Air Passenger Flow Communities Between Countries

SUN Xiaoqian[1,2], WANDELT Sebastian[1,2], DZIKUS Niclas[3] and LINKE Florian[3]

1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, P. R. China

2. National Key Laboratory of CNS/ATM, Beijing 100191, P. R. China
E-mail: sunxq@buaa.edu.cn, wandelt@informatik.hu-berlin.de

3. Institute of Air Transportation Systems, German Aerospace Center, Blohmstrasse 18, 21079 Hamburg, Germany
E-mail: niclas.dzikus@dlr.de, florian.linke@dlr.de

Abstract: In this study, we perform an empirical analysis of air transportation systems as networks of countries. We build three network variants based on worldwide ticket data from the year 2013 and identify patterns of air passenger flow between countries as induced by network centrality metrics. These metrics, understood as observables about countries, are then exploited to estimate the similarity of countries in the global network. We detect community structures and find that communities largely coincide with the continental division of the world. Our research contributes to a better understanding of international air passenger flow patterns and potential control mechanisms against natural or human intended disruptions.

Key Words: Air Transportation Networks; Passenger Flow; Community Detection

1 Introduction

Transportation infrastructure forms the backbone for tourism and enabling movements of all sorts of goods and people across countries [1, 2]. Particularly, aviation networks are an essential part of public transportation systems due to their convenience [3]. Complex network theory [4, 5] provides powerful tools to help us understand the structures and dynamics of air transportation systems. Airport networks [6] and air route networks [7] have been the main subjects of previous studies;. for instance, analyzing various centrality measures of individual airports [8], delay propagation [9], epidemics spreading [10], network robustness [11], and temporal evolution of the network [12, 13].

Analyzing the airport network is the natural choice for air transportation system analysis. City networks, where airports inside metropolitan regions are merged into a single node, incorporate the possibility of urban transit inside cites, and thus, provide a different view on the network. In this study we extend the idea of node aggregation, by analyzing the worldwide country network: A network which is obtained by merging all airports within a country into a single node. This is not only a natural extension of city-level networks; we believe that the role of country borders plays an integral role when it comes to global transportation, despite all attempts to simplify the process of crossing borders inside, for instance, Europe. Particularly, we expect that, in the future, countries will increasingly exploit their powers in global infrastructure networks to enforce their own political viewpoints. We list a few recent examples only:

- Refugee movements from south-east Europe to the north west in 2015 have led to a long-forgotten awareness of country borders all over Europe: Many European countries attempt to block the flow of refugees to (or through) their countries. The sole criteria for letting people enter a country is often based on the origin country of a person, which underlines the importance of the concept *country* in global transportation.
- During the peak of the Ukrainian crisis, Russia blocked its airspace for flights originating from European countries, leading to huge economic losses and delays for major European airlines.
- The outbreak of Ebola starting West Africa in 2014 has caused humanitarian disaster. Flight restrictions have been announced in order to control the international spread of this disease from these West African countries to other countries in the world. These restrictions were usually formulated based on *country borders*, e.g., not allowing people from Liberia and Guinea to cross, for instance, Spanish/US borders.

In this paper, we provide an empirical analysis of air passenger flow between countries and detect community structures for the year 2013. The main contributions of our paper are:

1) We build three network variants of international air passenger transportation: Physical network (network on a flight-level), reciprocal passenger flow network (an origin-destination network with equal flow between countries), and net positive passenger flow network (encoding the deviation of origin-destination flow).
2) We identify the role of countries by using standard network metrics for air transportation country networks.
3) Our detection of community structures in the country network shows that communities largely coincide with the continental division of the world. This is surprising, given that previous studies at the airport/city level suggested that transportation communities are independent

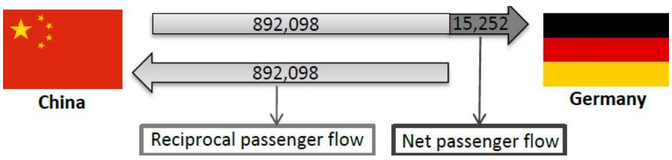

Fig. 1: An example of reciprocal passenger flow and net passenger flow between China and Germany in year 2013. In total, 892,098 passengers travelled in both directions, while 15,252 passengers create a net passenger flow from China to Germany.

