})$ considered in the following of the paper are always divided by $E(\mathbf{G}^{\text{ideal}})$ and therefore $0 \leqslant E(\mathbf{G}) \leqslant 1$. Though the maximum value $E = 1$ is reached only when there is an edge between each pair of vertices, real networks can nevertheless assume high values of E. This definition is valid both for unweighted and for weighted graphs, and very important for the purposes of this paper it can also be applied to disconnected graphs. The efficiency E has been calculated for many real networks [6,13] since it is well known that the harmonic mean is a better quantity to consider than the simple arithmetic mean [14] when one wants to characterize a mean flow-rate of information [12].

3. The critical components of a network

Having defined the efficiency E, and assuming it as a good quantity to characterize the average properties of a network, we are now ready to illustrate the method we have recently proposed in Ref. [15] to determine the network

[1] The formalism presented in this paper can be easily extended also to directed graphs.

critical components. Here we will only focus on the determination of the critical nodes, though the method for a generic subset (nodes, links and combination of nodes and links) of **G**. The general theory and all the details can be found in Ref. [15].

The main idea is to use as a measure of the centrality of a node i the drop in the network efficiency caused by the deactivation of the node. The importance $I(\text{node}_i)$ of the ith node of the graph **G** is therefore:

$$I(\text{node}_i) \equiv \Delta E = E(\mathbf{G}) - E(\mathbf{G} - \text{node}_i) \quad i = 1, \ldots, N, \tag{2}$$

whereby $\mathbf{G} - \text{node}_i$ we indicate the network obtained by deactivating node$_i$ in the graph **G**. The most important nodes, i.e. the *critical nodes* are the ones causing the highest ΔE.

Notice that the same procedure can be repeated for a link or for a generic subset $\mathbf{S} \in \mathbf{G}$ consisting of more than one node and/or edge in order to simulate multiple attacks at the same time [15]. Alternatively the method can be used as a guide in the network design, in particular to improve the characteristics of an already existing network [15].

3.1. Example 1: Protection of a communication network

As an illustrative example of how our method works in practice, we present an application to *communication* networks. Communication networks are ubiquitous and are among the most important infrastructure networks. Therefore it is of fundamental importance to identify and protect the critical points of a communication network. Here we focus in particular on the Internet, which is a physical communication network, i.e. a network where each link and node have a physical representation in space. An example of the effect of an attack on the Internet is that caused by the terrorist attacks of September 2001. In fact, immediately following the September 2001 attacks, the Internet experienced a significant drop in performance marked by increased packet loss and difficulty in reaching some Web Sites. We have considered various Internet Backbone maps from Ref. [16]. Here we present the study of Infonet, since the network in question has a small number of nodes $N = 94$ (indicating the cities with Internet access presence and transit), and links $K = 96$, and can be represented graphically in a clear way (see Figs. 1 and 2). The network consists of two main parts: the US and the European backbone respectively with $N_1 = 66$ and $N_2 = 28$ nodes, connected by means of three links. The efficiency of the network is $E = 0.2701$. As said previously we use our method to spot what are the critical nodes of the network, i.e. the nodes whose protection from terroristic attacks must be assumed as the first concern of a national policy.

