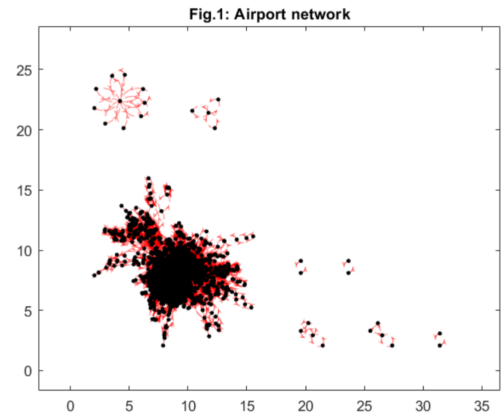
Airport network

Lincetto Riccardo 1156313

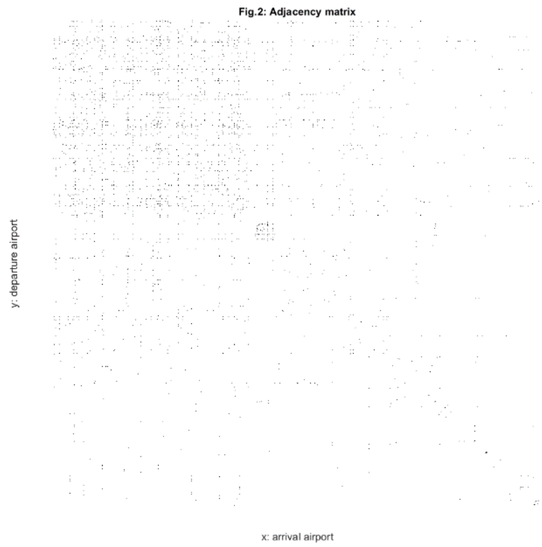
Network description

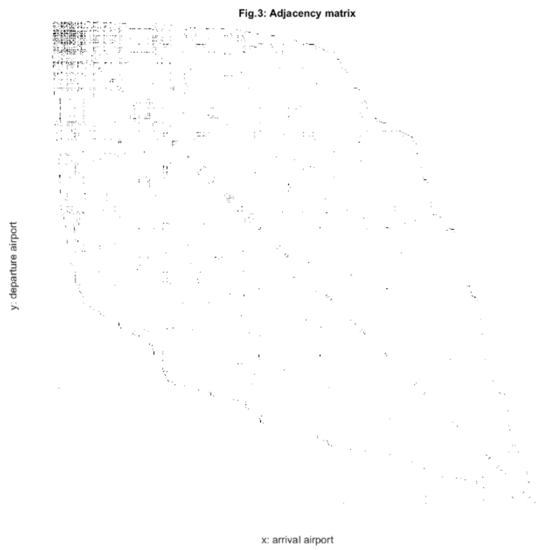
In this network nodes represent different airports and edges are placed between nodes if there exist a flight between corresponding airports. The dataset has been downloaded from kaggle.com as a CSV file format which contains the description of a set of 67663 flights in a global scale, including departure and destination airports, the airline company IDs and other information which are not needed for the creation of the network. The resulting graph is then directed, composed by 3425 nodes and 37595 edges which are not weighted: the number of flights associated to the network is less than the starting one because a great fraction of them can be operated by multiple airline companies, but in this case only the existence of a flight is needed to place a connection between two airports. The graph obtained is showed in Fig. 1 (check *graph.bmp* in the directory for a better resolution).



The first thing that it is possible to notice from the plot is that there exist a giant component and that there are other connected components, which are not connected to the former.

Even though the network is directed, the resulting graph is almost as if it were undirected: this can be better observed by looking at the adjacency matrix, which is not far from being symmetrical. Two graphical representations of the matrix are thus displayed in Fig. 2 and Fig. 3, with black points indicating the existence of a link between airports and white points indicating the opposite. The difference between the two representations consists of the order of nodes: the first figure has its nodes in alphabetical order while the second one has the nodes belonging to the giant component in front and then the other components. The second ordering was obtained running a breadth first search (BFS) algorithm and the structure of the matrix obtained should be the one of a block diagonal: this is hard to see anyway on the bottom right part of the Fig. 2 but still the quasi-symmetry can be appreciated from it (check *adj_matrix.bmp* and *adj_matrix_reordered.bmp* in the directory).

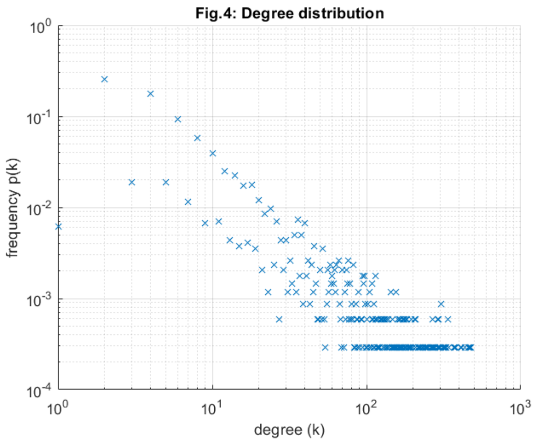




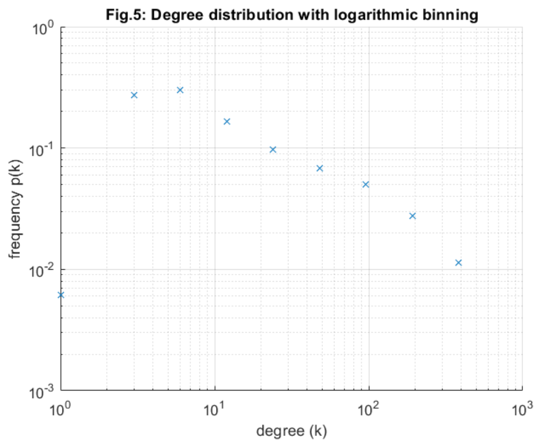
One more thing that can be observed from these two matrix representations is that the network is sparse, in fact the maximum number of edges which can be present (in case of a fully connected network) is in the order of 10^6 while the actual number of edges is in the order of 10^4.

Degree distribution

Defining the degree of a node as the sum of its in-degree and its out-degree, the histogram of the degree distribution is obtained as showed in Fig. 4



Data is a bit noisy because the number of nodes is not high enough to have a smooth representation of the curve, but still it is possible to appreciate the absence of isolated nodes and the presence of a few hubs in the network. The biggest hub in the network is Frankfurt Airport (FRA), which is in fact the biggest airport of Germany and the 13th busiest one in 2016 for number of passengers according to wikipedia, that in this network has degree 477. The average degree is instead 21.95 and its variance is 2336.5. Using a logarithmic binning to represent data it is possible to observe a linear behaviour in the logarithmic domain and thus the degree distribution follows a power law: the network has then the scale-free property and using the approximated formula gamma is found to be $\gamma = 1 + N / \left(\sum_{j=1}^{N} \ln k_j / k_{min} \right)$ = 1.4, where N is the number of nodes, k_j is the degree of node j and k_{min} is the minimum degree (which is equal to 1 in this network). Results of logarithmic binning can be observed in Fig. 5.



Connectivity

The network presents a giant component of 3378 nodes over 3425 and a total of 44 components, where each one is meant as a set of nodes strongly connected (i.e. linked in both directions). The diameter of the graph coincides with the one of the giant component and has value 14, while the average value is 4.1, which is a bit higher than expected since $\hat{d}_{avg} = \ln(N) / \ln(k_{avg}) = 2.6$. The average clustering coefficient is 0.6 and it is computed excluding the values for the nodes that have only one neighbor. Since the value of the clustering coefficient is quite high, the network is robust.