Complex Network Applications to the Infrastructure Systems: the Italian Airport Network case

J. Quartieri*, M. Guida*, C. Guarnaccia*, S. D'Ambrosio**, D. Guadagnuolo*

*Department of Physics "E. R. Caianiello" and Faculty of Engineering,
University of Salerno,

**Department of Mechanical Engineering and Faculty of Engineering ,
University of Salerno

Abstract: - In this paper the structure of the Italian Airport Network (IAN) has been analysed, describing it as a mathematical graph and investigating its topological properties by means of a non standard point of view based on the approach of the multiple addendials.

It turns out that this network has remarkable and interesting features, such as like the one to be a scale-free network, since both the degree and the betweenness centrality distributions follow a typical power-law behaviour, well known in literature like a Double Pareto Law.

Moreover, these distributions exhibit a number of interesting peculiarities, like the existence of some "hubs" in the network, i.e., in the graph theory's jargon, nodes with a very large number of links.

According to a preliminary investigation the IAN can be also considered a candidate to represent a small-world network as the mean distance L between reachable pairs of airports grows at most as the logarithm of the number of airports.

Key-Words: - Airport Network, Graph, Multiple Addendials.

1 Some mathematical background

With the aim to ease reading of this paper also to general people we report some useful considerations on some aspects of the mathematical approach to complex networks. This goal will be accomplished utilizing a non standard point of view, whose basic formulation has been already introduced in past papers by one of us [1].

1.1 Multiple addendials

Referring to [1] for more details, here we address just the main features of an "addendial".

An addendial of first degree is defined as the sum of the first integer numbers:

$$n/_1 \equiv n + (n-1) + ... + 2 + 1 = \frac{n(n+1)}{2} = \binom{n+1}{2} \qquad (1)$$

An addendial of "k"-th degree is, therefore, the sum of the first addendials of "k-1"-st degree:

$$n/_k = n/_{k-1} + (n-1)/_{k-1} + ... + 2/_{k-1} + 1/_{k-1} = \binom{n+k}{k+1} \qquad (2)$$

In a short notation:

$$n/_k = \sum_n n/_{k-1} \qquad (2a)$$

This relation together with the definition (1) suggests to put:

$$n \equiv n/_0 . \qquad (3)$$

Multiple addendials are integer functions of two integer variables and have very remarkable properties [1].

Besides (2a), which we can rewrite as:

$$\sum_n n/_k = n/_{k+1} \qquad (2b)$$

we immediately have:

$$\sum_k n/_k = (n+1)/_k \qquad (4)$$

$$\Delta_n n/_k = n/_{k-1} \qquad (5)$$

$$\Delta_k n/_k = (n-1)/_k \qquad (6)$$

Being $\Delta_j f_j \equiv f_j - f_{j-1}$, $\sum_j$ turns out to be just its inverse operator. In fact $\Delta_j \sum_j = \sum_j \Delta_j = I$ (7)

1.2 Maximum number of links

Let us, now, consider a network consisting of n distinct points (n-vertices or n-nodes) between which we can establish connections (edges or links) in such a way that any two points are connected by links (if each pair of nodes is connected by one and only one link, the network is named "simple"). In this work we deal only with simple networks. The points connected each other in this way will be called "neighbours". Such networks can be mathematically represented by graphs.

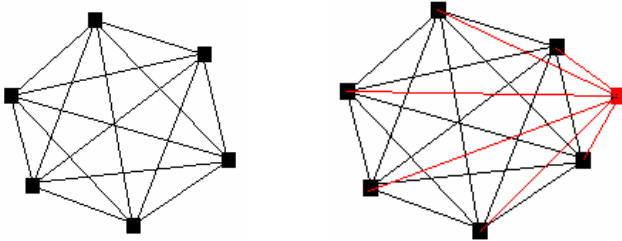Let us calculate how many links can be drawn among n assigned nodes: M(n).