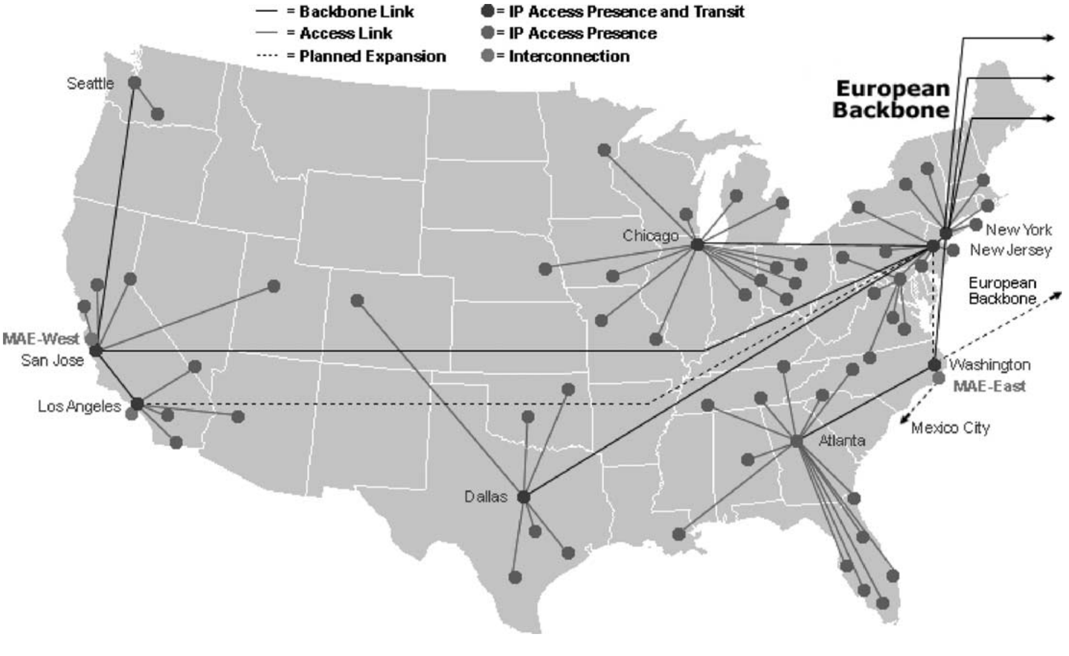

Fig. 1. The US part of the Infonet Internet backbone from [16]. See text for details.

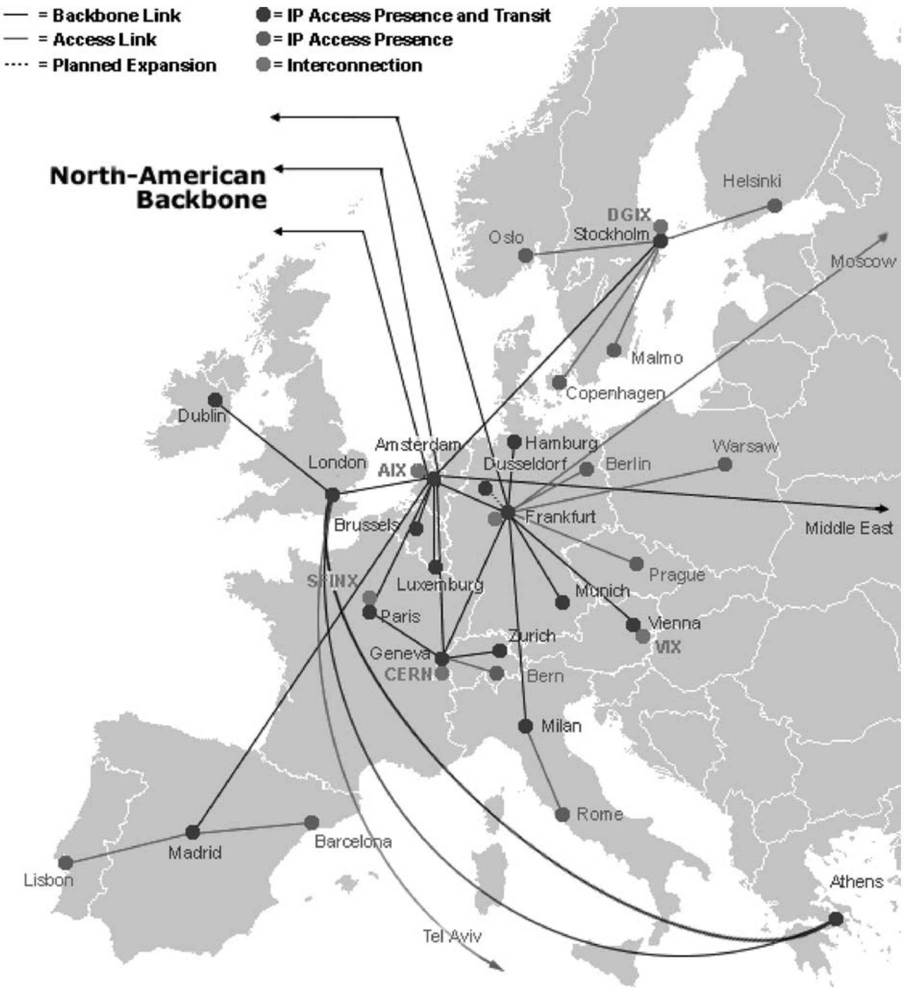

Fig. 2. The European part of the Infonet Internet backbone from [16]. See text for details.

We deactivate one by one any of the $N = 94$ nodes and we calculate the efficiency of the new network and the drop of the efficiency caused, as in formula (2). In Table 1 we report the critical nodes (the six most important nodes), in order of their importance. The numbers reported in table indicate that New Jersey and NYC are two key nodes, in fact their damage would disconnect the US from the European backbone, reducing by more than 50% the efficiency of the

Table 1
Effect of the deactivation of a node in the network of Infonet

	Removed node	$E(\mathbf{G} - \text{node}_i)$	$\Delta E/E$	k
1	New Jersey	0.1155	0.573	9
2	NYC	0.1270	0.530	9
3	Chicago	0.1947	0.280	15
4	Amsterdam	0.2051	0.241	9
5	Atlanta	0.2088	0.227	14
6	Washington	0.2152	0.203	2

The efficiency of the original network is $E(\mathbf{G}) = 0.2701$. The removed node is listed in the first column, the efficiency of the graph once the node is removed is reported in the second column, while the relative drop of efficiency is in the third column. In the last column we report, as an alternative measure of the importance of a node, the degree of (i.e. the number of links incident with) the removed node. The six most important nodes of the network are ordered according to the reduction of the efficiency they cause.

whole network. This result can somehow explain the significant drop in performance experienced in the aftermath of the terroristic attacks of the 11 September, and the difficulty in the connections between Europe and US.

