Eur. Phys. J. B (2016)
DOI: 10.1140/epjb/e2016-60887-x

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article

Weighted multiplex network of air transportation

Imre Varga[a]

Department of Informatics Systems and Networks, University of Debrecen, 4028 Debrecen, Hungary

Received 14 November 2015 / Received in final form 26 February 2016

Abstract. In several real networks large heterogeneity of links is present either in intensity or in the nature of relationships. Therefore, recent studies in network science indicate that more detailed topological information are available if weighted or multi-layer aspect is applied. In the age of globalization air transportation is a representative example of huge complex infrastructure systems, which has been analyzed form different points of view. In this paper a novel approach is applied to study the airport network as a weighted multiplex taking into account the fact that the rules and fashion of domestic and international flights differ. Restricting study to only topological features and their correlations in the system (disregarding traffic) one can see reasons why simple network approximation is not adequate.

1 Introduction

Through the researches of this century a new field of science got into the focus of attention. Network science tries to describe and understand how the units of a large complex system interact or connect to each other. Theoretical models [1–5] are developed to characterize the structural properties and dynamics of broad range of real world networks [6–11] having emergent behavior. Description of networks become more detailed by introducing weights of interactions [12–14] or the multilayer network aspect [15–21] or both [22].

Transportation systems, telecommunication networks, electrical grids or other interdependent critical infrastructures have remarkable effect to economy and our everyday life. These networks play a part in several dynamical processes such as spreading of diseases or information, cascading failures and so forth [23–26]. In order to predict and understand these processes first we have to analyze the structure of the underlying networks.

The aim of this paper is to characterize the structural properties and the correlations of air transportation network from special point of views. One of the most basic classification of flights based on the country of source and destination airports. The key is not the distance, but the conditions of domestic flights and flying abroad can be very different (duty, passport control, visa, language, etc.). But what about the network structure? Are there differences between the structures of domestic and international air transport networks? Are there correlations between them? To answer these questions a multiplex description is applied. Additionally not just the existence of a connection between two given airports can be important, but some kind of intensity of their relationship also (e.g. on a popular link more airlines operate).

In Section 2 the weighted multiplex description of global air route network is presented and the terminology is introduced. In Section 3 the topological analysis and its results are shown in regard to correlations. The paper is closed with conclusions in Section 4.

2 A weighted multiplex approach of air transportation

In this work the data source of the world-wide air transportation system was used provided by OpenFlight [27]. The dataset contains the source and destination airports (and their countries) of non-stop direct flight routes of airlines. Almost 3200 airports are connected by more than 66 500 directed routes of 540 airlines in 226 countries of the world.

In order to proceed the general analysis of the system it is considered as a graph, where the vertices are separate airports (not cities with one or more airports). Almost all routes between airport pairs in the dataset are symmetrical, i.e. if there is a direct flight from A to B then there exists a flight from B to A as well. This work is restricted to only symmetrical cases representing the connections of nodes by undirected links. While numerous airlines (with several flights) can operate between two given airports links can be considered to be weighted. In the aspect of this topological study the strength of connection can be captured better by the number of different airlines, than the number of flights from the source to the destination in a given time interval. Thus the $w_{ij} \in \mathbb{N}$ weight of a link between nodes i and j is measured by the number of airlines operate between them. This means if

