Physica A 401 (2014) 201–206

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Network analysis reveals patterns behind air safety events

Massimiliano Zanin *

Innaxis Foundation & Research Institute, José Ortega y Gasset 20, 28006, Madrid, Spain
Faculdade de Ciências e Tecnologia, Departamento de Engenharia Electrotécnica, Universidade Nova de Lisboa, Lisboa, Portugal

HIGHLIGHTS

- A complex network representation of air traffic is proposed.
- Aircraft are mapped into nodes, pairwise connected according to their distance.
- Resulting topologies are able to forecast the appearance of unsafe situations.

ARTICLE INFO

Article history:
Received 10 October 2013
Available online 24 January 2014

Keywords:
Complex networks
Air transport
Data mining

ABSTRACT

Complex networks have been extensively used to study the topological and dynamical characteristics of transportation systems, although far less attention has been devoted to the analysis of specific problems arising in everyday operations. In this work, the use of a network representation is proposed for studying the appearance of Loss of Separation events, a kind of safety occurrence in which two aircraft violate the minimal separation while airborne. The topological analysis of networks representing the structure of traffic flows allows identifying situations in which the probability of appearance of such events is increased. Beyond these specific results, this work demonstrates the usefulness of the complex network approach in the analysis of operational patterns and occurrences.

1. Introduction

Complex networks [1–4] have extensively been used in the last decade as an instrument for modeling and analyzing complex systems, *i.e.* those systems composed of a large number of elements interacting in a non-linear way [5]. Among the large set of examples that can be found in the Literature [6], transportation systems have been the focus of numerous researches, the air transport being no exception—see, for instance, Refs. [7–9]. These studies can be categorized in two families, according to their main scope [10]: the characterization of the topology arising from direct connections between pairs of airports on one side, and the study of simple dynamical processes on top of the connection network, *e.g.* the appearance and propagation of congested states, on the other. A third family that has largely been neglected is the study of operational problems, that is, the use of complex networks as an instrument for the analysis of adverse situations that arise in the daily operation of the system. Notice that, on the contrary, this has extensively been done in other fields: in biomedicine, for instance, beyond the analysis of the topology created by interactions between genes, complex networks have been used to tackle specific problems, as detecting those genes responsible for different diseases [11,12].

In this work, complex networks are used to describe the status of the airspace when two aircraft are expected to cross close to each other, a situation known as *Loss of Separation* (LoS). Such events are of major importance in Air Traffic Management, as they may lead to serious accidents (*mid-air collisions*) when not correctly managed. In normal conditions,

* Correspondence to: Innaxis Foundation & Research Institute, José Ortega y Gasset 20, 28006, Madrid, Spain. Tel.: +34 606 993761.
E-mail addresses: massimiliano.zanin@ctb.upm.es, mzanin@innaxis.org.