Fig. 1a and Fig. 1b

Then, suppose we already have (n-1) connected nodes (Fig. 1a). Among them we have M(n-1) links. If we add one more node we have to insert (n-1) new links (Fig. 1b). In formula:

$M(n) = M(n-1) + (n-1);$

$M(n) - M(n-1) = (n-1);$

$\Delta M(n) = (n-1)$

Applying the operator Σ to both sides:

$$\Sigma\Delta \ M(n) = \Sigma(n-1)$$

Simplifying the two inverse operators in the l.h.s., and recalling (3) and (2b), we finally have:

$$M(n) = (n-1)/_1 + const$$

In order to determine the value of the constant we impose the initial value condition: $f(2)=1$, which means that between only two nodes we can draw just one link, and we have:

$1 = M(2) = 1/_1 + const = 1 + const \quad \rightarrow \quad const = 0$

So $M(n) = (n-1)/_1$ which also means:

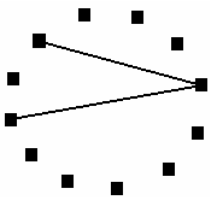$$M(n) = \binom{n-1}{2} = \frac{(n-1)n}{2} = (n-1) + ... + 2 + 1$$

For future reference we put $(n-1)/_1 \equiv X$
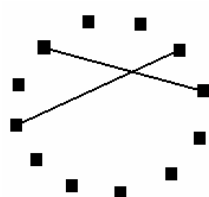
1.3 Maximum number of s-links

Now we want to extend this analysis to a wider class of mathematical objects. Let us begin calculating the maximum number of 2-links $M(n,2)$.

We call "2-link" a sub-graph with only two links (in sequence or not) in an n-nodes graph:

In sequence: Not in sequence:

Before, we saw that: $M(n) \equiv M(n,1) = (n-1)/_1 \equiv X$

Appling the "zero rule" for independent procedures, we can say:

in a set of n-nodes, with no link, we can insert the first link in X different ways. Having placed the first link, we can put the second one in (X-1) ways. So, if the two links were distinguishable we would have $M(n,2)_{dist} = X(X-1)$, but, as we are interested in studying indistinguishable links, we have, theferore, to divide by 2! (permutations):

$$M(n,2) = \frac{X(X-1)}{2!} \ .$$

Iterating, we have:

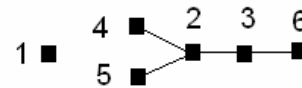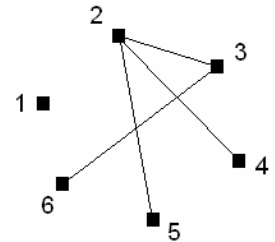$$M(n,s) = \frac{X(X-1)...(X-s+1)}{s!} = \binom{X}{s} \text{ and, finally:}$$

$$M(n,s) = \binom{(n-1)|_1}{s}$$

Let us now introduce few characteristic features useful in the study of complex networks.

Adjacency matrix

A square matrix $A(n \times n)$, whose elements a_{ij} take value 1 *if there is a link from node i to node j* and take value 0 otherwise, is termed the **adjacency matrix**. For the example below it is:

As each link is not directed (undirected network) the matrix A turns out to be naturally symmetric.

Connection degree

The connection degree k_i of a node i is the number of other nodes to which it is connected, or, shortly, the number of its neighbours:

$$k_i = \sum_{j=1}^{n} a_{ij}$$

Distance matrix

The **distance matrix** $D(n \times n)$ has as elements $\{d_{ij}\}$, which are the number of steps, along the shortest path, linking every pair of connected nodes.

The mean distance of a connected network is defined as:

$$L_{conn} = \frac{\frac{1}{2}\sum\limits_{i,j=1}^{n} d_{ij}}{\#_{pairs}} = \frac{\frac{1}{2}\sum\limits_{i,j=1}^{n} d_{ij}}{X_{conn}} \ .$$

The last equality holds because $\#_{pairs}$ equals the maximum number of links, X_{conn}.

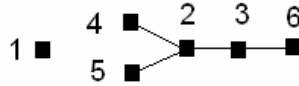When X_{conn} is written in terms of the number of connected nodes, n_{conn}, we have:

$$L_{conn} = \frac{1}{n_{conn}(n_{conn}-1)} \sum\limits_{i,j=1}^{n} d_{ij}$$

and L_{conn} is also named the **characteristic distance**.

