Spatial Aggregation of Worldwide Air Transportation Networks

Xiaoqian Sun[12] and Sebastian Wandelt[12,*]

Abstract

Air transportation, as the dominant mode for long-distance passenger travels, plays a critical role for a sustainable and greener future. Therefore, the analysis and understanding of air transportation systems is crucial. We investigate air transportation networks and their layers as induced by spatial aggregation. Existing studies analyze network layers individually, ignoring interactions/couplings between the layers. Here, we aim to provide a unified understanding of the transitions among network layers, based on the worldwide air transportation networks for the year 2015. We derive aggregated network instances at six different levels. While few nodes are important at all levels of aggregation, others only become important for few aggregation levels. Moreover, we report our analysis results of different layers' robustness.

Keywords: Air transportation networks; Aggregation hierarchy; Network robustness

1 Introduction

With steadily increasing long-distance travel demand, air transportation networks become an essential part of transportation infrastructures in modern societies [Zanin and Lillo2013, Cook et al. 2015]. In order to better understand their structures and dynamics, a significant amount of research has been conducted, for instance, on topological properties of air navigation route networks [Sun and Wandelt2014, Sun et al. 2017b], evolution of airport networks [Azzam et al. 2013, Sun et al. 2015], network resilience against airport closures or flight cancellations [Wei et al. 2014, Wandelt et al. 2015, Sun et al. 2017a], as well as multiplex properties of airline networks [Zanin2015]. It is nevertheless by and large recognized that modeling the network around individual airports only provides a limited view on air transport in general [O'Connor and Fuellhart2016]. This is especially relevant in metropolitan areas, with increasing long-distance mobility demand being served by more than one airport. Therefore, the conceptual notion of Multi-Airport Region (MAR) is often used in air transport research, defined as a group of two or more major commercial airports in a metropolitan region [O'Connor and Fuellhart2016]. For instance, intentional airport substitution inside a MAR is an efficient way to mitigate supply demand imbalances. The concept of MAR also facilitates the study of resilience of air transport systems: if an airport in a MAR was disabled, alternative airports could be identified as backups. Recently, the idea of airport aggregation has been further taken to the level of provincial and country networks [Wandelt and Sun2015]. Investigation and comparison of different aggregation levels of air transportation networks, regarding their topological features and connectivity, can reveal hidden network properties.

In this study, we propose a new view on air transportation networks, taking different levels of aggregation into consideration. While other studies look at single such networks separately, we aim to provide a unified understanding of the transitions between airports/regions among network layers. As a case study, we investigate the worldwide air transportation network for the year 2015. Based on a

1 School of Electronic and Information Engineering, Beihang University, No. 37 Xueyuan Road, Beijing, China.
2 Beijing Key Laboratory for Network-based Cooperative ATM, Beijing, China.
* Corresponding author; E-mail: wandelt@buaa.edu.cn.

consistent global dataset, we first build the traditional airport network, for then obtaining all other aggregation levels through a contraction operation. In total, we derive aggregated network instances at six different levels: airports, cities, spatial distance of 100 km, spatial distance of 200 km, regional and country network. The aim of the analysis is to understand how the worldwide air transportation network evolves from fine-grained to coarse aggregation levels. Our results highlight some important facts, the main ones being that:

1. The network structure varies significantly among different aggregation levels, and specifically, aggregated networks have higher clustering coefficients and shorter average path length.
2. The importance of nodes gradually changes among aggregation levels, while few nodes are important in all of them. We also find that node degree and betweenness are weakly correlated at all aggregation levels.
3. Resilience analysis suggests that networks at higher aggregation levels are more robust than the fine-grained counterparts. Moreover, airports in Asia and Europe are often more important for the network robustness than other airports.
4. Most network communities follow geographical boundaries with few exceptions. The number of communities and their structures do not change significantly from fine-grained to coarse aggregation levels, which indicates that the community structure is a rather stable property of air transportation networks.

The remainder of our study is organized as follows. Section 2 presents our methodology to construct multi-scale air transportation networks with different aggregation levels. In Section 3, we present the results of multi-scale worldwide air transportation networks for the year 2015. Finally, conclusions are drawn in Section 4.

2 Data and Methodology

We extract the global air traffic data from Sabre Airport Data Intelligence (ADI, http://www.airdi.net) to build the aggregated worldwide air transportation network for year 2015. The data set contains information on a yearly basis for all commercial direct flights, including: origin/destination airports, number of passengers, revenue, average fare, and traffic type (scheduled or charter). This data is used to reconstruct the traditional airport network as follows: an airport is a node if it is either an origin or destination of any flight; and a link is created between two nodes if there is at least one direct flight between the corresponding airports. In general, all other aggregation levels are obtained by a contraction operation on the original airport network. Given a network with nodes N and links L, we define a mapping m, such that the domain of m is N and the range of m is the set of transformed nodes N_t. The links L_t are obtained by applying the mapping to the nodes of each link in the original network, *i.e.* if $(a,b) \in L$, then we have $(m(a), m(b)) \in L_t$. Below we describe how we derive the mapping m for all aggregation levels.

- **Airport network:** The mapping $m_{airport}$ is simply an identity mapping, *i.e.* all airports in the network are preserved.
- **City network:** The mapping m_{city} maps each airport to the city it belongs to. The information about the cities of the airports comes from the Sabre dataset.
- **Distance-based network of 100 km:** The mapping $m_{havd100}$, as inspired by previous research works on MARs [Bonnefoy2008, Bonnefoy et al. 2010, Wittman2014, O'Connor and Fuellhart2016], is created as follows. First, we sort the airports according to the total number of passengers in the year 2015, including both inbound and outbound passengers. Next, we iterate over all airports in the descending order of passengers and compute the Haversine distance to all

other airports. Once the distance to the larger airport is below a threshold δ, we assign that airport as belonging to the MARs of the larger airport. Finally, we mark the airport as being assigned to avoid reassigning it in the future.