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2014.01.032

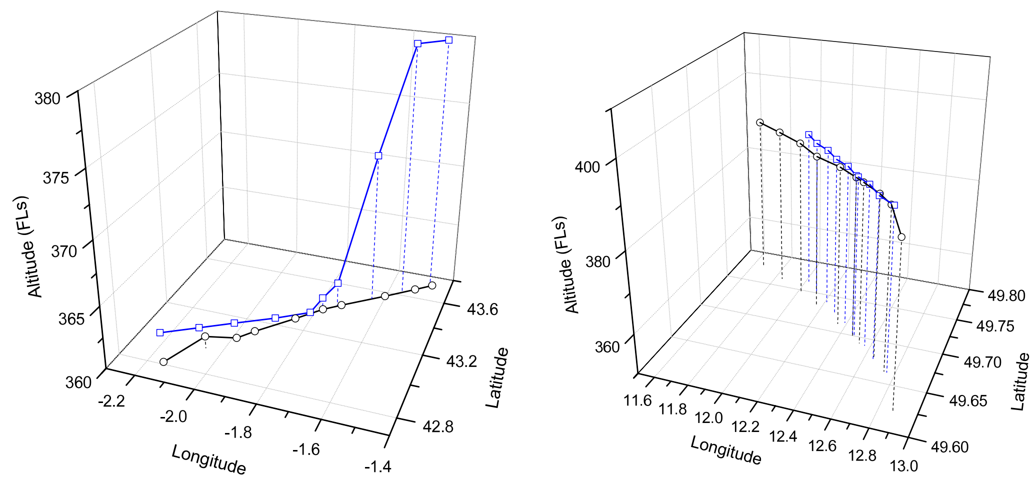

Fig. 1. *Example of safe and unsafe events:* left graph depicts the trajectories of two aircraft that were going to reach a *Loss of Separation* condition, but in which the intervention of the controller solved the conflict; notice how, between the third and four point (from left to right), the speed of the black aircraft was reduced, as indicated by the smaller length of the corresponding segment. Right graph depicts a similar situation where the LoS is not avoided. Both graphs have been constructed using radar recording of the 1st of March 2011. Longitudes and latitudes are expressed in degrees, while the altitude in *Flight Levels* (hundreds of feets).

two different barriers are acting to prevent such events [13]: first, the intervention of Air Traffic Controllers, whose mission is to guide the aircraft through optimal and safe trajectories, and second, different automatic tools, *e.g.* the *Traffic Collision Avoidance System*, which alert both the pilots and the controller of a potentially unsafe situation. In spite of this, and due to the high congestion of some airspaces, LoSs may still happen.

Two types of events are here considered: (i) those events that may have evolved in an incident, but were avoided by the intervention of pilots and controllers, and thus resulted in a *safe* situation; and (ii) those that actually ended up in a separation loss, *i.e.* in a (potentially) *unsafe* situation. Example of both types of events are represented in Fig. 1. In both cases, the two aircraft are expected to reach a close position in space at a given time. In Fig. 1 left, the black aircraft is slowed down between the third and fourth point of its trajectory, resulting in a safe distance with respect to the blue aircraft; on the other hand, Fig. 1 right reveals how the actuation of the controller is not enough to prevent an insufficient separation between the two vehicles. Using information about trajectories of all aircraft crossing the European airspace between March and December 2011, 100.032 events have been detected, 4.316 of which have been classified as *unsafe*—see Sections 2.1 and 2.2 for additional details.

The analysis of each one of these 100.032 events is performed by creating a network representation of the traffic in the region of interest. Trajectories of involved flights are *rewinded* 2 min back in time, in order to analyze the characteristics of the airspace that have led to the appearance of the event, and aircraft positions are mapped into a complex network following the method proposed in Fig. 2. Specifically, a node is associated to an aircraft when the latter lays within an *outer radius* centered in the safety event (blue dashed circle of Fig. 2); pairs of nodes are then connected whether their respective distance is lower than an *inner radius* (gray dashed circles of Fig. 2). Notice that such representation mimics the way Air Traffic Controllers manage aircraft: all flights within a given airspace are monitored, but special attention is devoted to those pairs whose distance falls below a given threshold [14].

In what follows one hypothesis will be tested, *i.e.* that an analysis of the topology of these networks may unveil information about the mechanisms behind the appearance of unsafe events. This is achieved by classifying events through standard data mining algorithms, using the topological metrics extracted from the networks as input parameters: a significantly high classification score implies that such network representations encode information able to forecast the appearance of unsafe situations. Beyond this specific aim, the work here presented aims at highlighting how complex networks can be used for the study of operational problems, a promising field that has not yet been tackled by means of statistical physics techniques.

2. Data preparation

2.1. Data set description

Trajectories data have been extracted from the *ALL_FT+* data set, collected by the EUROCONTROL PRISME group. It includes information about planned and executed trajectories for all flights crossing the European airspace, with a mean average resolution of 2 min. The data set covers the period from 1st March to the 31st December 2011, including a total of 10.3 million flights.

