



# Topological data analysis for aviation applications

Max Z. Li<sup>a,\*</sup>, Megan S. Ryerson<sup>b,c</sup>, Hamsa Balakrishnan<sup>a</sup>

<sup>a</sup> Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>b</sup> Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA, USA

<sup>c</sup> Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA



## ARTICLE INFO

### Keywords:

Aviation data  
Topological data analysis  
Persistent homology  
Airline networks

## ABSTRACT

Aviation data sets are increasingly high-dimensional and sparse. Consequently, the underlying features and interactions are not easily uncovered by traditional data analysis methods. Recent advancements in applied mathematics introduce *topological methods*, offering a new approach to obtain these features. This paper applies the fundamental notions underlying topological data analysis and persistent homology (TDA/PH) to aviation data analytics. We review past aviation research that leverage topological methods, and present a new computational case study exploring the topology of airport surface connectivity. In each case, we connect abstract topological features with real-world processes in aviation, and highlight potential operational and managerial insights.

## 1. Introduction

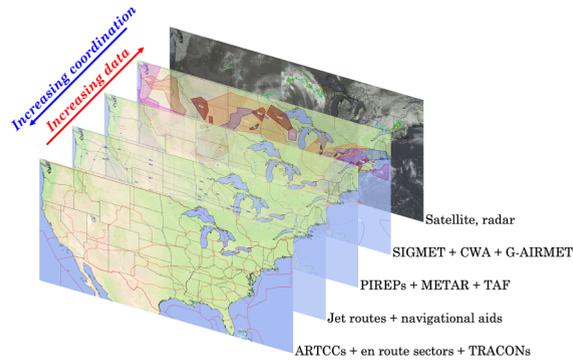
The sustained growth of air traffic worldwide has been accompanied by an influx of aviation data (Li and Ryerson, 2019). The 1300 commercial airlines and their nearly 32,000 in-service aircraft operated 41.9 million flights between more than 3700 airports in 2017, transporting 4.1 billion passengers across 45,000 routes (Air Transport Action Group, 2018). Each of these components of the air transportation system is also a source of data. The interconnectedness of the system is reflected in its data: Embedded sensors onboard individual aircraft not only report data pertaining to that specific aircraft, but also keep track of its interactions with other assets in the air and on the surface. It is in this context that the “Flight 4.0” era of smart and connected technologies has emerged, coupling traditional aviation paradigms with concepts such as the Internet of Things (IoT) and cyber-physical systems (Durak, 2018). The application of these concepts to aviation has already produced promising results (Mott and Bullock, 2018; Chatterjee et al., 2017).

While the ever-growing quantity of data is, barring computational tractability, generally welcomed, the increasing complexity of the data poses a challenge (Cook et al., 2015). The complexity of aviation data reflects the interdependencies within the air transportation network. We provide a more in-depth review and discussion of big data in aviation in Section 1.1. Additionally, many aviation data sets are high-dimensional and sparse; graph-based models that only encode pairwise relationships do not have the flexibility to capture higher-order relational information. We review elements of complexity within the aviation system, as well as previous research investigating network-level modeling in aviation in Section 2.

Given the prevalence of such data sets in aviation that resist accurate representation and analysis via classical statistical, graph-theoretic, and other data analysis methods, we propose switching to a global, topologically-driven perspective to extract pertinent features. Recent advances in applied mathematics introduce *topological methods* – most prominently *topological data analysis* and

\* Corresponding author.

E-mail addresses: [maxli@mit.edu](mailto:maxli@mit.edu) (M.Z. Li), [mryerson@design.upenn.edu](mailto:mryerson@design.upenn.edu) (M.S. Ryerson), [hamsa@mit.edu](mailto:hamsa@mit.edu) (H. Balakrishnan).



**Fig. 1.** The architectural layers of the NAS with respect to convective weather; each layer provides its own set of data that are closely related to data from another layer.

*persistent homology* (TDA/PH) – that offer a compelling new direction from which these salient features may be obtained (Carlsson, 2009; Ghrist, 2014). TDA/PH uses results from algebraic topology to extract information such as clusters, holes, voids, and complexes from high-dimensional data sets; these topological methods have already found a diverse range of data science applications, a sampling of which we review in Section 2.