---

[a] e-mail: varga.imre@inf.unideb.hu

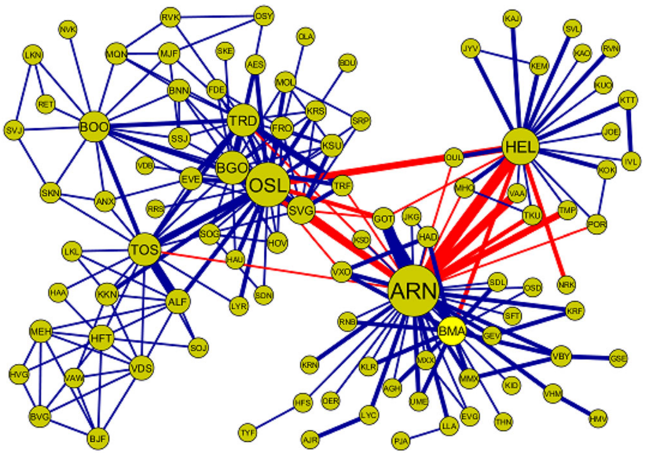

Fig. 1. The undirected weighted multiplex network of the Norwegian, Swedish and Finnish airports. Nodes are labeled by the IATA code of airports. The size of nodes is proportional to their degree. Nodes are connected by two different links representing domestic (dark blue) and international (light red) routes. The links are undirected due to the symmetric flights. The width of the links is determined by the weight of direct routes i.e. the number of different airlines operating direct non-stop flights between the two airports.

$w_{ij} > 1$, one has chance to choose among more airlines to travel from airport i to j. By the way the traveling terms can also change due to the race condition. The presence of more airlines between two airports indicates more prestige of the link measured by weight. Cardillo et al. [18] treated airlines as layers of the network so from this point of view this w_{ij} weight of a link gives the number of airline-layers where node i and j is connected in their model. In order to take into account the differences between international and domestic routes two types of links are introduced. Based on this the airport network is a multiplex of two network layers. One of them contains routes between airports of the same country while the other refers to international routes. A small part of the system is illustrated in Figure 1.

2.1 General formalism

The system consists N_N nodes representing airports. Both the *International* and the *Domestic* layer of the multiplex can be specified by an adjacency matrix

$$A^{[\alpha]} = \left\{a_{ij}^{[\alpha]}\right\}, \tag{1}$$

where $\alpha \in \{Int., Dom.\}$ and $a_{ij}^{[\alpha]} = 1$ if nodes i and j are connected and $a_{ij}^{[\alpha]} = 0$ otherwise for all $i, j = 1, \ldots, N_N$. The vector of these adjacency matrices $\mathbf{A} = \{A^{[Int.]}, A^{[Dom.]}\}$ describes the total multiplex network. The degree of the multiplex nodes is also a vector

$$\mathbf{k}_i = \left\{k_i^{[Int.]}, k_i^{[Dom.]}\right\}, \tag{2}$$

where $k_i^{[\alpha]}$ is the number of neighbors of node i in layer α. The total number of links in layer α is $N_L^{[\alpha]} = 1/2 \sum_i k_i^{[\alpha]}$. Important to mention that $k_i^{[\alpha]}$ is interpreted in a weighted network that is more available airlines between a pair of airports does not change the degree, however changes the opportunities of travelers.

The aggregated topological adjacency matrix can be defined as $A = \{a_{ij}\}$, where $a_{ij} = \max(a_{ij}^{[\alpha]})$. This aggregated network contains $N_L = \sum_\alpha N_L^{[\alpha]}$ links.

3 Comparing international and domestic layers

The full multiplex of airports has $N_N = 3182$ nodes and $N_L = 18\,797$ weighted links ($N_L^{[Int]} = 9956$ in international layer and $N_L^{[Dom]} = 8841$ in domestic layer). The average degree of nodes is $\langle k \rangle = 11.814$, while the average weight of links is $\langle w_{ij} \rangle = 1.821$ in the aggregated network.

First results of the analysis say that almost 2/3 of nodes have connections only in their country and 192 airports are available only from abroad, while 987 nodes are connected to others by both types of links. The aggregated network contains 8 separate clusters of airports with a giant component, which covers 99.1% of nodes. From the point of view of clusters the two layers of the system are completely different. Layers contain clusters of airports, where there are no connections between clusters. Nodes which have connections in the given layer form $n_c^{[Int.]} = 8$ and $n_c^{[Dom.]} = 147$ separate (smaller or larger) clusters. Of course, all countries define an own cluster in the *Domestic* layer, while *International* layer contains a dominant giant component. Naturally all pairs of airports are either in the same country or in different countries so a special property of the system is that there are no multi-links or overlapping edges.