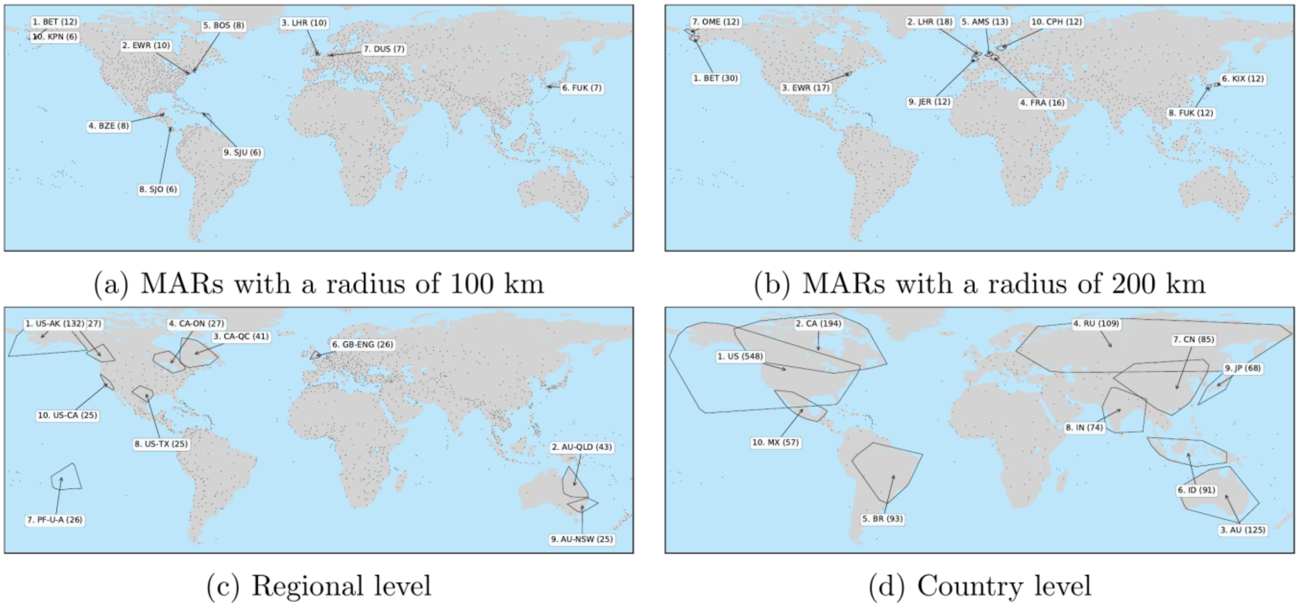
- **Distance-based network of 200 km:** The mapping $m_{havd200}$ is created similarly to $m_{havd100}$, with a threshold of $\delta = 200km$.
- **Regional network:** The mapping m_{region} is based on m_{city}, and maps each city to its ISO 3166-2 code. ISO 3166-2 is a short, unique alphanumeric code representing sub-national administrative territories of all countries in the world. In practice, the cooperating/competing regions can improve the overall system performance [Zhang et al. 2011b]. Each code consists of two sub-codes: a first ISO 3166-1 alpha-2 code of the country, and the second part containing up to three alphanumeric characters. For instance, the code US-AK stands for United States-Alaska.
- **Country network:** The mapping $m_{country}$ maps each airport to the country it belongs to, based on the region as codified by the ISO 3166-1 alpha-2 code. Following the previous example, the code US stands for United States.

Network metrics	**airport**	**city**	**havd100**	**havd200**	**region**	**country**
Number of nodes	3097	3034	2228	1373	1442	231
Number of links	50694	45701	38199	26948	29559	5675
Network density	0.01057	0.00993	0.0154	0.02861	0.02845	0.21363
Average node degree	32.7	30.1	34.3	39.3	41.0	49.1
Network radius	4	4	4	3	3	2
Network diameter	7	7	7	6	5	3
Average clustering coefficient	0.61278	0.61951	0.62552	0.62191	0.6281	0.69392
Average shortest path length	2.84307	2.83462	2.70962	2.56049	2.49874	1.82564

Table 1: Topological properties of worldwide air transportation networks at six different aggregation levels.

3 Results

We report the results of our experimental evaluation, following the methodology outlined in Section 2. We provide an overview on standard topological properties of the network at different scales in Section 3.1. Section 3.2 discusses how the importance of nodes gradually changes from fine-grained aggregations to coarser levels. In Section 3.3, we analyze the resilience of all networks against random failures and targeted attacks. The community structure of all networks is compared in Section 3.4.



(a) MARs with a radius of 100 km

(b) MARs with a radius of 200 km

(c) Regional level

(d) Country level

Figure 1: Top ten nodes based on the number of airports in air transportation networks with four different aggregation levels. Large regions are highlighted with their convex hull as a black line.

3.1 Network overview

First, we discuss and compare several relevant topological properties for the six network aggregation levels. The results are summarized in Table 1. As one may expect, the airport and city networks have the largest number of nodes; and, with further aggregation, the network's size and link density respectively decreases and increases. The country network, the coarsest aggregation, has 231 nodes and a 20 times higher density than the original network. Accordingly, the average node degree is increasing from around 30 to almost 50. Radius, diameter, and average shortest path length of the network are gradually decreasing with higher aggregation levels, meaning that the nodes are becoming closer to each other. The clustering coefficient is rather stable, ranging from 0.61 to 0.69.