For the previous example we obtain:

$$\mathbf{L_{conn}} = \frac{\mathbf{35}}{\mathbf{20}}$$
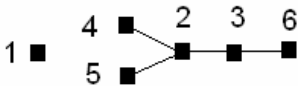
The diameter D of the network can be defined as the length of the maximum "shortest path". $D = max(\ d_{ij}\ )$. In the previous example D=3.

Betweennes or centrality

The betweenness or centrality B_i of node i is defined as the number of all "shortest paths" linking any two different nodes and passing through i.


e.g. $B_2 = 5$, $B_6 = 0$, $B_3 = 3$

We define the normalized betweenness $b_i = B_i / \langle B \rangle$, where $\langle B \rangle$ represents the average betweeness of the network.

Clustering coefficient

the clustering coefficient of node i, C_i, $C_i \in [0,1]$, can be defined as the following ratio, where :

the **numerator** is equal to the number of pairs made of neighbours of i which are themselves neighbours, i.e., neighbours of each other or equivalenty themselves connected by a link.

the **denominator** is equal to the number of all the possible pairs of neighbours that could be in principle constructed from the k_i neighbours of node i, i.e. $(k_i - 1)|_1$:
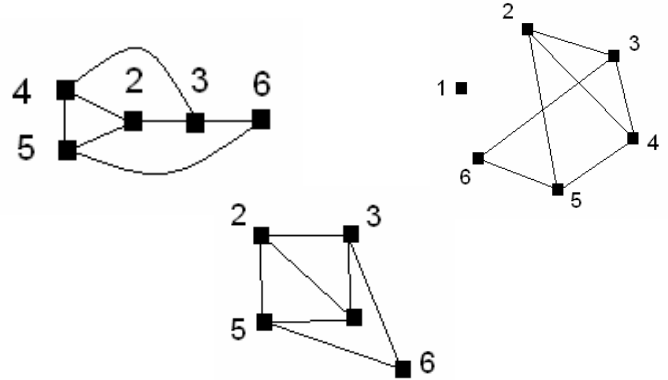
$$k_2 = 3 \quad k_3 = 3 \quad k_4 = 3 \quad k_5 = 3 \quad k_6 = 2$$

$$C_2 = 2/3 \quad C_3 = 1/3 \quad C_4 = 2/3 \quad C_5 = 1/3 \quad C_6 = 0$$

2 The Italian Airport Network

We investigated the topological properties of the Italian Airport Network (IAN), representing it like a mathematical graph where to each airport is associated a node and pairs of nodes that are just connected by non-stop passenger flights are linked together. In order to accomplish that, we have studied the data derived from the OAG Max database [2], compiled by OAG Worldwide (Downers Grove, IL) including all the scheduled flights and scheduled charter flights of the world's airlines both for big aircrafts (air carriers) and small aircrafts (air taxis) for the period June 1, 2005, to May 31, 2006 (period (o)). At the end of this construction, whose full details are contained in a previous paper by one of us [3], the data corresponding to IAN have been extracted and the resulting directed graph has turned out to be made of 42 nodes and a total number of links (non-stop flights) equal to 310 (Fig. 2).

Following the definitions given in sect. 1, let us consider the adjacency matrix, $A(n \times n)$, whose elements a_{ij}, in the case of the IAN, take value 1 *if on any day of the week* there is a flight *from node i to node j* by any service provider and 0 otherwise. In order to simplify the following analysis we will restrict ourselves just to data pertaining to flights available in the period (o). Similar investigations related to other periods of time are described in details in [3].
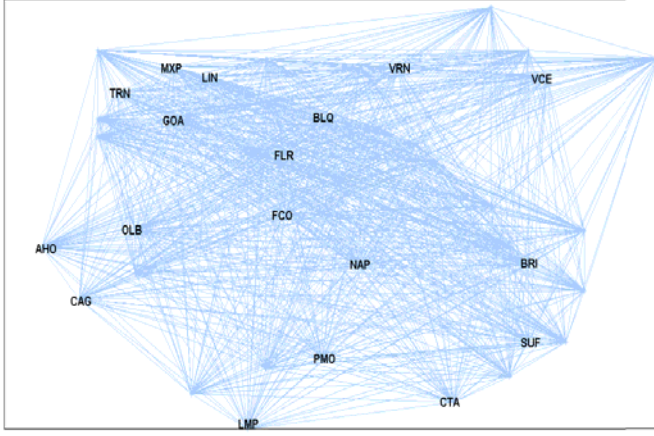
Fig.2: The Italian Airport Network (IAN)