While most of airports have only a few connections, some huge airports have hundreds of links. This network is scale-free as found earlier [8,12] for air transportation networks. The degree distribution of each layer obeys power-law form with an exponential cut-off

$$P(k) \propto k^{-\gamma} \exp(-k/k_x), \tag{3}$$

where k_x is a cut-off distance (see the inset of Fig. 2). The exponent γ and the cut-off distance k_x are definitely different in the two layers.

Beside the degree distribution, the distribution of link weights is also an important property of the system. From the point of view of this quantity the layers do not differ. The weight distribution is exponential with the same value of the coefficient in the exponent (see Fig. 2). As the figure shows there are direct routes between airports where more than 10 airlines are present in both layers. Nodes can be characterized by their strength as well, written as:

$$s_i = \sum_{j=1}^{N_N} a_{ij} w_{ij}, \tag{4}$$

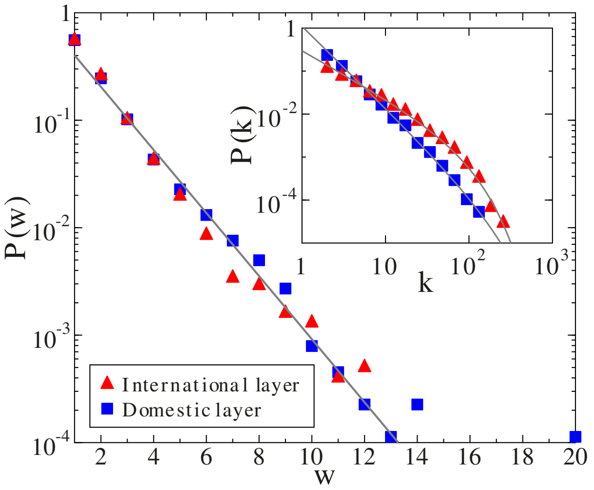

Fig. 2. Exponential distribution of linkweights in a semi-log plot. The solid line represents the fit of the distribution for the aggregated network with $P(w) = 0.79e^{-0.68w}$ ($R^2 = 0.98$). Inset: degree distribution proves scale-free behavior of layers. The solid lines indicate fitting by equation (3), where $\gamma^{[Dom.]} = 1.9$ and $\gamma^{[Dom.]} = 1.1$. Logarithmic binning is used in the plot.

that is as the sum of the weights of links of node i. Strength in this way is a kind of weighted degree. In order to study the correlation among the weights and degree the strength-degree correlation is plotted in the inset of Figure 3. As it is visible the strength of node i depends on its degree naturally. This dataset can be well fitted by the $s_i = \langle w_{ij} \rangle \, k_i$ form. This means that there is a correlation between the strength and the degree of node i, but there is no correlation between the weights of links of node i and its degree.

The degree correlation can be qualified by the local weighted average nearest neighbors degree [12], defined as:

$$k^{w}_{nn,i} = \frac{1}{s_i} \sum_{j=1}^{N_N} a_{ij} w_{ij} k_j. \tag{5}$$

The average of this quantity over all nodes with degree k as a function of degree k is represented in the main panel of Figure 3. As one can see the two layers act in radically different ways. In the *Domestic* airroute network layer nodes tend to connect to other nodes with similar degree, so increasing function indicates weighted assortativity. In the same time in *International* layer neither this correlation nor negative correlation can be observed.

To combine the topological and weight information $c^{w[\alpha]}_i$ weighted clustering coefficient of node i was introduced [12] in layer α in the following form

$$c^{w[\alpha]}_i = \frac{1}{s^{[\alpha]}_i (k^{[\alpha]}_i - 1)} \sum_{m,n} \left(a^{[\alpha]}_{im} a^{[\alpha]}_{mn} a^{[\alpha]}_{ni} \frac{w^{[\alpha]}_{im} + w^{[\alpha]}_{in}}{2} \right), \tag{6}$$

where node i has more then one connections in the given layer ($k^{[\alpha]}_i > 1$). Naturally in unweighted case where $w_{ij} = 1$ we get back the traditional topological clustering coefficient of a layer $c^{w[\alpha]}_i = c^{[\alpha]}_i$. One of the metrics of layer α is the average weighted clustering coefficient