### 1.1. Aviation and big data

We briefly discuss some key characteristics of aeronautics-specific big data to provide context (Burmester et al., 2018). The first aspect is *volume*, characterized by the increasing quantities of interdependent data, rendering traditional data analysis methods inadequate in some settings. A single day of air traffic operations within the US National Airspace System (NAS) generates (1)  $O(10^4)$  reports pertaining to arrival and departure information, boundary crossing updates, oceanic reports, planned positions, and actual positions; (2)  $O(10^5)$  records of flight management information, flight plan-specific data, aircraft specifications, and flight route data; and (3) more than  $O(10^6)$  observations related to flight tracks, airways, centers, fixes, sectors, and waypoints (Comitz et al., 2013). The interdependencies within and interconnectedness of data elements also present scenarios in which topological methods may be useful. Furthermore, various inconsistencies, errors, and sparsities that are present in the data result in high levels of variability (Burmester et al., 2018). Due to the inherent ability of TDA/PH to search for global topological invariants within a data set, the methods we discuss in Section 4 are robust against noise within aviation data sets. In a sense, the overall shape of the data should not change in a dramatic way if slightly perturbed, either by introducing noise or by inducing sparsities (Ghrist, 2014).

Fig. 1 illustrates that even in the limited example of convective impacts on the NAS, the complexity of the system is evident. The forward layers in Fig. 1 that describe various NAS airspace partitions are highly organized and structured, while the last layer containing raw meteorological data and weather radar returns is comparatively less structured, but with higher data dimensions and quantities.

Taking the layers described in Fig. 1 into account with other features within the aviation network (airports as nodes, air routes as edges, origin-destination delays, ground stops, Ground Delay Programs, Airspace Flow Programs, etc.), a highly complex and high-dimensional network begins to surface. It is often important to keep track of many tuples of relationships between airports, airlines, and other aviation stakeholders, not just pairwise relations. A graph-theoretic analysis of this network is limited to dyadic relationships. By contrast, tools based on TDA/PH enable an analysis of the overall topology of NAS data sets. We will motivate the case for introducing TDA/PH as a complementary data analysis tool in aviation data analytics in the succeeding sections.

### 1.2. Manuscript outline

We provide an encompassing literature review of network-level models and complexity science in aviation, as well as successful applications of TDA/PH, in Section 2. We highlight specific contributions of our work in Section 3, then present a mathematically rigorous but concrete introduction to TDA/PH and its algebraic-topological foundations in Section 4. To motivate the applicability of TDA/PH in aviation, we conduct an in-depth review of three aviation research works with a topological flavor in Section 5, before presenting our novel computational case study where we apply TDA/PH to airport surface operations in Section 6. We holistically round out our work with a detailed discussion of potential operational and managerial insights derivable from using TDA/PH in aviation data analytics (Section 7). To encourage the continued exploration of TDA/PH in aviation, we provide an array of future research directions in Section 8. We summarize the main contributions of our work and our vision for TDA/PH in aviation data science in Section 9.

## 2. Literature review

In a comprehensive look at applying complexity science to air transportation, Cook et al. (2015) explores how high-dimensional

and sparse data sets, along with operational, equipment, and weather uncertainty lead to tremendous challenges in modeling, predicting, and improving the performance of the air transportation system. Examples of complexity factors throughout the air transportation system include number of aircraft, aircraft proximity measures, and density indicators (Djokic et al., 2010). The ability to extract insights and patterns from high-dimensional and complex data sets – both in aviation and other industries – has the potential to inform the creation of better models and predictive analytics (Govindan et al., 2018; Oracle, 2012; Ayasdi, 2015; FlightGlobal, 2019). Furthermore, the specific adoption of TDA/PH by several data science industries has already yielded novel insights in areas such as healthcare and retail consumer behavior (Ayasdi, 2015).