We have performed additional experiments to analyze whether these derived networks can be considered as small-world. In a small-world network [Watts and Strogatz1998], nodes tend to form a large number of triangles, thus indicating a dense micro-scale connectivity; but, at the same time, this does not affect the average distance between pairs of nodes, which scales sub-linearly with the size of the network. This yields a simple two-fold criterion to assess the small-world property of a network: first, an average clustering coefficient significantly higher than an equivalent random network, constructed with the same number of nodes and links; second, a similar average shortest path length [Watts and Strogatz1998]. Our experiments show that all networks considered in this study are small-world. The average shortest path length of the air transportation network ranges from 1.8 (country) to 2.8 (airport, city); while the clustering coefficients are from 0.62 (airport, city) to 0.69 (country). For the random network counterparts, however, we observe clustering coefficients between 0.01 (airport, city) and 0.21 (country), i.e. significantly lower than in the real-world networks.

Figure 1 reports the top 10 nodes in the aggregated networks, ranked according to the number of airports in each of them mapped. The highest-ranked city is London, which is served by five (rather large) airports: STN (London Stansted), YXU (London Municipal Airport), LTN (London Luton Airport), LGW (London Gatwick Airport) and LCY (London City Airport). However, London is a rather exceptional case of a multi-airport city, given that the majority of cities only have one or two airports. The aggregation by Haversine distance is dominated by airport BET (Bethel Airport), located in Alaska, US, followed by LHR (London Heathrow, UK) and EWR (Newark Liberty International Airport, New Jersey, USA). In general, at the Haversine-distance aggregation level, we already find several larger airport regions. At the regional level, Alaska (US-AK) is ranked first, containing 132 airports. This latter result deserves further discussion. The size of areas identified by ISO 3166-2 codes is rather heterogeneous: while US-AK covers around 1.7 million km^2, other regions are significantly smaller - for instance DE-BR (Brandenburg in Germany) only covers 30,000 km^2. Therefore, the aggregation based on such regions should be understood from an administration-induced point of view, where the (regional) government can make decisions on the development of their airports, or at least provide incentives to set a direction for future development. Similarly, the nodes aggregated at country level are to be understood as under control of the national government and national authorities. Not surprisingly, US is ranked top in our list with more than 500 active airports used for passenger transportation. Altogether, it can be seen that the country ranking according to the number of airports largely coincides with their respective sizes, since larger countries often need more airports to ensure an efficient passenger transportation.