In the same table we also report k the degree of each node, i.e. the number of links incident with the node to show that for a generic network the most connected node is not always the most important [9]. In fact in the case of Infonet, although Chicago and Atlanta are the two most connected nodes, the deactivation of one of these nodes and consequently of its links only reduces of 20% the efficiency of the whole network. This simple result explains that the best strategy is to find and protect the nodes with high $I(\mathbf{S})$, nodes which are often apparently less important than the others, since these are less connected (small k), rather than all the hubs, i.e., the nodes that apparently are the most important because they have the largest number of connections. Deactivating a node means deactivating all its links, and that can have a rather strong consequences on the network. We can also study some finer effects on the network by removing a small number of edges [15]: the results show for instance that NYC–New Jersey is a fundamental link. In fact breaking this single connection will result results in a drop of the 38% of the efficiency of the network. The existence of such a link, and its uniqueness implies either an intense policy of protection of this link from attacks, or a strategic expansion of the network with the addition of new links [15].

3.2. Example 2: Disrupting networks of terrorists

As a second example we consider the connections network of the hijackers and related terrorists of the September 2001 attacks. Of course mapping networks after an event is relatively easy, while the real problem in this case is to map covert networks to prevent criminal activity, a task that can be much more difficult. The network reported in Fig. 3 was reconstructed by Valdis Krebs using public released information taken from the major newspapers [17]. The network has $N = 34$ nodes, representing the 19 hijackers and other 15 associates who were reported to have had direct or indirect

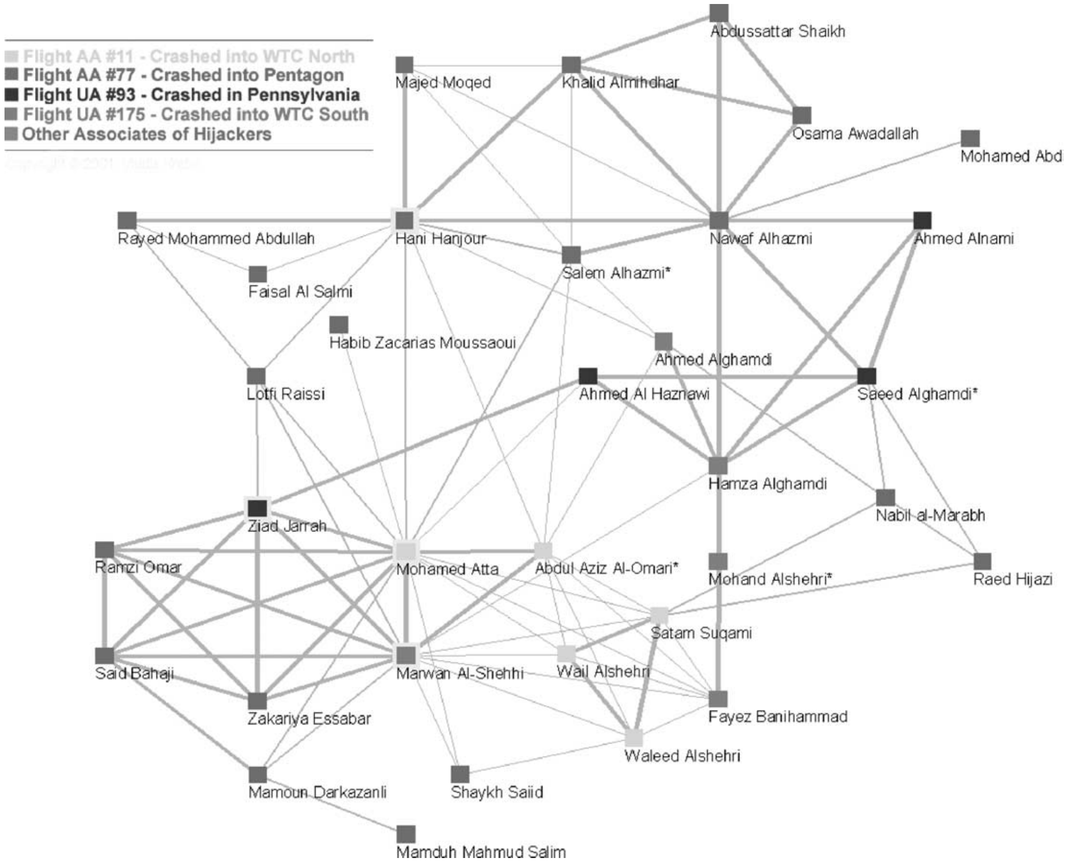

Fig. 3. Connections network of the hijackers and related terrorists of the September 2001 attacks. (Figure taken from [17]). The nodes represent the terrorists and the links represent the interactions between terrorists. See text for details.