Previous research focuses on a broader view of the air transportation system, choosing to examine the entirety of a given air route network using network science (Zanin and Lillo, 2013). Through such a wide scope of analysis, insights such as the varying levels of dynamic complexity and how that variability plays into delay and fuel inefficiencies have been uncovered (Simić and Babić, 2015; Rocha, 2017). These insights help address another facet of airspace complexity: the impacts of different stakeholders – airport operators, airlines, air navigation service providers (Schaar and Sherry, 2010) – on the air transportation network as a whole. Each stakeholder has a different perspective of the system, resulting in a different valuation of system utility and performance efficiency (Kotegawa et al., 2014). While these insights are critical quantitative foundations upon which routing decisions and airspace utilization should be made at the stakeholder level, it may also be necessary to infer global features and characteristics regarding various aviation subsystems from local data.

More recently, in line with the increasing ubiquity of big data in aviation, aviation data science research has shifted towards dealing with high-dimensional and multifaceted data sets. New predictive models in aviation using methods such as deep belief networks (DBNs) harness such complex data sets to estimate flight delays (Yu et al., 2019); however, such methods cannot identify nor explicitly discern triadic and higher-order relationships, a concept that naturally arises in TDA/PH. Topological motifs analyzed in Du et al. (2018) offer a proxy of higher-order interactions between airport delays, but ultimately still relies on underlying pairwise Granger causality tests. Other previous works use analogues of hypergraphs via a multi-layered network representations in order to approximate higher-order relationships between airports (Belkoura et al., 2016; Du et al., 2016); however, each individual network is still restricted to node-edge airport pairs, and is not flexible enough to admit multi-airport interactions. TDA/PH utilize so-called *simplicial complex representations*, and is not constrained to pairwise relationships, as we will see in Section 4.

TDA/PH leverage tools from algebraic topology to provide topologically qualitative and global information about a data set using local information, while also naturally allowing for higher-order relationships. While this subject is slowly transitioning from pure mathematics to applications in engineering, the usage of topological structures such as networks and graphs in air transportation research is much more mature. In the scope of air route networks, graph-based approaches have been used to characterize the complexity of airspace sectors (Hongyong et al., 2015), to quantify the connectivity and robustness of the route network structure (Wei et al., 2014; Zhou et al., 2019), and to find correlations between stakeholder metrics and the network topology of the air route network (Kotegawa et al., 2014). In addition, Yousefi and Zadeh (2013) applied methods in fluid flow dynamics to characterize en route airspace corridors, resulting in a quantitative benefit assessment of new Area Navigation (RNAV)-enabled “Q” jet routes.

TDA/PH have already found a diverse range of applications in engineering, particularly in problems involving large, high-dimensional data sets with non-trivial degrees of connectivity and loops. These topological methods have opened new avenues in networked neuroscience (Giusti et al., 2016), as well as in path-planning for autonomous robots (Bhattacharya et al., 2013, 2015) and modeling coverage and exploration behaviors for teams of multiple robots (Bhattacharya et al., 2014). Topological methods have also been used to extract persistent features within fluid flow (Kasten et al., 2011), identify pertinent structures within LiDaR data sets (Keller et al., 2011), and characterize the complex connections and networks within the airways of the human lung (Szymczak, 2011).

### 3. Contribution of work

Our work in this paper embraces the fact that aviation, like many other fields, is entering the era of big data. Such an abundance of high-dimensional trajectory data lends itself well to augmentation routines and topologically-qualitative methods from the emerging toolbox of TDA/PH. We have identified the need for complementary data analytics tools that offer insights into global, topological features within complex and high-dimensional aviation data sets in Sections 1.1 and 2. To directly address this methodological gap, the first major contribution of our work is to provide a consistent, formal, and concrete introduction to TDA/PH within the aviation domain. This is accomplished through Section 4, where we use an aviation-based example to build up the algebraic-topological notions required for TDA/PH. Our second major contribution is bridging the connection between TDA/PH and potential actionable operational and managerial insights within the air transportation system. We accomplish this through providing an in-depth review of previous research work (Section 5), presenting a new computational airport surface case study (Section 6), discussing how TDA/PH yields insights directly relatable to physical processes and operations within the aviation domain (Section 7), and providing a plethora of future research directions in aviation data analytics that leverage TDA/PH (Section 8).