This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61521091.

Table 1: The top five countries ranked by nine weighted metrics in three network variants: Degree (D_W), in/out degree (Di_W, Do_W), betweenness (B_W), closeness (C_W), current flow betweenness (CFB_W), current flow closeness (CFC_W), clustering coefficient (CC_W), and eigenvector (EV_W) centralities, separately. The link weight is the number of passengers.

Rank	**D_W**	**Di_W**	**Do_W**	**B_W**	**C_W**	**CFB_W**	**CFC_W**	**CC_W**	**EV_W**
					Physical network				
1	GB	GB	GB	GB	GB	US	GB	TW	GB
2	US	US	US	US	ES	AE	DE	PH	ES
3	DE	DE	DE	AE	US	GB	US	ID	DE
4	ES	ES	ES	DE	DE	FR	ES	SG	US
5	CN	CN	CN	FR	CA	DE	FR	IE	FR
					Reciprocal passenger flow network				
1	US	US	US	GB	GB	US	GB	GB	GB
2	GB	GB	GB	US	US	GB	US	DE	ES
3	DE	DE	DE	ES	ES	FR	DE	US	DE
4	ES	ES	ES	FR	DE	DE	ES	ES	US
5	CN	CN	CN	DE	CA	CN	FR	FR	IT
					Net positive passenger flow network				
1	US	US	CN	GB	KR	US	US	SY	DE
2	CN	DE	IN	US	GB	FR	GB	US	US
3	DE	SA	GB	FR	CN	DE	DE	DE	JP
4	GB	TH	KR	CN	IT	CN	CN	KH	GB
5	SA	IT	US	IT	FR	GB	FR	GB	TH

of borders.

4) We investigate the correlation between topological properties and socio-economic measures. We find a strong correlation between weighted network degree and the GDP of a country.

The remainder of this paper is organized as follows. Section 2 provides an overview of the country network and its topological properties. The detection of community structures in the country network is presented in Section 3. The paper concludes with Section 4.

2 Overview of the Country Network

2.1 Data Preparation

We build a weighted country network based on the ticket data from the Sabre Airport Data Intelligence (ADI, http://www.airdi.net) in the year 2013, the ticket data is stored monthly and contains three types of information:

1) Origin/Destination airports
2) Up to three connecting airports
3) The number of passengers who bought this ticket in one month.

In the country network, all airports of a country are aggregated into one node, the link between two countries is weighted by the total number of traveled passengers (the sum of all passengers traveling between airports of these two countries). In the current study, we consider links only which have more than 100,000 passengers traveling per year; setting a threshold is a common way to prioritize passenger-relevant connection in transportation. After performing all steps of our data pre-processing, we obtain 164 nodes in the country network. Below, we derive three instance for this air passenger country network.

Table 2: Three variants of worldwide air transportation country network, where ”pax” stands for passenger; ASPL represents average shortest path length.

	Nodes	Links	Density	ASPL
Physical network	164	4,496	0.168	1.989
Reciprocal pax flow network	164	13,602	0.509	1.492
Net positive pax flow network	164	6,850	0.301	1.766

2.2 Three Network Variants of the Country Network

In [14, 15], it is argued that for a comprehensive study of air transport, both aspects should be taken into account: direct flight connections and passenger flow from origin to destination. Adapted from the terminology in the migration field [16], we define the number of passengers flying in both directions as the reciprocal passenger flow; and the net passenger flow is the number of passengers flying in one direction and do not return. As illustrated in Fig. 1, in the year 2013, 892,098 passengers traveled in both directions, while 15,252 passengers create a net passenger flow from China to Germany. We define three network variants as follows.

- **Physical network**: a directed network, where a ticket with multiple hops is accounted for each hop separately.
- **Reciprocal passenger flow network**: a directed network with equal origin-destination passenger flow between two countries.
- **Net positive passenger flow network**: a directed network with net weights of origin-destination pairs and there is at most one link between any two countries.