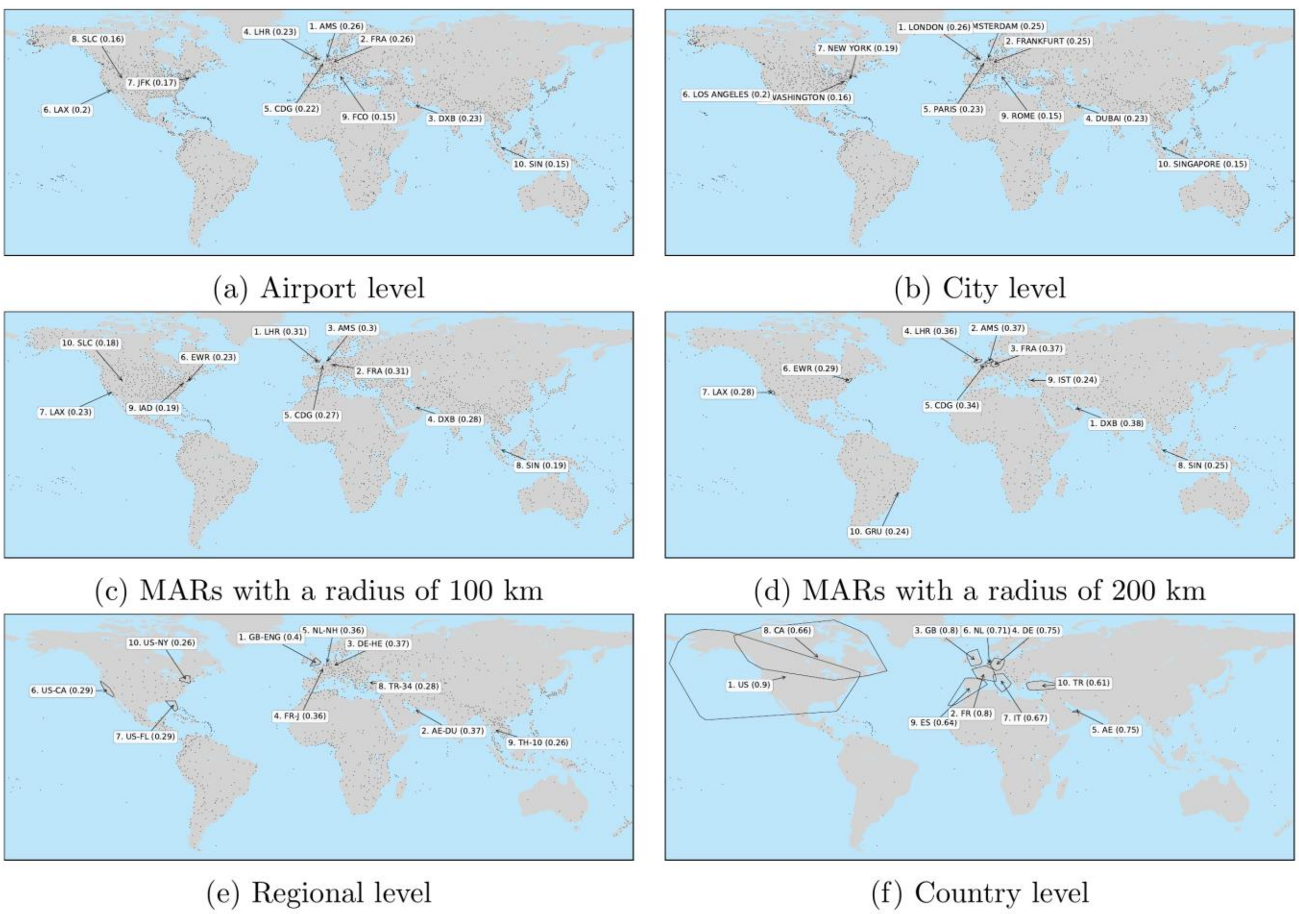


Figure 2: Top ten nodes based on degree centrality in air transportation networks with six different aggregation levels. Large regions are highlighted with their convex hull as a black line.

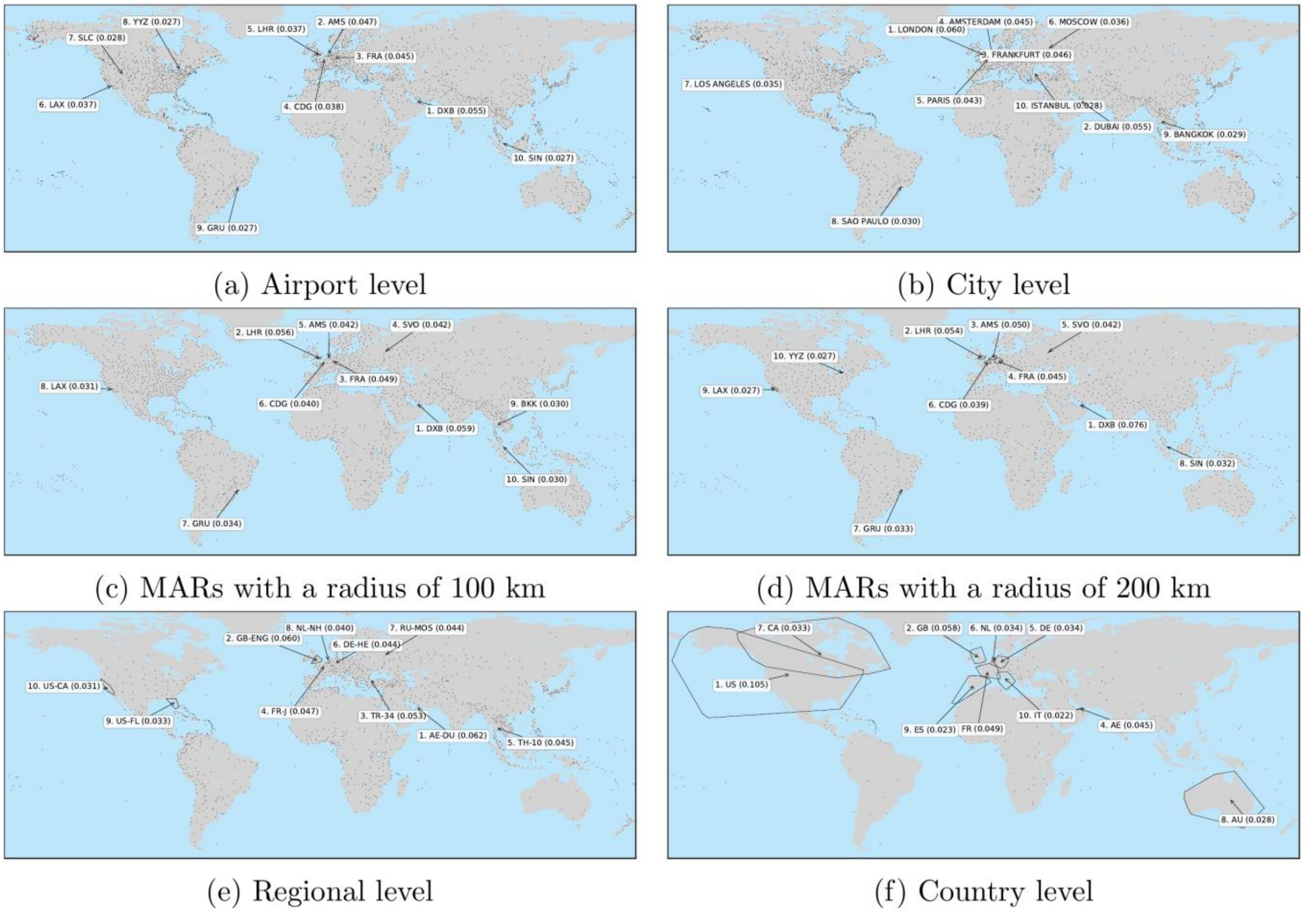


Figure 3: Top ten nodes based on betweenness centrality in air transportation networks with six different aggregation levels. Large regions are highlighted with their convex hull as a black line.

3.2 Node importance

After having discussed some global properties of the reconstructed networks, we here descend to the micro-scale to analyze the importance of individual nodes. Several standard metrics are available to quantify the node importance from a complex network point of view, the most important and well-known being the degree and the betweenness centrality. The degree centrality of a node reflects its number of direct neighbors, normalized by the total number of nodes. Passengers of an airport with a high degree centrality can reach more airports within one step, *i.e.* using a single flight. The betweenness centrality, on the other hand, measures how frequently a node appears on the shortest paths of a network.