We emphasize that TDA/PH does not replace, but instead complement existing topological and graphical methods employed in aviation research. We surveyed a sampling of aviation research that employed network-based and graph-theoretic models in Section 2; the array of important and impactful results from Section 2 are based off of dyadic relationships. TDA/PH complement these approaches by capturing triadic and higher-order relationships (Ramanathan et al., 2011; Giusti et al., 2016; Sizemore et al., 2018; Xia, 2018). This indicates that one challenge regarding applying TDA/PH to aviation data sets is first identifying whether or not characterizing pairwise relationships are enough. For example, in studies of causal interactions between airport delays, it may be desirable to only examine pairwise airport interactions via Granger causality (Du et al., 2018); however, if the goal is instead to

capture hub-and-spoke dynamics with stronger forms of causality such as structural equation models (SEMs), then higher-order relationships may be necessary. Another challenge is the need for aviation-specific subject matter expertise in mapping topological results to real-world processes and phenomenon; we address this challenge directly in Section 7.

#### 4. From simplices to holes: fundamental concepts from algebraic topology

TDA/PH are built up from a collection of fundamental ideas, definitions, and concepts borrowed from *algebraic topology*. We introduce some fundamentals in this section, now that we are equipped with the contextual understanding of the types of aviation data sets that may be well-suited for TDA/PH. A high-level explanation of algebraic topology is that it is a rigorous way of comparing spaces, and keeping track of non-trivial topological features within these spaces. In other words, the goal is to find *holes* in spaces. Imagine two spaces where one is a disc, and another is a disc with a hole punctured in the middle of it. In the first space (the whole disc), if we envision moving a rubber band around in the space, there are no restrictions to how we move this rubber band. In other words, there is only one configuration for this rubber band. However, in the second space (the punctured disc), there are two configurations for this rubber band: When the rubber band surrounds the punctured hole, and when the rubber band does not. The rubber band cannot transition between the two configurations without breaking, due to the hole in the space of the punctured disc.

Algebraic topology provides a framework drawn from abstract algebra and topology for this natural intuition of holes within spaces. TDA/PH provide the bridge between data and the kinds of topological spaces that can be examined through concepts from algebraic topology. We build up to a rigorous definition of holes and higher-dimensional holes (Definition 7) in this section, culminating in an algebraic structure endowed with the ability to keep track of holes between spaces (homology groups; Definition 13) that formalizes our rubber band example. With an understanding of the various topological and algebraic foundations, the natural question then is how these abstract notions translate to aviation data science applications (Sections 5 and 6) and specifically how topological features such as holes map to physical processes and phenomena within the aviation domain (Section 7), translating to potential operational and managerial insights.

In this section, we present the background required to apply TDA/PH methods. Using an illustrative example, we develop the algebraic-topological construct known as a *simplicial complex*, and define various algebraic objects on the simplicial complex. Finally, we define *homology groups*, whose sizes inform us whether or not there are holes in our dataset. For a deeper discussion of algebraic topology and TDA/PH, we recommend other references (Hatcher, 2002; Edelsbrunner and Harer, 2009; Ghrist, 2014).

##### 4.1. Simplicial complexes

We begin our exploration of TDA/PH by considering a particular way of structuring data that has its roots in algebraic topology. Table 1 shows a sample data set containing a collection of major airports within the US, as well as a list of associated features. The airports are Chicago O’Hare International Airport (ORD), Chicago Midway International Airport (MDW), New York-LaGuardia Airport (LGA), Newark Liberty International Airport (EWR), San Francisco International Airport (SFO), and Miami International Airport (MIA). The features need not share contextual commonalities: information regarding metroplexes are geographic in nature, hub characteristics are airline-dependent, and Ground Delay Programs (GDPs) are tactical features related to air traffic flow management (ATFM). It would be difficult to capture the relationships and structure in this data with a graph that models only pairwise relations.

Using the vertex labels for each airport given by the leftmost column in Table 1, we define the set of vertices  $\mathcal{V} := \{v_0, v_1, v_2, v_3, v_4, v_5\}$  along with a superset  $\mathcal{K}$  containing sets of vertex relations:

**Table 1**

Example data set containing airports along with the associated features. UAL and AAL denote United Airlines and American Airlines; Departure delays, Miles-In-Trail (MIT) and Minutes-In-Trail (MINIT), Flow Constrained Areas (FCA), routing flow constraints (e.g., wind routes), and Ground Delay Programs (GDPs) are all traffic management initiative (TMI) features.