We would like to have an overview of the three network variants. As shown in Table 2, the reciprocal passenger flow network is the densest one and has the smallest average shortest path length. Our three variants of country networks help to understand the importance of each country in global air transportation systems from different perspectives.

2.3 Standard Centrality Measures in the Country Network

Different centrality measures often generate different importance rankings, Table 1 presents the rankings of the countries according to the nine weighted network centrality measures. A general trend is that United Kingdom (GB) ranks first regarding the majority of the network metrics in the physical network (7 out of 9 metrics) and reciprocal passenger flow network (5 out of 9 metrics). Also, United States

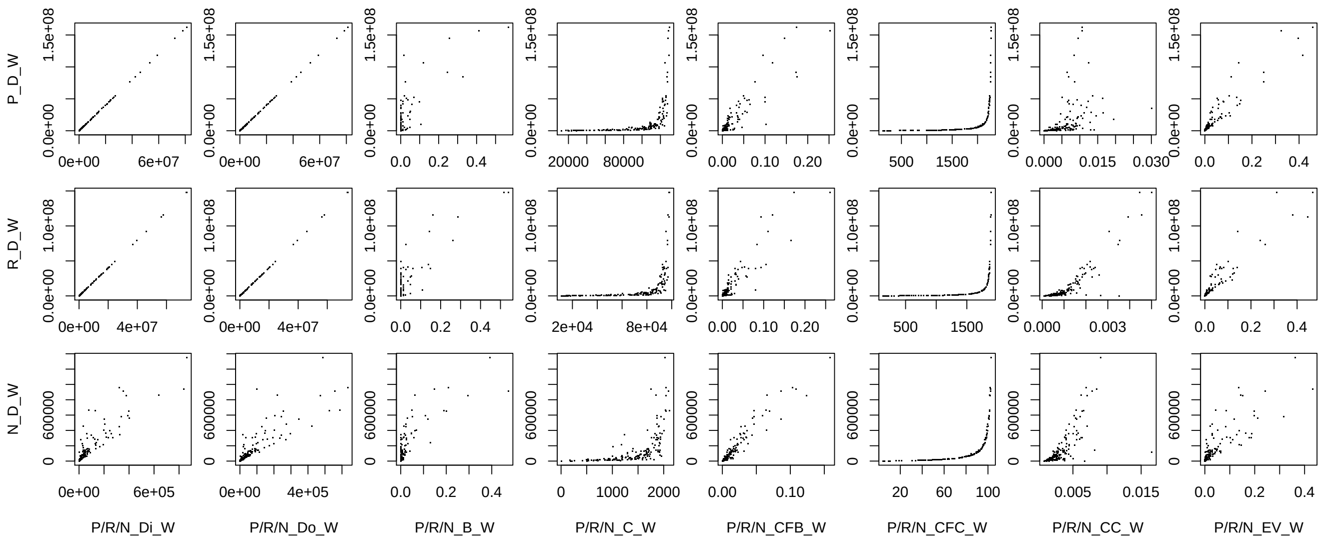

Fig. 2: Scatterplot of weighted degree against other weighted metric values, using the same symbols as in Table 1.