We first discuss the degree centrality. Figure 2 presents the top ten nodes ranked according to this metric, for the six aggregated air transportation networks. One can observe that the majority of important nodes is located in Western Europe and Northern America. London and Amsterdam, for instance, are extremely influential according to their degree centrality. We also report the results for betweenness centrality. Figure 3 presents the top ten nodes ranked according to this metric. One can observe that the majority of important nodes is located in Western Europe and Northern America.

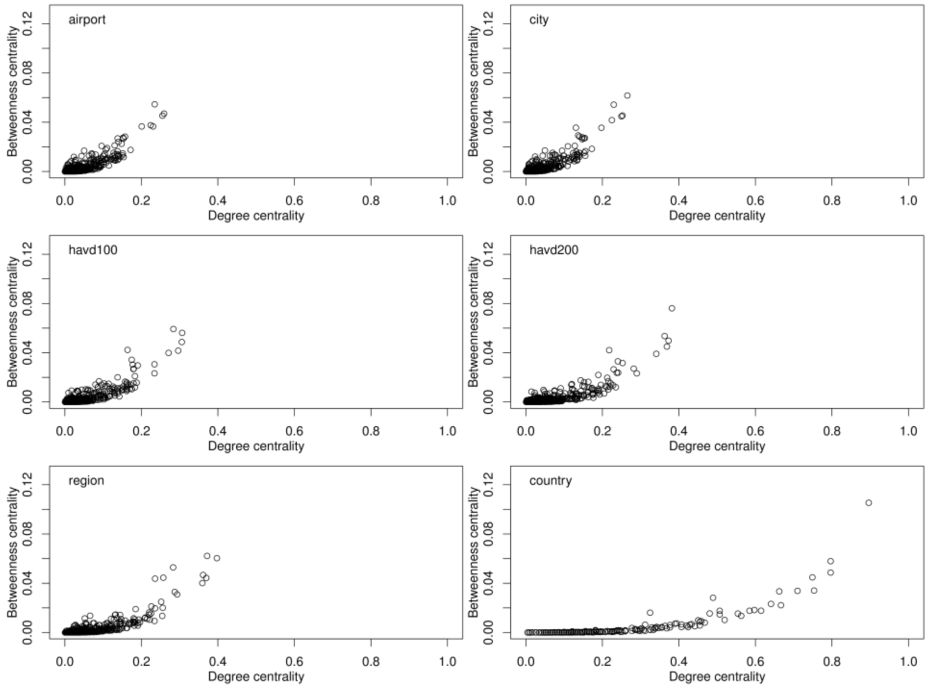


Figure 4: Scatter plot of degree centrality vs. betweenness centrality in the six aggregated air transportation networks. There is no strong correlation between the degree centrality of a node and its betweenness centrality in the networks, expect for the case of country network.

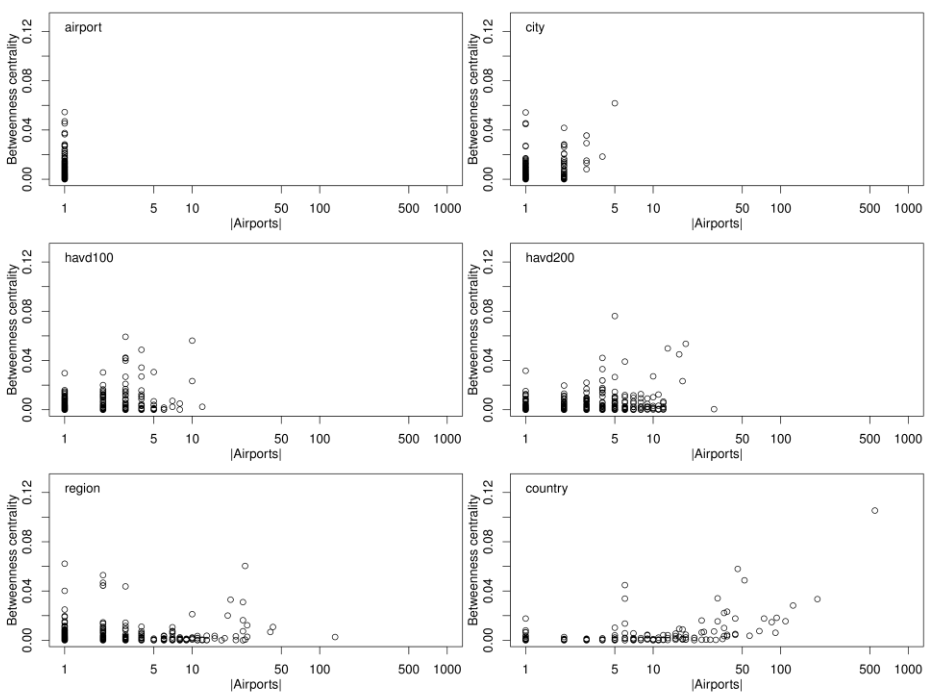


Figure 5: Scatter plot between the number of airports and betweenness centralities in the six aggregated air transportation networks at a semi-log scale. With the levels of aggregation going from single airports to country entities, betweenness centralities do not scale up.