Then, we calculate for each node i the connection degree k_i, that is the number of other airports to which airport i is connected by a non-stop flight, and the corresponding cumulative distribution cumulative distribution $P(>k)$, which gives the probability that a node has k or more connections to other nodes and it is defined as:

$$P(>k) = \sum_{k'=k}^{\infty} p(k'), \text{ where } p(k) \text{ is the probability density,}$$

$$p(k') = k' / \sum_{1}^{n} k_i$$

On a log-log scale the normalized cumulative distribution $P(>k)$ versus the degree k looks like (Fig. 3).
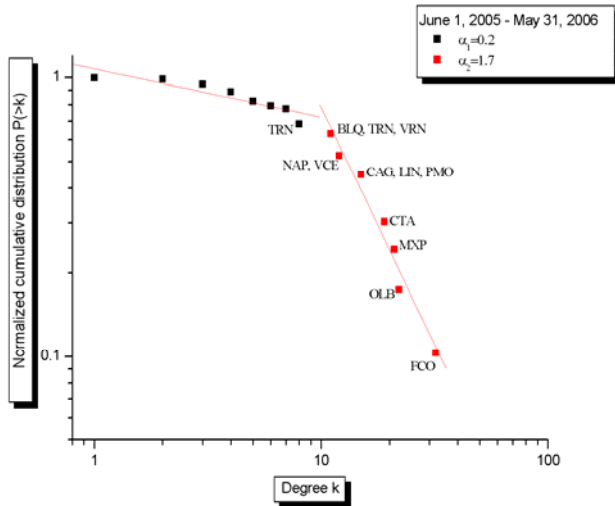
Fig.3: The normalized cumulative distribution $P(>k)$ versus the degree k for period (o).

The data distribution suggest that $P(>k)$ follows a Double Pareto Law [4]

$$P(>k) \propto \begin{cases} k^{-\alpha_1} & \text{for } k \leq k_c \\ k^{-\alpha_2} & \text{for } k > k_c \end{cases},$$

where $k_c = 9$ and $(\alpha_1, \alpha_2) = (0.2, 1.7)$.

This behaviour is typical of many complex networks, termed "scale-free" [5].

The fact that the exponent α_2 of the degree distribution is greater than one suggest the possibility that the structure of our network is fractal as proposed in [6]. The fact that networks are scale-free was shown to have important implications on network robustness to random failures and vulnerability to attack [7]. The distribution reportedin Fig. 3 clearly shows the existence in the Italian Airport Network of so-called "hubs", i.e. nodes with a large number of connections: in our case Fiumicino (FCO), Olbia (OLB), Milano-Malpensa (MXP) and Catania (CTA).

The degree of a node is a importance source on its own but, however, does not provide complete information on the role played by the single nodes in the network. To detail a bit more the role of each node inside the network let us adapt, according to the definitions already given in sect. 1, the so-called "betweenness centrality" of cities, introduced in [8-9].

Besides the betweenness B_i of each airport i it is convenient, as done in [3, 10], to introduce the corresponding normalized betweenness $b_i = B_i / \langle B \rangle$, where $\langle B \rangle$ represents the average betweeness of the network and compute the cumulative distribution $P(>b)$ as

$$P(>b) = \sum_{b'=b}^{\infty} p(b'),$$

where $p(b)$ is the probability density. For the period considered, we plot, on a log-log scale, the normalized cumulative distribution $P(>b)$ versus the normalized betweenness b (Fig. 4). As for the degree distribution (Fig. 3) the normalized cumulative distribution $P(>b)$ follows a Double Pareto Law. This behaviour turns out to be again typical of a scale-free network [4].
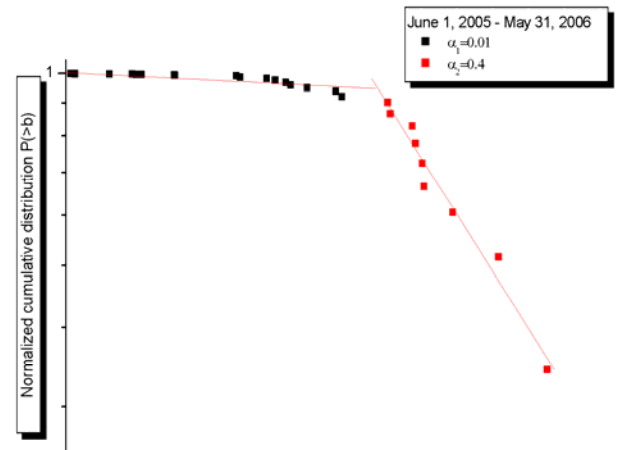
Fig.4: The normalized cumulative distribution $P(>b)$ versus the normalized betweenness b for period (o).