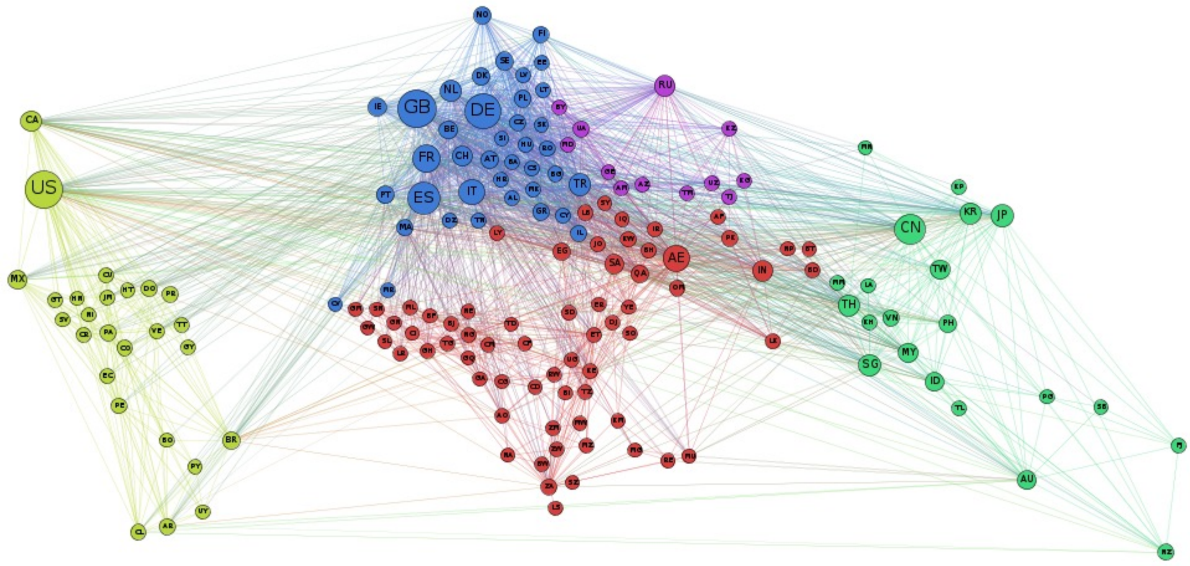

Fig. 3: Community structure for the physical network. Each colour represents one modularity class, the size of a node is proportional to its weighted degree. Communities in the physical network largely coincide with the division of countries into continents. This shows that intra-continental traffic is prevalent, with few exceptions only. Most notably, countries from North-Africa belong to the European community.

(US) is often ranked close to United Kingdom.

We take the net positive passenger flow network as an example to explain the practical meaning of these metrics. United States (US) receives the highest number of passengers from the world (Di_W); While China (CN) has the highest number of passengers leaving this country (Do_W). When considering the neighbour's connections of a country (EV_W), Germany (DE) outranks United States (US) receiving the highest number of passengers. This is because Germany has several strong neighbours, such as United Kingdom, Italy, and France. These neighbours also receive large number of passengers which enables higher value of eigenvector centrality for Germany. United States (US) has the highest scores of current flow betweenness (CFB_W) and current flow closeness (CFC_W). South Korea (KR) is closest to all other countries where people only travel to and do not return (C_W); while Syria (SY) is most clustered among these countries (CC_W). The latter observation about Syria is rather interesting, since our data is for the year 2013; thus; much earlier than the (current) peak of the refugee movement. This indicates that analysis about the country level are valuable for detecting population-moving trends early.

In addition, it is also interesting to check the relationship among weighted network metrics. Fig. 2 presents the scatter plot between weighted degree against another eight weighted metric values. It can be observed that in the physical and reciprocal passenger flow networks, weighted degree and in/out degree are linearly correlated; while weighted degree and current flow closeness have strong positive correlation in three network variants.