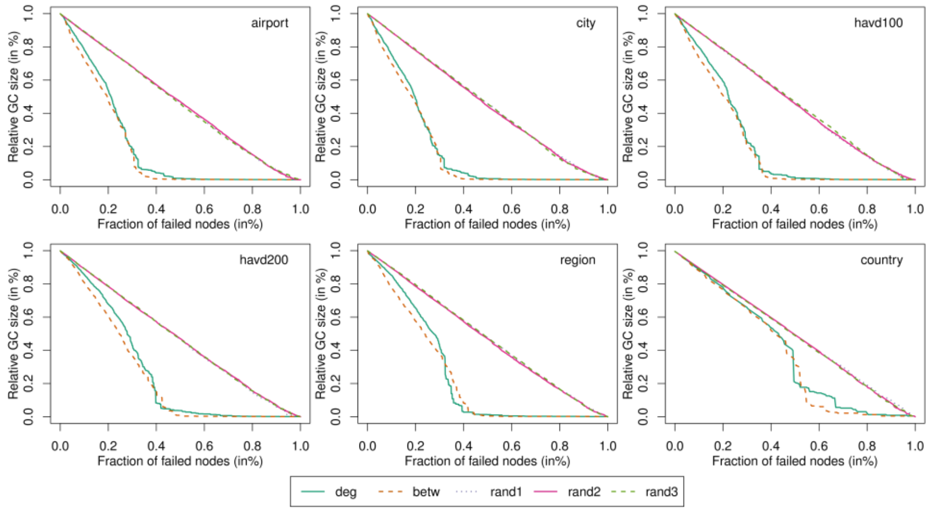
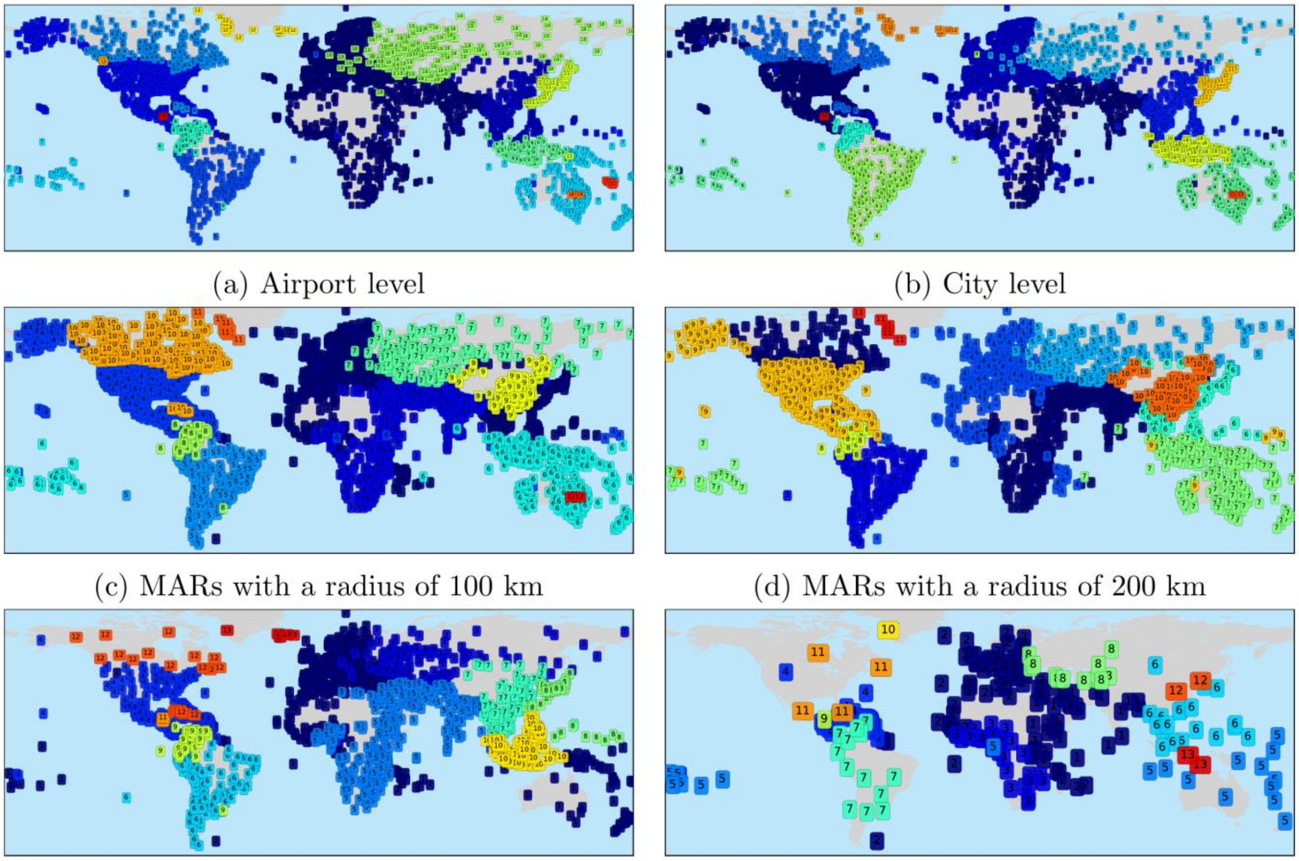


Figure 6: Robustness analysis of the six aggregated air transportation networks: Intentional attacks (degree-based (deg), betweenness-based (betw), and the number of airports for an aggregated node) as well as random failures (rand1, rand2, and rand3 are three random attacks to the network).

Figure 4 presents the scatter plot of the degree *vs.* the betweenness centrality in the six aggregated air transportation networks. All aggregation layers present a weak linear correlation (R^2 between 0.64 and 0.71) between both centralities, with the exception of a quadratic relationship in the case of the country network (R^2 of 0.90); both metrics thus agree on the identification of the most important hubs. An interesting question is whether the betweenness centrality of a node is strongly connected to the number of airports aggregated into that node; in other words, do more airports lead to higher centrality values? Figure 5 shows the scatter plots between the number of airports and betweenness centralities in the six aggregated air transportation networks at a semi-log scale. Surprisingly, no correlation is observed, thus indicating that putting more airports into a region does not necessarily increase the importance of that region.