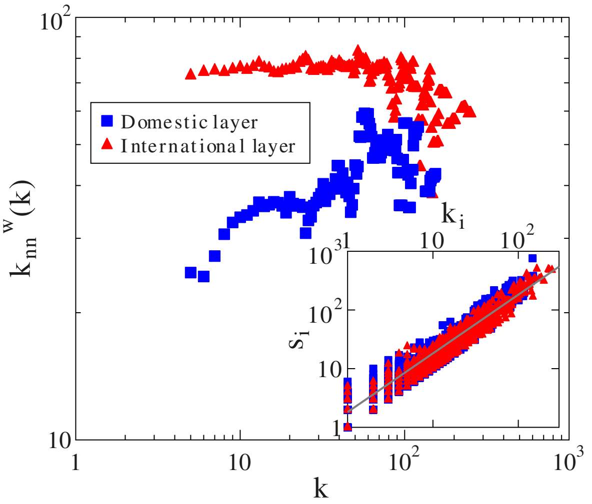

Fig. 3. Weighted average nearest neighbor degree as a function degree shows assortativity in *Domestic* layer, while the average neighbors degree does not depend on the degree of the given node in *International* layer if an airport has less than 100 connections (in order to avoid statistical fluctuations simple moving average is plotted). Inset: strength-degree correlation of nodes. Solid line illustrates a linear dependence, where the slope is the value of $\langle w_{ij} \rangle$. Thus, just the average value of weights has influence to node strength independently of actual weights of the links of a given node.

$$C^{w[\alpha]} = \frac{1}{N_N} \sum_i c^{w[\alpha]}_i. \tag{7}$$

As it is known the average topological clustering coefficient $C^{[\alpha]}$ is smaller than $C^{w[\alpha]}$, if links with large weights tend to form triplets, while in uncorrelated (randomized) network $C^{[\alpha]} = C^{w[\alpha]}$. In this air transportation multiplex $C^{w[Int.]} = 0.356 \pm 0.323$ and $C^{w[Dom.]} = 0.475 \pm 0.423$, so they are clustered networks. Both values are a bit above the unweighted $C^{[\alpha]}$ value, but the differences are smaller than the margin of errors. In this way the correlation between topology and weights cannot be significant.

Due to economic reasons most travelers choose routes between two given airport minimizing the number of transfer at internal airports. This is why the L_{ij} shortest path length between node i and j is an important quantity in this network. The diameter of a network can be defined as $D = \max(L_{ij})$. In the *Domestic* and *International* multiplex layer the diameter is $D = 10$ and $D = 8$, respectively, while the average shortest path length over nodes $\langle L \rangle$ is a bit above 3.0 in both cases. The cumulative distribution of the shortest path length of these small-word networks is shown in Figure 4.

To measure the importance of airport m normalized betweenness centrality $c_B(m)$ can be introduced, which shows how many percentage of the shortest paths of cluster from node i to j pass through node m ($i, j = 1, \ldots, N_N$ and $i, j \neq m$). The average value of c_B differs within the two layers of the multiplex, $\langle c^{[Dom.]}_B \rangle = 0.0409 \pm 0.15$ and