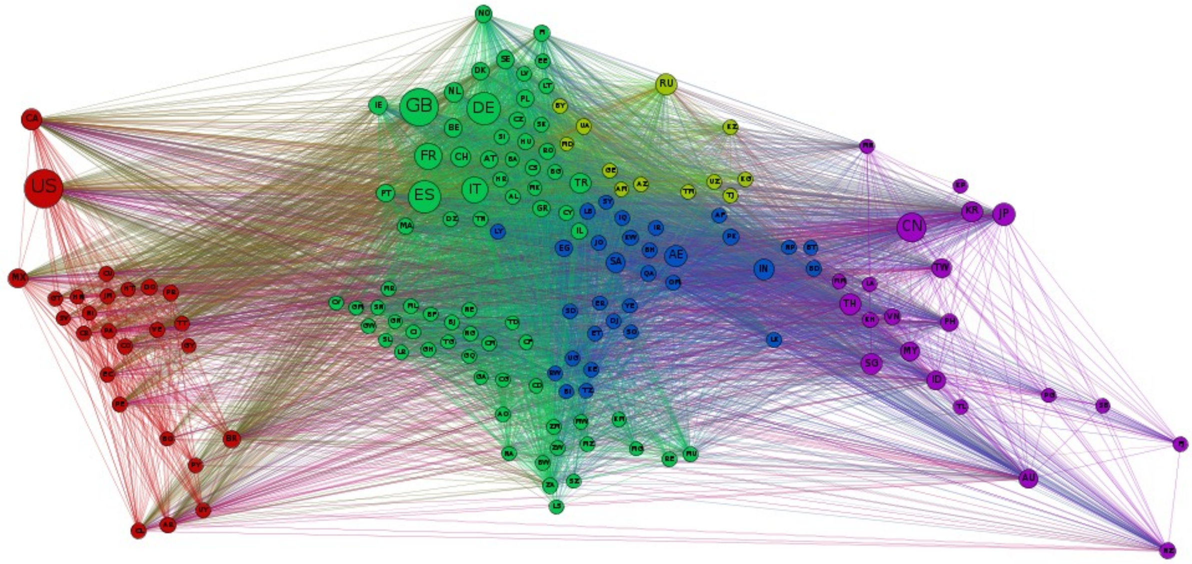

Fig. 4: The community structures for the reciprocal passenger flow network. Each colour represents one modularity class; the size of the node is proportional to its weighted degree. The top 5 ranked nodes based the weighted degree are US (United States), GB (United Kingdom), DE (Germany), ES (Spain), and CN (China).

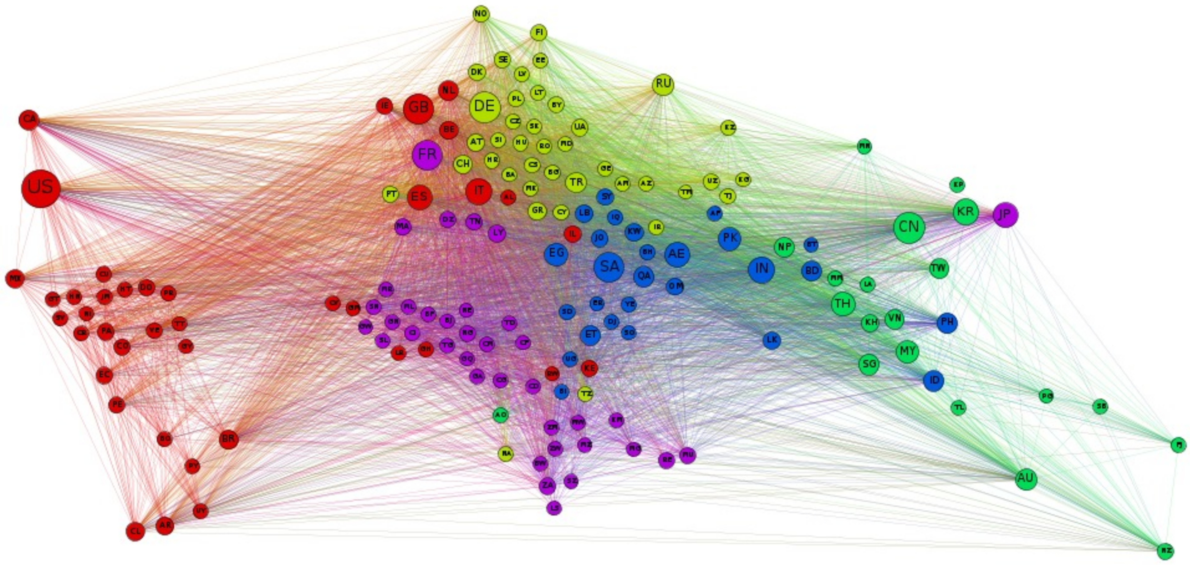

Fig. 5: The community structures for the net positive passenger flow network. Each colour represents one modularity class, the size of the node is proportional to its weighted degree. The top 5 ranked nodes based the weighted degree are US (United States), CN (China), DE (Germany), GB (United Kingdom), and SA (Saudi Arabia).