3.3 Network robustness

Air transport, like all other critical infrastructures, is required to maintain the highest fault tolerance [Freeman et al. 2013, Zhang et al. 2011a, Lee et al. 2014], since disruptions have huge economic and societal impacts [Ball et al. 2006]. For instance, due to an overnight snowstorm on March 12, 2013, Frankfurt airport was closed and airlines cancelled about 700 flights. The 2010 ash cloud over Europe, caused by eruption of Icelandic volcano Eyjafjallajökull, is estimated to have caused losses of approx. 3.3 billion Euro for large European airlines [Mazzocchi et al. 2010]. Therefore, analysis of the robustness is a critical and important issue in air transport, especially for the deployment of future improvements. In this study, we quantify the robustness of the network against random failures and targeted attacks by measuring the size of the giant component, as often performed in related studies [Lordan et al. 2014, Wang et al. 2014].



(a) Airport level (b) City level

(c) MARs with a radius of 100 km (d) MARs with a radius of 200 km

(e) Regional level (f) Country level

Figure 8: Community structures in the six aggregated air transportation networks with the Louvain method, different colors represent different communities. The communities of the networks largely coincide with the administrative boundaries of regions. The number of communities is stable for all aggregation levels.

Figure 6 presents the results of the robustness analysis of the six aggregated air transportation networks, for both intentional and random attacks. Intentional attacks disable nodes based on certain network metrics: we here consider node degree and betweenness, as previously defined. Essentially, the robustness curve of the networks for random failures is along the diagonal line, which means that worldwide air transportation networks are rather robust against random failures. Under intentional attacks, as induced by degree and betweenness rankings, the network breaks down into isolated components much faster. In general, networks at higher aggregation levels, e.g., distance-based MARs and country networks, are more robust than the fine-grained counterparts, e.g., airport and city networks. One possible explanation is that the aggregated networks have a higher link density, see Table 1.

3.4 Communities

In complex networks, the notion of community has been developed to identify nodes which are densely connected within a group, yet sparsely connected with other groups in the network. The detection of communities helps to understand the underlying structures of the network and identifying hidden properties between nodes [Palla et al. 2005, Zhou2003]. In this study, we use the widely-known Louvain method [Blondel et al. 2008] to identify communities in the network. The Louvain method is a greedy one, in that it attempts to optimize the modularity of partitions in the network using a sequential procedure, which starts with smaller communities and gradually analyzes larger ones. Figure 8 presents the community structures in the six aggregated air transportation networks. We find that the communities of the networks largely coincide with the administrative regional boundaries. The number of communities is stable for all aggregation levels (usually around 10 communities are identified by the Louvain method). One interesting insight is how the West-African coast belongs to the community of Western Europe in all aggregation levels, while the rest of Africa is strongly connected to the Arabian region, with connections up to India.

4 Conclusions

In this study, we have explored the complex network properties of the worldwide air transportation network at six aggregation levels. At each level, nodes have either been aggregated according to a distance measure, or to the corresponding administrative boundary. The major findings are summarized below.

1. **Network structure:** The network structure varies significantly between different aggregation levels. The number of nodes ranges from 231 in country network to 3097 in the airport network. The country network is much denser (21%) than the city network (1%).
2. **Node importance:** The importance of nodes in aggregated networks changes gradually with the aggregation level. While few nodes are important at all levels of aggregation, *e.g.* LHR/London/Great Britain or JFK/New York/US, others only become important at some specific scales, *e.g.* SIN/Singapore or Canada.
3. **Network robustness:** The degree and the betweenness of a node provide necessary information to reduce the functionality of the network, as measured by the size of the giant component. In general, networks at higher aggregation levels, *e.g.* distance-based Multi-Airport regions and country networks, are more robust than the fine-grained counterparts, *e.g.* airport and city networks.
4. **Communities:** Previous research works were not conclusive on whether communities induced by the air transportation structure coincide with geographical and political boundaries. We find that indeed most communities follow geographical boundaries, with few exceptions.

Our current study focuses on topological properties of the worldwide air transportation networks at six different aggregation levels; as well as how the networks evolve from fine-grained to coarse granularity. In future works, the analysis could be complemented with weight information, using for instance information about the number of passengers or number of flights for each direct connection. The temporal evolution of the multi-scale properties of the network could also be studied further, in order to derive policy implications [Gil-Alana et al. 2013]. Moreover, additional measures of network robustness could be used [Chow et al. 2015, Pien et al. 2015, Wandelt et al. 2017]. Finally, our research on network aggregation can be taken over to other modalities, e.g., public transit networks from GTFS-encoded datasets [Wandelt et al. 2016a] or railway networks [Wandelt et al. 2016b].

Acknowledgement

This study is supported by the National Natural Science Foundation of China (Grant No. 61650110516 and Grant No. 61601013).

References

Azzam, M., Klingauf, U., and Zock, A. (2013). "The accelerated growth of the worldwide air transportation network." *The European Physical Journal Special Topics*, 212, 35–48.

Ball, M., Barnhart, C., Nemhauser, G., Odoni, A., and Laporte, G. (2006). "Air transportation: Irregular operations and control." *Transportation*, (14), 1–67.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment*, 10, P10008.

Bonnefoy, P. (2008). "Scalability of the air transportation system and development of multi-airport systems: A worldwide perspective." Ph.D. thesis, Massachusetts Institute of Technology,

Bonnefoy, P., de Neufville, R., and Hansman, R. (2010). "Evolution and development of multiairport systems: Worldwide perspective." *Journal of Transportation Engineering*, 136(11), 1021–1029.