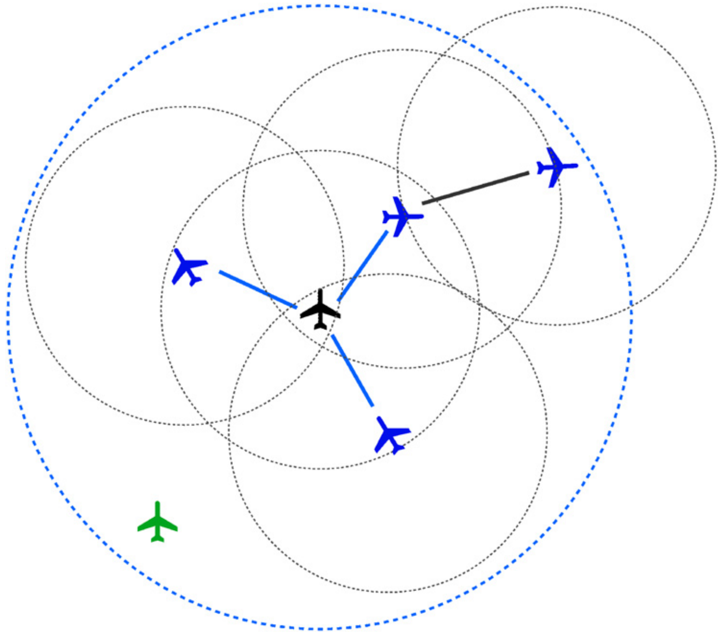

Fig. 2. *Network reconstruction procedure:* the proposed network reconstruction method starts by identifying the LoS event being studied, here represented by the gray aircraft in the center of the figure. Next, those aircraft that are within a given radius (called *outer radius*, blue dashed circle) are mapped into nodes of the network. Finally, pairs of nodes are connected if their corresponding aircraft are within a given distance from each other (*inner radius*, gray dashed circles). Notice that the green aircraft is within the outer radius, and therefore a node is associated to it; nevertheless, as it is far away from any other aircraft, that node remains disconnected. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Event detection algorithm

Potential LoSs were identified by projecting the intentions of each aircraft, as declared before the departure of the flight, starting from the actual position recorded by radars, and by detecting if two aircraft may break the established separation minima in the near future. Notice that this is equivalent to the surveillance task performed by any controller: given the actual position of aircraft, forecast their future intentions and detect potential unsafe situations [14].

Following this definition, these events include both situations that resulted in a safety-related condition, *i.e.* a reduction of the separation between aircraft, and situations that might have resulted in similar conditions, but in which the intervention of the controllers (or of the pilots) solved the problem before its appearance. By analyzing the real evolution of both flights, all events have been classified into these two groups, respectively *safe* and *unsafe*.

Due to the characteristics of the considered data set, *e.g.* the presence of errors and incongruences, an additional filter has been applied, in order to eliminate those events that are reflecting unreal situations. Thus, flights and events fulfilling one of the following conditions have been discarded:

- flights whose radar and planned trajectories are exactly the same, indicating that real radar information was missing or was artificially created;
- events for which the final real separation is lower than 20 s. This is usually due to a low spatial or temporal resolution in the radar information available;
- flights whose real trajectories include physically impossible segments, like supersonic velocities.

A total of 100.032 events passed this selection, 4.316 of which represent unsafe conditions.

3. Calculation and results

At the beginning of the network reconstruction process, two parameters have to be manually set, *i.e.* the *outer* and *inner* radii, respectively representing the size of the airspace monitored by the controller, and the distance at which two aircraft start to be an attention spot. The determination of these two parameters by means of operational considerations is not straightforward as they depend on several factors, like the procedures enforced in the considered airspace, or the presence of ascending/descending flights near airports. The definition of an inner radius is even more complex, as it depends on the specific controller managing the operation, his/her experience, *et caetera*.

Table 1
Optimal parameter selection. Classification score obtained for different combinations of the outer and inner radii—see Section 3 for details. For clarity's sake, results are coded in different colors: green, above 0.57; orange, between 0.56 and 0.57; and red below 0.56. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Top: inner radius Left: outer radius	6 NM	9 NM	12 NM	15 NM	18 NM
15 NM	0.5720	0.5648	0.5728	0.5645	—
30 NM	0.5615	0.5775	0.5760	0.5693	0.5613
45 NM	0.5608	0.5735	0.5710	0.5480	0.5510
60 NM	0.5555	0.5590	0.5575	0.5550	0.5425

Table 2
Statistical analysis of topological metrics: columns report the average and standard deviation of each metric, as detected in safe and unsafe events, and the p-value of a two-tailed t-test.