3 Community Structures in the Country Network

In complex networks, communities are groups of nodes which are densely connected within the group and sparsely connected with other groups. The detection of communities helps to understand the underlying structures of the network and hidden properties between nodes [17]. We use the Louvain method [18] to identify communities in the country network, which is a greedy method attempting to optimize the modularity of partitions in the network.

Fig. 3 presents the results of community detection for the physical network, where different colours represent different modularity classes and the size of a node is proportional to its weighted degree. The codes of countries are according to `http://www.worldbank.org`. We can observe that there are five communities among the 164 countries and the communities largely coincide with the continental division of the world.

In the modularity class 0 (in lime colour), 15 countries from North America and 11 countries from South America are clustered together, indicating that air traffic between North America and South America is stronger than to the rest of the world. Similarly, 6 countries from Oceania are grouped together with other 15 Asian countries in the modularity class 1 (in green colour), with the indication of strong air traffic between Oceania and Asia.

	Modularity class – physical					Modularity class – reciprocal					Modularity class – net-positive				
Continent	**0**	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**	**0**	**1**	**2**	**3**	**4**
AF				47	5			40	12		6	2	8	35	1
AS		15	8	19	2		15	2	19	8	1	10	19	1	13
EU			4		32			32		4	7	28		1	
NA	15					15					15				
OC		6					6								6
SA	11					11					11				
SUM	**26**	**21**	**12**	**66**	**39**	**26**	**21**	**74**	**31**	**12**	**40**	**40**	**27**	**37**	**20**

Fig. 6: Summary of the community structures in the three networks. The cell in green colour indicates that countries from the same continent are in the same community; while the cell in yellow colour indicates that irregularities happen in the sense that countries from different continents appear in the same community.