Another property of the IAN network is that it is a small-world network.

According to Watts and Strogatz' [11-12] a network is a small-world network when the mean distance among reachable pairs of nodes L grows at most as $\log_e n$, where n is the number of nodes.

In order to check that, we calculate L, as it has been defined previously.

For the IAN, it turns out to be that

$$L = 1.97 \quad \text{and} \quad \log_e n = 3.74 \text{ for period (o).}$$

These figures indicate that L is closer to 3.74 rather than to the number of nodes, n = 42 and show that according to the definition [11-12], the Italian Airport Network is a "small-world" network.

Another quantity that is worth to check in order to characterize the topological properties of a network, is the so-called network's clustering coefficient C, defined previously.

If we calculate C, according to the definition, for the IAN we find that, for the period (o) it turns out to be: $C = 0.10$.

4 Conclusion

In this paper the topological properties of the system of the Italian airports have been investigated from a non standard approach to the theory of complex networks. The associated network, the so-called Italian Airport Network (IAN) have been constructed associating a node to each airport and an edge to each non-stop connecting passenger flight operating between two different airports in any of the seven days of the week.

The topological properties of the resulting network have been carefully examined leading to the evidence of a scale-free behaviour in the degree of connection distribution, in which the existence of "hubs", i.e. nodes with a high value of the degree of connection, clearly manifests. However, the scale-free behaviour turned out to be a little bit different from the ones already reported in the literature for other air transportation networks, suggesting that the formation mechanism model underlying the IAN could be pretty different from the ones proposed so far.

The topological properties of IAN have been investigated and confirmed (e.g., the scale-free characteristics) considering the data available in different period of time related to different seasons of the year. In these cases the well-known tourist vocation of some Italian locations really makes the difference. From such a kind of analysis some interesting and precious results for professional operators of the National Airport Infrastructure can be easily extracted.

Furthermore, IAN satisfies all the requirements prescribed by the definition of small-world networks.

References:

[1] Quartieri J., "Some mathematical considerations arising in the study of a Degenerate Harmonic Oscillator", *Lett. Nuovo Cimento*, Vol. 35, N. 14, 1982, p.p. 433-436.

[2] http://oagdata.com/solutions/max.aspx.

[3] Guida M. and Funaro M., "Topology of the Italian Airport Network", *Chaos, Solitons & Fractals*, Vol. 31, 2007, p.p. 527-536.

[4] Li W., Cai X., *Phys. Rev. E*, Vol. 69, 2004, 046106.

[5] Bàrabasi A. L. and Albert R., *Science*, Vol. 286, 1999, p.p. 509-512.

[6] Song C., Havlin S. and Makse H.A. (2005), "Fractal growth of complex networks : repulsion between hubs". ArXiv: cond-mat/0507216 v1.

[7] Crucitti P., Latora V. and Marchiori M., "Model for cascading failures in complex networks", *Phys. Rev. E*, Vol. 69, 2004, 045104(R).

[8] Freeman, L. C. (1977), *Sociometry*, Vol. 40, 35–41.

[9] Newman, M. E. J. (2001), *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.,* Vol. 64, 016131, 016132.

[10] Guimerà R., et al. (2005), "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles", *Proc. of the Natl. Acad. of Sci. USA,* Vol. 102, No. 22, 7794-7799.

[11] Watts D.J., Strogatz S.H. (1998) "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393, 440-442.

[12] Watts D.J. (1999), "Small Worlds - The Dynamics of Networks Between Order and Randomness". Princeton University Press.