Table 2
Effect of the deactivation of a node in the terroristic network of the September 2001 attacks

	Removed node	$E(\mathbf{G} - \text{node}_i)$	$\Delta E/E$	k
1	Mohamed Atta	0.4291	0.150	16
2	Salem Alhazmi	0.4484	0.112	8
3	Hani Hanjour	0.4554	0.098	10
4	Mamoun Darkazanli	0.4586	0.091	4
5	Marwan Al-Shehhi	0.4587	0.091	14
6	Nawaf Alhazmi	0.4611	0.086	9
7	Hamza Alghamdi	0.4646	0.080	7
8	Satam Suqami	0.4656	0.077	8
9	Abdul Aziz Al-Omari	0.4667	0.075	9
10	Fayez Banihammad	0.4710	0.067	7

The efficiency of the original network is $E(\mathbf{G}) = 0.5047$. The removed node is listed in the first column, the efficiency of the graph once the node is removed is reported in the second column, while the relative drop of efficiency is in the third column. In the last column we report, as an alternative measure of the importance of a node, the degree of (i.e. the number of links incident with) the removed node. The ten most important nodes of the network are ordered according to the reduction of the efficiency they cause.

interactions with the hijackers, and $K = 93$ links. In the network map in the figure the hijackers are color coded by the flight they were on. The associates of hijackers are represented as dark gray nodes. The gray lines indicate the reported interactions with thicker lines indicating a stronger tie between two nodes. To individuate the critical nodes i.e. the terrorists who played key roles in the network, we deactivate one by one any of the $N = 34$ nodes and we calculate the efficiency of the new network and the drop of the efficiency caused, as in formula (2). In Table 2 we show the ten most important nodes ranked according to our measure; the degree of each node k is also reported. The most important role in the network was played by Mohamed Atta who was on the flight AA-11 crashed into the World Trade Center North and who was the terrorist having the largest number (16) of direct contacts with the other terrorists. Notice the presence in the table of terrorists having a small degree as Hamza Alghamdi ($k = 7$) who was on the flight UA-175 crashed into the World Trade Center South, and also the important role played by associates of the hijackers as Mamoun Darkazanli. Of course, having a partial knowledge of the terroristic organization networks and individuating the critical components of the network, i.e. the terrorist to target in order to disrupt the terrorist organization, can help a lot to prevent criminal activity.

4. Conclusions

In this paper we have presented a method to identify the critical nodes of a network, i.e. the nodes crucial for the efficient functioning of the network. We have shown two possible applications of the method namely finding the nodes of a communication network to protect from attacks, and finding the key terrorists to target in order to disrupt a terroristic organization network. The method can also be used to identify the critical links of the network and to improve the efficiency of a communication network [15].

Acknowledgement

We thank Andrea Rapisarda for a critical reading of the manuscript.

References

[1] Watts DJ, Strogatz SH. Nature 1998;393:440.
[2] Strogatz SH. Nature 2001;410:268.
[3] Albert R, Barabási A-L. Rev Mod Phys 2002;74:47.
[4] Newman MEJ. SIAM Rev 2003;45:167.
[5] Latora V, Marchiori M. Phys Rev Lett 2001;87:198701.
[6] Latora V, Marchiori M. Europ Phys J B 2003;32:249.

[7] Latora V, Marchiori M. The architecture of complex systems. Santa Fe Institute for Studies of Complexity, Oxford University Press; 2003.

[8] Elliott E, Kiel LD. A complex systems approach for developing public policy toward terrorism: an agent-based approach. This issue.

[9] Albert R, Jeong H, Barabási A-L. Nature 2000;406:378, Correction: Nature 2001;409:542.

[10] Holme P, Kim BJ, Yoon CN, Han SK. Phys Rev E 2002;65:056109.

[11] Crucitti P, Latora V, Marchiori M, Rapisarda A. Physica A 2003;320:622.

[12] Smith J. Commun ACM 1988;31:1202.

[13] Latora V, Marchiori M. Physica A 2002;314:109.

[14] Marchiori M, Latora V. Physica A 2000;285:539.

[15] Latora V, Marchiori M. Vulnerability and protection of critical infrastructures to be submitted. cond-mat.

[16] http://navigators.com/isp.html.

[17] Krebs VE. Connections 2002;24:43, See also http://www.orgnet.com/hijackers.html.