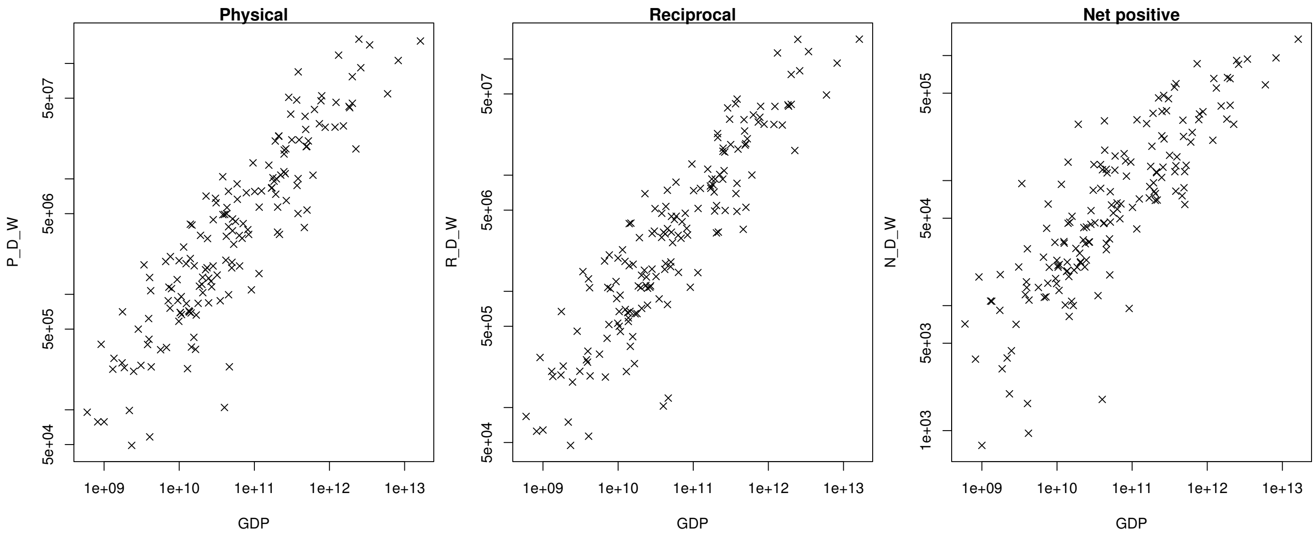

Fig. 7: GDP versus weighted degree at log-log scale for the three networks. A power law correlation is observed: $P\_D\_W = 0.0133 * GDP^{0.7824}, R^2 = 0.81$ (left: Physical); $R\_D\_W = 0.0081 * GDP^{0.7957}, R^2 = 0.82$ (middle: Reciprocal); $R\_D\_W = 0.019 * GDP^{0.605}, R^2 = 0.71$ (right: Net Positive).

We find out that in the modularity class 2 (in magenta colour), 12 out of 15 former Soviet countries are clustered together, with Russia in the central position. Note that the other three former Soviet countries: Estonia (EE), Latvia (LV), and Lithuania (LT) are in the cluster of European countries (in blue colour). Although the collapse of the Soviet Union happened two decades ago, there are still close connections among these former Soviet countries in global air transportation.

It is interesting to note that in the modularity class 3 (in red colour), several Western-Asian countries are grouped with the majority of African countries. We think that the increasing economic ties between Asia and Africa enable strong mobility connections [19, 20].

Most European countries are clustered in the modularity class 4 (in blue colour), which suggests strong intra-European air traffic. We discover that in the European community, there are two Asian countries (Israel and Turkey) and five Northern-African countries (Algeria, Cape Verde, Mauritania, Morocco, and Tunisia). The strong economic ties between Israel, Turkey, and Europe explain well why they are in the same community. Among the five Northern-African countries, Algeria, Mauritania, and Tunisia were French colonies; Cape Verde was one Portuguese colony; Morocco was colonized by France and Spain. Besides for having been colonies, most European people prefer short flights into *warm* Northern-Africa for vacation than longer flights to the Southern-Africa. The economic cooperation and industrial modernization enables close relations of these countries with Europe.

In Fig. 4 and Fig. 5, we present the results of community detection for the reciprocal and net positive passenger flow network, respectively. We observe similar patterns that countries from the same continent are often clustered together. In the reciprocal passenger flow network, the countries from North America and South America are still in the same community. The situations hold for the countries from Oceania-Asia and the former Soviet countries. However, most African countries join the European community; while the remaining African countries cluster together with Asian countries. In the net positive passenger flow network, several European countries and a few African countries join the group of North/South American countries; while the former Soviet countries clustered together with the European community. Interestingly, France and Japan join the African community. Several African countries are former colonies and there is a long history of people moving from/to Africa/France. Japan has recently increased its effort to adopt China's successful investment strategy in Africa [21].

Fig. 6 summarizes the results of the community structures in the three networks. Among the three networks, the North/South American community and the Asia/Oceania community maintain stable; while the European community and African community are not stable.

Furthermore, we also look at the relationship between the topological properties of a country and its socio-economic characteristics. For instance, Fig. 7 shows the scatter plot of GDP and weighted degree for the 164 countries in the three network variants, respectively. There is a power law correlation between a country's GDP and its weighted degree ($P\_D\_W = 0.0133 * GDP^{0.7824}, R^2 = 0.81; R\_D\_W = 0.0081 * GDP^{0.7957}, R^2 = 0.82; R\_D\_W = 0.019 * GDP^{0.605}, R^2 = 0.71$). It is interesting to investigate further other socio-economic factors, such as unemployment rate and political situations [23].

4 Conclusions

Our study provides an empirical analysis of international air passenger flow between countries. We constructed three network variants: Physical network, reciprocal passenger flow network, and net positive passenger flow network. We identified the role of countries by using standard network metrics for air transportation networks. Our detection of community structures in the country network showed that communities largely coincide with the continental division of the world, a finding different to what can be observed about airport/city-level networks. The construction of the worldwide country air transportation network helps us to identify the role of each country in the global air transportation systems. It also reveals the travel patterns of international passenger flows between countries and provides a platform for efficient modeling of global diseases spreading, such as Ebola [22], since the scale of the network is reduced by two orders of magnitudes (from more than 8,000 nodes to around 200 nodes). With the worldwide country air transportation network as a starting point, there are several interesting topics for future research, among which the relationship between the network properties of a country and its socio-economic factors should be analysed further.