Vertex	Airport	Features				
$v_0$	ORD	Chicago metroplex	UAL hub	AAL hub	FCA TMI	Wind route TMI
$v_1$	MDW	Chicago metroplex	Secondary airport			
$v_2$	LGA	New York metroplex	Secondary airport			
$v_3$	EWR	New York metroplex	UAL hub	Departure delays	MINIT TMI	Wind route TMI MIT TMI
$v_4$	SFO	Ongoing GDP	UAL hub	Departure delays	FCA TMI	MIT TMI
$v_5$	MIA	Ongoing GDP	AAL hub	Departure delays	MINIT TMI	

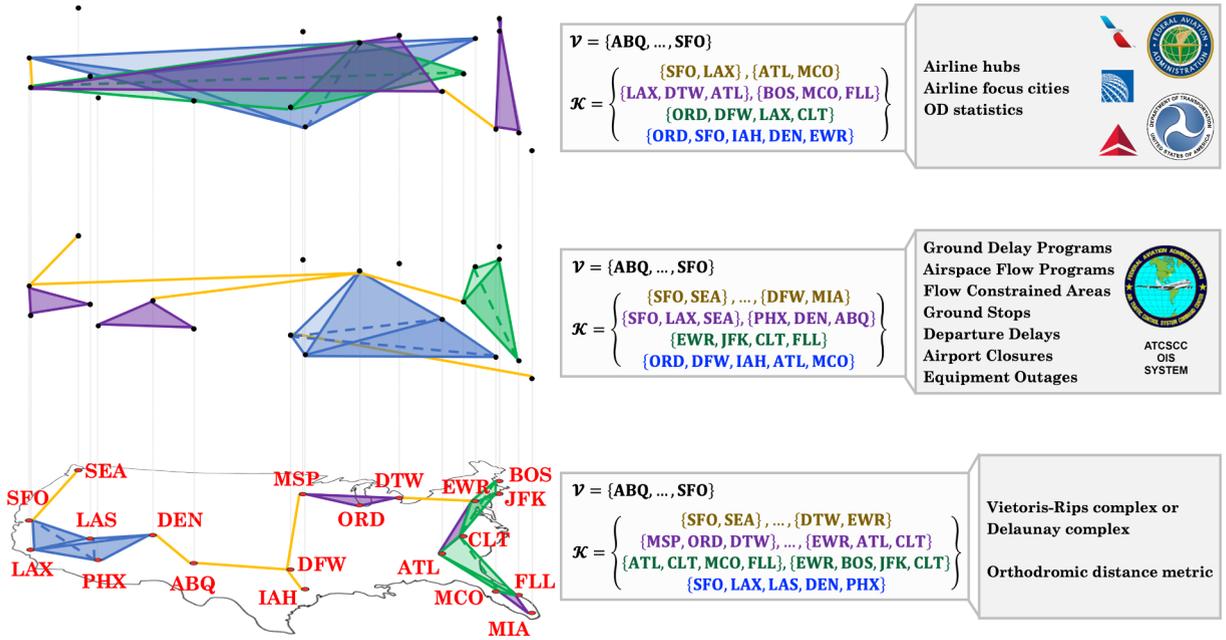


Fig. 2. Example of multi-layered simplicial complexes corresponding to different types of aviation data.

$$\mathcal{K} := \{ \text{sets of vertices } \in \mathcal{V} \text{ with same features} \}$$

$$= \left\{ \begin{array}{l} \{v_0\}, \{v_1\}, \{v_2\}, \{v_3\}, \{v_4\}, \{v_5\} \\ \{v_0, v_1\}, \{v_2, v_3\}, \{v_1, v_2\}, \{v_4, v_5\}, \{v_0, v_5\}, \{v_3, v_4\}, \{v_3, v_5\}, \{v_0, v_4\}, \{v_0, v_3\} \\ \{v_0, v_3, v_4\}, \{v_3, v_4, v_5\} \end{array} \right\}. \tag{1}$$