Topological metric	Average (safe)	Average (unsafe)	Std. deviation (safe)	Std. deviation (unsafe)	T-Test p-value
Number of nodes	16.97	14.58	14.18	11.75	0.0
Link density	0.071	0.091	0.058	0.071	0.0
Maximum degree	3.114	2.924	3.061	2.813	0.0
Entropy of the degree distribution	1.033	0.999	0.734	0.644	0.003
Clustering coefficient	0.202	0.204	0.210	0.214	0.597
Efficiency	0.094	0.121	0.086	0.116	0.0
Small-worldness	1.628	1.687	4.083	4.211	0.36
Number of connected components	8.578	7.654	3.730	3.556	0.0
Size of the giant component	5.607	4.890	7.042	5.708	0.0

In order to define the best pair of outer and inner radii, the methodology proposed in Ref. [15] has been applied, which is based on the idea that the score obtained in a classification task can be used as a proxy of the relevance of a network representation. The following steps are considered:

1. given a pair of values for the outer and inner radii, 100.032 networks are created, each one corresponding to a safety event. Values range between 15 and 60 NM for the outer and between 6 and 18 NM for the inner radius—see Table 1 for all combinations;
2. from each network, a set of nine topological metrics have been extracted, covering the most important aspects of its micro- and macro-scale; they include the number of nodes, link density, maximum degree, entropy of the degree distribution [16], clustering coefficient, efficiency [17], small-worldness [18], the number of connected components and the size of the giant component;
3. all possible pairs of topological metrics have then been used to feed a classification algorithm, whose aim was to correctly discriminate the unsafe from safe events. The algorithm used in the classification was a Support Vector Machine [19,20], due to its simplicity and high performances in real applications. As the original data set contains more safe than unsafe events, the classification has been performed on synthetic data sets comprising the same number of events for both categories. Furthermore, and in order to obtain a better representation of the validity of the analysis, a leave-one-out cross-validation technique has been applied [21];
4. finally, the best score obtained for each pair of outer and inner radii has been recorded, and used as a proxy of the quantity of information codified in the networks.

The results of this optimization process are presented in Table 1. The best classification score (0.5775) is obtained for the combination of outer/inner radii of 30/9 NM; given an average air speed of 500 knots, this is equivalent to a time separation of 3.6 and 1 min, indicating that controllers mostly work on the short term.

The topological properties of the networks resulting from this optimal combination of parameters have been further analyzed. With the aim of understanding which network metrics can better describe the differences between (and forecast the appearance of) safe and unsafe events, a statistical analysis has been performed on each one of them, in order to detect significant variations between both groups. Table 2 reports, along with the average and the standard deviation of each metric, the p-value of a *two-tailed t-test* [22]; high values of the p-value (*e.g.* higher than 0.05) indicate that the two distributions (corresponding to safe and unsafe events) have the same mean, and that therefore they are not likely to be used to discriminate both groups of events.

Among the seven topological metrics that have passed the *t-test*, four of them, *i.e.* the number of nodes, the efficiency, the number of connected components and the size of the giant component, have been further analyzed. Fig. 3 reports the two probability distributions of each one of them, corresponding to safe (green bars) and unsafe (red bars) events. By comparing