References

[1] G. Bagler, "Analysis of the airport network of india as a complex weighted network," *Physica A: Statistical Mechanics and its Applications*, vol. 387, no. 12, pp. 2972–2980, 2008.

[2] B. F. Santos, S. Limbourg, and J. S. Carreira, "The impact of transport policies on railroad intermodal freight competitiveness the case of belgium," *Transportation Research Part D: Transport and Environment*, vol. 34, no. 0, pp. 230 – 244, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361920914001564

[3] Woolley-Meza *et al.*, "Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements," *The European Physical Journal B*, vol. 84, no. 4, pp. 589–600, 2011.

[4] A.-L. Barabasi, "Network science," *Philosophical Transactions of the Royal Society A: Mathematical,Physical and Engineering Sciences*, vol. 371, 2013.

[5] M. Newman, *Networks-An Introduction*. Oxford University Press, 2010.

[6] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: Anomalous centrality, community structure, and cities global roles," *PNAS*, vol. 102, no. 22, pp. 7794–7799, 2005.

[7] X. Sun and S. Wandelt, "Network similarity analysis of air navigation route systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 70, no. 0, pp. 416–434, 2014.

[8] S. Paleari, R. Redondi, and P. Malighetti, "A comparative study of airport connectivity in china, europe and us: Which network provides the best service to passengers?" *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 2, pp. 198 – 210, 2010, selected Papers from the 12th {ATRS} Conference.

[9] P. Fleurquin, J. J. Ramasco, and V. M. Eguiluz, "Systemic delay propagation in the us airport network," *Scientific reports*, vol. 3, 2013.

[10] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, p. 1337, 2013.

[11] S. Wandelt, X. Sun, and X. Cao, "Computationally efficient attack design for robustness analysis of air transportation networks," *Transportmetrica A: Transport Science*, vol. 11, no. 10, pp. 939–966, 2015.

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

[13] X. Sun, S. Wandelt, and F. Linke, "Temporal evolution analysis of the european air transportation system: air navigation route network and airport network," *Transportmetrica B: Transport Dynamics*, 2014.

[14] S. Lehner, "Separate yet interdependent networks: The structure and function of european air transport," in *Complex Networks IV*. Springer, 2013, vol. 476, pp. 109–120.

[15] S. Lehner and V. Gollnick, "Function-structure interdependence of passenger air transportation: Application of a systemic approach," *networks*, vol. 11, p. 14, 2014.

[16] R. Manduca, "Domestic migration networks in the united states," Master's thesis, Department of Urban Studies and Planning, MIT, 2014.

[17] G. Petri, P. Expert, H. J. Jensen, and J. W. Polak, "Entangled communities and spatial synchronization lead to criticality in urban traffic," *Scientific reports*, vol. 3, 2013.

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

[19] B. Sautman and Y. Hairong, "Friends and interests: China's distinctive links with africa," *African Studies Review*, vol. 50, no. 03, pp. 75–114, 2007.

[20] C. Issawi *et al.*, *An economic history of the Middle East and North Africa*. Routledge, 2013.

[21] C. Rose, "Discourses on japan and china in africa: Mutual mis-alignment and the prospects for cooperation," *Japanese Studies*, vol. 32, no. 2, pp. 219–236, 2012.

[22] M. F. C. Gomes, Pastore, L. Rossi, D. Chao, I. Longini, M. E. Halloran, and A. Vespignani, "Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak," *PLoS Currents*, 2014.

[23] Z. Huang, X. Wu, A. J. Garcia, T. J. Fik, and A. J. Tatem, "An open-access modeled passenger flow matrix for the global air network in 2010," *PLoS ONE*, vol. 8, no. 5, p. e64317, 05 2013.