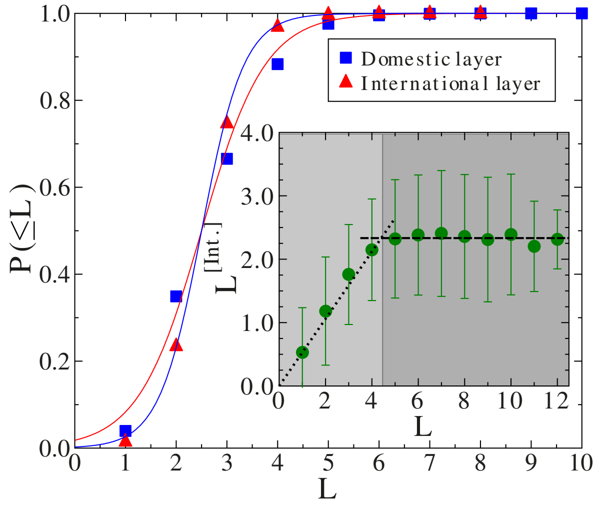

Fig. 4. Cumulative shortest path length distribution. Only a few percentage of shortest paths are longer than the half of the diameter of the network. Solid curves just illustrate logistic functional form. Inset: average number of *International* layer link in shortest paths of the aggregated network as a function of the total path length. A linear and a constant regimes exist. The slope of the dotted line fitting the former regime is the average ratio of international links in all shortest paths ($N_L^{[Int.]}/N_L = 0.530$). On the average long air routes contain less than 2.5 direct international connections, indicated by dashed line.

$\langle c_B^{[Int.]}\rangle = 0.0034 \pm 0.04$. In *Domestic* layer c_B is one order of magnitude larger than in the *International* layer because the former contains many small clusters. Within small clusters there are less shortest paths and more nodes play local central role. In order to explore the relationship among node degree and the normalized betweenness centrality their correlation coefficient R^2 is determined. Its value is $R^2 = 0.0124$ and $R^2 = 0.0267$ in *Domestic* and *International* layers, respectively. Correlation is not found, so not only more-connected airports can more-central and vice versa as it is shown by Guimerà et al. [8,9].

In the aggregated network the length of a general shortest path can be written as $L = L^{[Int.]} + L^{[Dom.]}$, where $L^{[\alpha]}$ is the number of flights in layer α along this shortest path of the aggregated network. The ratio of international and domestic hops depends on the length of path. The average number of direct international links along a general path with length L as a function of the total path length has two separate regimes (see Fig. 4 inset). If the path length L is larger than a crossover path length $L_x \approx 4$ only the number of domestic flights is increasing. Statistically one can reach all destination from every airport by not more than 2.5 international flights. Long routes contain several domestic transfers.

3.1 Correlation between layers

Assortativity/dissortativity is an important feature of a simplex network or a layer of multiplex [28]. In order to characterize correlation between layers of multiplex interlayer degree correlation was introduced as

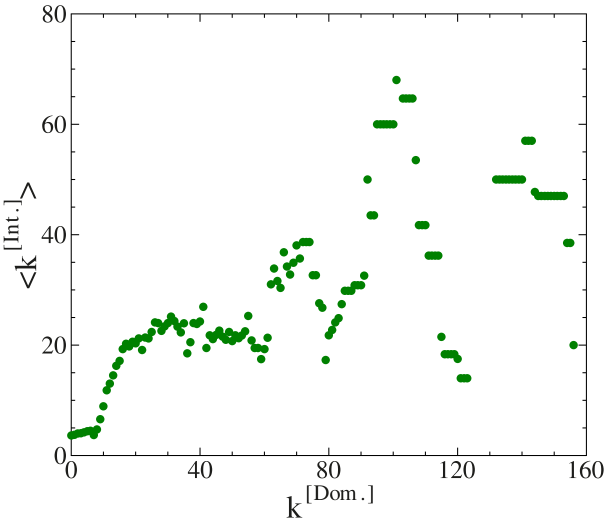

Fig. 5. The average degree in *International* layer as a function of degree in *Domestic* layer. To reduce the large fluctuations moving average is represented.

$$\rho = \frac{\left\langle k^{[Int.]}k^{[Dom.]}\right\rangle - \left\langle k^{[Int.]}\right\rangle\left\langle k^{[Dom.]}\right\rangle}{\sigma_{k^{[Int.]}}\sigma_{k^{[Dom.]}}}, \tag{8}$$

where $\sigma_{k^{[\alpha]}}$ is the standard deviation of degrees in layer α [15,29]. The value of this Pearson correlation coefficient between the two layers is $r = 0.271$, which indicates weak positive correlation [21]. This means that hubs of *International* layer are probably hubs of *Domestic* layer, as well.