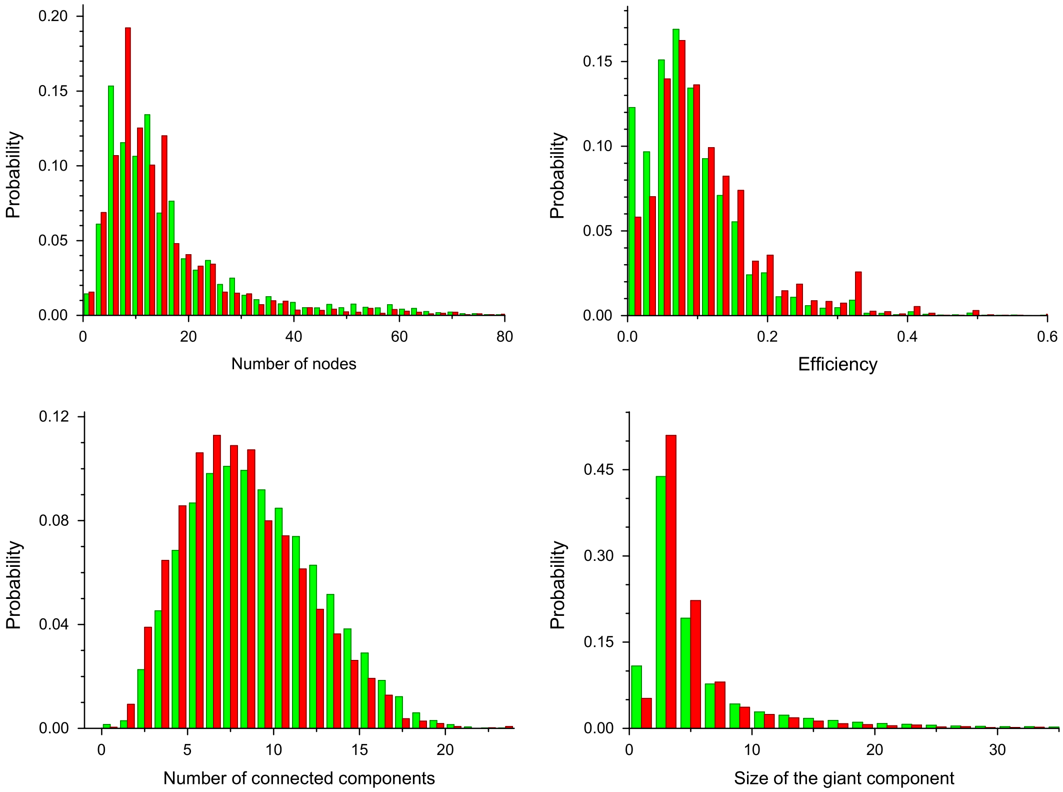

Fig. 3. *Topological metrics histograms:* histograms corresponding to the four topological metrics studied: the number of nodes, efficiency, the number of connected components, and the size of the giant component. Green (red) bars correspond to the value obtained for safe (unsafe) events. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the height of the two bars for a given metric value, it is possible to derive which kind of event is more probable, thus yielding insights about the conditions that increase the probability of appearance of unsafe situations. According to Fig. 3, unsafe events are more frequent in the middle range of number of nodes, *i.e.* of aircraft. This confirms classical results obtained in the Air Traffic Management: when the traffic level is high, the controller is aware of the risk, and adopts special techniques to avoid LoSs, like for instance holding circuits; on the other hand, when the traffic significantly reduces, his/her attention is likely to drop [14]. Furthermore, unsafe events are associated to network showing a lower number of connected components, of greater size, and with a higher internal cohesion (as indicated by the higher efficiency). From an operational point of view, this indicates situations in which the controller has to monitor groups of aircraft flying close together: his/her attention has then to split among different attention spots, with a consequently reduction in the situation awareness, and finally in traffic management efficiency. On the other hand, situations likely to evolve into a safe condition are characterized by evenly distributed flights across the airspace, thus forming a higher number of small clusters.

4. Conclusions

In conclusion, this work proposes a method for creating complex networks representing the interactions between aircraft crossing an airspace. Each network represents the status of the airspace *prior* to the appearance of a Loss of Separation event, with nodes being associated to flights, and pairs of them being connected when the distance between the corresponding aircraft is below a threshold.