Chow, A. H., Szeto, W., Wang, D. Z., and Waller, S. T. (2015). "Quantitative approaches to resilience in transport networks." *Transportmetrica A: Transport Science*, 11(9), 751–753.

Cook, A., Blom, H. A., Lillo, F., Mantegna, R. N., Micciche, S., Rivas, D., Vázquez, R., and Zanin, M. (2015). "Applying complexity science to air traffic management." *Journal of Air Transport Management*, 42, 149–158.

Freeman, P., Seiler, P., and Balas, G. J. (2013). "Air data system fault modeling and detection." *Control Engineering Practice*, 21(10), 1290 – 1301.

Gil-Alana, L. A., Barros, C. P., and Assaf, A. G. (2013). "The us airline industry: persistence and breaks." *Transportmetrica A: Transport Science*, 9(8), 742–752.

Lee, S. Y., Yoo, K. E., and Park, Y. (2014). "A continuous connectivity model for evaluation of hub-and-spoke operations." *Transportmetrica A: Transport Science*, 10(10), 894–916.

Lordan, O., Sallan, J. M., Simo, P., and Gonzalez-Prieto, D. (2014). "Robustness of the air transport network." *Transportation Research Part E: Logistics and Transportation Review*, 68(0), 155–163.

Mazzocchi, M., Hansstein, F., Ragona, M., et al. (2010). "The 2010 volcanic ash cloud and its financial impact on the european airline industry." *CESifo Forum*, Vol. 11, Ifo Institute for Economic Research at the University of Munich, 92–100.

O'Connor, K. and Fuellhart, K. (2016). "Airports and regional air transport markets: A new perspective." *Journal of Transport Geography*, 53, 78 – 82.

Palla, G., Derényi, I., Farkas, I., and Vicsek, T. (2005). "Uncovering the overlapping community structure of complex networks in nature and society." *Nature*, 435(7043), 814–818.

Pien, K.-C., Han, K., Shang, W., Majumdar, A., and Ochieng, W. (2015). "Robustness analysis of the european air traffic network." *Transportmetrica A: Transport Science*, 11(9), 772–792.

Sun, X., Gollnick, V., and Wandelt, S. (2017a). "Robustness analysis metrics for worldwide airport network: A comprehensive study." *Chinese Journal of Aeronautics*, –.

Sun, X. and Wandelt, S. (2014). "Network similarity analysis of air navigation route systems." *Transportation Research Part E: Logistics and Transportation Review*, 70, 416–434.

Sun, X., Wandelt, S., and Linke, F. (2015). "Temporal evolution analysis of the european air transportation system: air navigation route network and airport network." *Transportmetrica B: Transport Dynamics*, 3(2), 153–168.

Sun, X., Wandelt, S., and Linke, F. (2017b). "On the topology of air navigation route systems." *Proceedings of the Institution of Civil Engineers - Transport*, 170(1), 46–59.

Wandelt, S. and Sun, X. (2015). "Evolution of the international air transportation country network from 2002 to 2013." *Transportation Research Part E: Logistics and Transportation Review*, 82, 55–78.

Wandelt, S., Sun, X., and Cao, X. (2015). "Computationally efficient attack design for robustness analysis of air transportation networks." *Transportmetrica A*, 11(10), 939–966.

Wandelt, S., Sun, X., Zanin, M., and Havlin, S. (2017). “QRE: Quick robustness estimation for large complex networks.” *Future Generation Computer Systems*, –.
Wandelt, S., Sun, X., and Zhu, Y. (2016a). “Lossless compression of public transit schedules.” *IEEE Transactions on Intelligent Transportation Systems*, 17(11), 3075–3086.
Wandelt, S., Wang, Z., and Sun, X. (2016b). “Worldwide railway skeleton network: Extraction methodology and preliminary analysis.” *IEEE Transactions on Intelligent Transportation Systems*, PP(99), 1–11.
Wang, H., Huang, J., Xu, X., and Xiao, Y. (2014). “Damage attack on complex networks.” *Physica A: Statistical Mechanics and its Applications*, 408(0), 134–148.
Watts, D. J. and Strogatz, S. H. (1998). “Collective dynamics of ’small-world’ networks.” *Nature*, (393), 440–442.
Wei, P., Chen, L., and Sun, D. (2014). “Algebraic connectivity maximization of an air transportation network: The flight routes addition/deletion problem.” *Transportation Research Part E: Logistics and Transportation Review*, 61, 13–27.
Wittman, M. D. (2014). “An assessment of air service accessibility in u.s. metropolitan regions, 2007-2012.” *Report no.*, International Center for Air Transportation (ICAT), Department of Aeronautics & Astronautics, Massachusetts Institute of Technology.
Zanin, M. (2015). “Can we neglect the multi-layer structure of functional networks?.” *Physica A: Statistical Mechanics and its Applications*, 430, 184–192.
Zanin, M. and Lillo, F. (2013). “Modelling the air transport with complex networks: A short review.” *The European Physical Journal Special Topics*, 215(1), 5–21.
Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., and Chen, C. (2011a). “Data-driven intelligent transportation systems: A survey.” *Intelligent Transportation Systems, IEEE Transactions on*, 12(4), 1624–1639.
Zhang, X., Zhang, H., Huang, H.-J., Sun, L., and Tang, T.-Q. (2011b). “Competitive, cooperative and stackelberg congestion pricing for multiple regions in transportation networks.” *Transportmetrica*, 7(4), 297–320.
Zhou, H. (2003). “Distance, dissimilarity index, and network community structure.” *Physical review e*, 67(6), 061901.