The average degree in *International* layer $\langle k^{[Int.]}\rangle$ as a function of degree in *Domestic* layer $k^{[Dom.]}$ shows also low positive correlation by its increasing trend (see Fig. 5). Ranking the airports by degree in both layers the weak correlation becomes self-evident. Only 6 airports are both in the top 50 most connected airports of the two separate layers.

4 Conclusions

A study of world-wide air transportation network is presented, which points out the differences of the route networks of *Domestic* and *International* flights by considering the global system as a weighted multiplex with two layers. The effects of weights and the relationship of the weights and the topology is highlighted in order to realize the differences of the layers. It was found that the simplex airport network hides many details of this complex system. Layers are relevant entities of the network, because they are different from each other and different from the aggregated network as well. On of the most important results is that only the multiplex approach can tell us that statistically only 2 or 3 passport controls are necessary during a long travel containing more than 10 direct fights. In this way taking layers (in macroscopic scale) and weights (in micro scale) into account leads to a better description of emergent systems.

References

1. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
2. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
3. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
4. M.E.J. Newman, *Networks: an Introduction* (Oxford University Press, Oxford, 2010)
5. I. Varga, in *Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014* (Springer, 2015), p. 323
6. L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Proc. Natl. Acad. Sci. **97**, 11149 (2000)
7. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW* (Oxford University Press, Oxford, 2003)
8. R. Guimerà, S. Mossa, A. Turtschi, L.A.N. Amaral, Proc. Natl. Acad. Sci. **102**, 7794 (2005)
9. R. Guimerà, L.A.N. Amaral, Eur. Phys. J. B **38**, 381 (2004)
10. F. Liljeros, C.R. Edling, L.A. N Amaral, H.E. Stanley, Y. Aberg, Nature **411**, 907 (2002)
11. E. Bullmore, O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009)
12. A. Barrat, M. Barthélémy, R. Pastor-Satorras, A. Vespignani, Proc. Natl. Acad. Sci. **101**, 3747 (2004)
13. A. Barrat, M. Barthélemy, A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004)
14. G. Bagler, Physica A **387**, 2972 (2008)
15. K.-M. Lee, B. Min, K.-I. Goh, Eur. Phys. J. B **88**, 48 (2015)
16. V. Nicosia, V. Latora, Phys. Rev. E **92**, 032805 (2015)
17. V. Nicosia, G. Bianconi, V. Latora, M. Barthelemy, Phys. Rev. Lett. **111**, 058701 (2013)
18. A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, S. Boccaletti, Sci. Rep. **3**, 1344 (2013)
19. F. Battiston, V. Nicosia, V. Latora, Phys. Rev. E **89**, 032804 (2014)
20. K.-M. Lee, J.Y. Kim, W.-K. Cho, K.-I. Goh, I.-M. Kim, New J. Phys. **14**, 033027 (2012)
21. B. Min, S.D. Yi, K.-M. Lee, K.-I. Goh, Phys. Rev. E **89**, 042811 (2014)
22. G. Menichetti, D. Remondini, P. Panzarasa, R.J. Mondragn, G. Bianconi, Plos One **9**, 6 e97857 (2014)
23. G. Kocsis, I. Varga, Infocomm. J. **6**, 45 (2014)
24. S.V. Buldyrev, R. Parshani, G. Paul, H.E. Stanley, S. Havlin, Nature **464**, 1025 (2010)
25. G. Kocsis, F. Kun, J. Stat. Mech. Theor. Exp. **10**, P10014 (2008)
26. J. Ø. H. Bakke, A. Hansen, J. Kertész, Europhys. Lett. **76**, 717 (2006)
27. Source of airport, airline and route datasets: `http://openflights.org/data.html`, accessed in August 2015
28. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
29. J.Y. Kim, K.-I. Goh, Phys. Rev. Lett. **111**, 058702 (2013)