The analysis of the resulting networks allows to recover known results, as for instance the importance of the level of traffic to be managed by a controller. In addition, such approach unveils aspects of the problem that were previously unknown, or that at least were not mathematically formalized, as the adverse effect created by multiple compact groups of aircraft, which split the attention of the controller among multiple targets. Such insights may be used in the future for the construction of an improved warning system, able to detect unsafe situations before their appearance.

Beyond the specific results here reported, this work represents an example of the use of complex network for modeling and analyzing operational problems in air transport. By highlighting the characteristics of patterns created by interacting elements, complex network theory may be used in the future to tackle other problems in which such interactions are essential drivers. Some of them include the analysis of other safety events, like *runway incursions* (where one vehicle enters the runway without authorization [23]) or *level bust* (when an aircraft fails to reach the assigned flight altitude); but also global phenomena, like the characterization of the resilience of the system after the appearance of errors or equipment failures.

Acknowledgments

This work is co-financed by EUROCONTROL acting on behalf of the SESAR Joint Undertaking (the SJU) and the EUROPEAN UNION as part of Work Package E in the SESAR Programme. Opinions expressed in this work reflect the authors' views only and EUROCONTROL and/or the SJU shall not be considered liable for them or for any use that may be made of the information contained herein.

References

[1] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47.
[2] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.
[3] L.A. Amaral, J.M. Ottino, Complex networks, Eur. Phys. J. B 38 (2004) 147–162.
[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[5] P.W. Anderson, et al., More is different, Science 177 (1972) 393–396.
[6] L.d.F. Costa, O.N. Oliveira Jr., G. Travieso, F.A. Rodrigues, P.R. Villas Boas, L. Antiqueira, M.P. Viana, L.E. Correa Rocha, Analyzing and modeling real-world phenomena with complex networks: a survey of applications, Adv. Phys. 60 (2011) 329–412.
[7] A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, Proc. Natl. Acad. Sci. USA 101 (2004) 3747–3752.
[8] R. Guimera, S. Mossa, A. Turtschi, L.N. Amaral, The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles, Proc. Natl. Acad. Sci. 102 (2005) 7794–7799.
[9] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, The role of the airline transportation network in the prediction and predictability of global epidemics, Proc. Natl. Acad. Sci. USA 103 (2006) 2015–2020.
[10] M. Zanin, F. Lillo, Modelling the air transport with complex networks: a short review, Eur. Phys. J. Spec. Top. 215 (2013) 5–21.
[11] A.-L. Barabási, N. Gulbahce, J. Loscalzo, Network medicine: a network-based approach to human disease, Nature Rev. Genet. 12 (2011) 56–68.
[12] M. Zanin, S. Boccaletti, Complex networks analysis of obstructive nephropathy data, Chaos 21 (2011) 033103.
[13] M.S. Nolan, Fundamentals of Air Traffic Control, Cengage Learning, 2010.
[14] H. Blom, S. Stroeve, M. Everdij, M. Van der Park, Human cognition performance model to evaluate safe spacing in air traffic, Human Fact. Aerosp. Saf. 3 (2003) 59–82.
[15] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. García-Prieto, F. del Pozo, E. Menasalvas, S. Boccaletti, Optimizing functional network representation of multivariate time series, Sci. Rep. 2 (2012).
[16] B. Wang, H. Tang, C. Guo, Z. Xiu, Entropy optimization of scale-free networks robustness to random failures, Phys. Stat. Mech. Appl. 363 (2006) 591–596.
[17] V. Latora, M. Marchiori, Efficient behavior of small-world networks, Phys. Rev. Lett. 87 (2001) 198701.
[18] M.D. Humphries, K. Gurney, Network small-world-ness: a quantitative method for determining canonical network equivalence, PLoS One 3 (2008) e0002051.
[19] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[20] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge university press, 2000.
[21] V.N. Vapnik, Statistical Learning Theory, Wiley, 1998.
[22] J.E. Freund, G.A. Simon, Modern Elementary Statistics, vol. 12, Prentice-Hall Englewood Cliffs, New Jersey, 1967.
[23] G. Singh, C. Meier, Preventing runway incursions and conflicts, Aerosp. Sci. Technol. 8 (2004) 653–670.