The elements of  $\mathcal{K}$  are interpreted using the shared features among the vertices, as shown in Table 1. For example, an element in  $\mathcal{K}$  containing one vertex corresponds to the feature of the vertex  $v_i$  representing a certain airport (e.g.,  $\{v_0\} \in \mathcal{K}$  is associated with the feature of representing ORD), an element in  $\mathcal{K}$  containing two vertices indicate features shared by an airport pair (e.g.,  $\{v_0, v_1\} \in \mathcal{K}$  is associated with the feature that both ORD and MDW are in the Chicago metroplex), and so on. The algebraic structure provided by  $\{\mathcal{V}, \mathcal{K}\}$  is known as a *simplicial complex*. Formally, an (abstract) simplicial complex is defined in the following manner in terms of  $\mathcal{V}$  and  $\mathcal{K}$ :

**Definition 1.** An (abstract) simplicial complex, given a set of vertices  $\mathcal{V}$  and superset  $\mathcal{K}$  containing sets of vertex relations, is an algebraic structure such that if  $\sigma \in \mathcal{K}$  and  $\tau \subseteq \sigma \Rightarrow \tau \in \mathcal{K}$ .

Alternate ways of forming simplicial complexes from data include flag complexes and Vietoris-Rips complexes (Sizemore et al., 2018). We provide an illustrative example in Fig. 2 of additional simplicial complexes that could be formed from different types of aviation domain data.

We can decompose our simplicial complex into smaller *simplices*, and begin visualizing it from a geometric perspective. To this end, we consider a geometrically-inspired definition of a *k-simplex* (Hatcher, 2002):

**Definition 2.** A *k-simplex*, denoted as  $\Delta^k$ , is the convex hull of  $k + 1$  affine (non-collinear) points. The *standard k-simplex* can be defined explicitly as:

$$\Delta^k = \{ (t_0, \dots, t_k) \in \mathbb{R}^{k+1} \mid \sum_i t_i = 1, \quad t_i \geq 0, \forall i \}. \tag{2}$$

We provide an illustration of the standard 2-simplex  $\Delta^2$  in Fig. 3. In order to begin visualizing the simplicial complex given by the data in Table 1, we need to discuss the convention regarding *standard orientations* on the *k-simplices*. Fig. 4 provides an illustration of *k-simplices* for  $k = 0, 1, 2$ , and 3. The edges connecting each vertex in the *k-simplices* for  $k \geq 1$  are oriented with respect to the vertex indices. For example, given a 1-simplex  $\Delta^1$  with  $\{v_i, v_j\} \in \mathcal{K}$  and  $i < j$ , the standard orientation points from  $v_i$  to  $v_j$ . This extends to the labelings on the 2- and 3-simplex as well in Fig. 4. We can formalize the notion of a standard orientation on a *k-simplex* by the following definition:

**Definition 3.** Suppose we have a *k-simplex*  $\sigma_{0, \dots, k}$  with the indices written in a strictly-ascending order, i.e.  $0 < \dots < k$ . This ordering is the *standard orientation* on a *k-simplex*, and can be visualized via the decomposition  $0 < 1, 1 < 2, \dots, k - 1 < k$  consisting of 1-simplices  $\sigma_{0,1}, \sigma_{1,2}, \dots, \sigma_{k-1,k}$  where the oriented edge for  $\sigma_{i,j}$  exits the 0-simplex  $v_i$  and enters the 0-simplex  $v_j$ , given that  $j = i + 1$  and

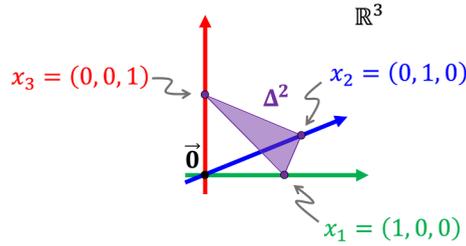


Fig. 3. The standard 2-simplex, denoted as  $\Delta^2$ . Note that  $(x_1, x_2, x_3)$  spans  $\mathbb{R}^3$ .

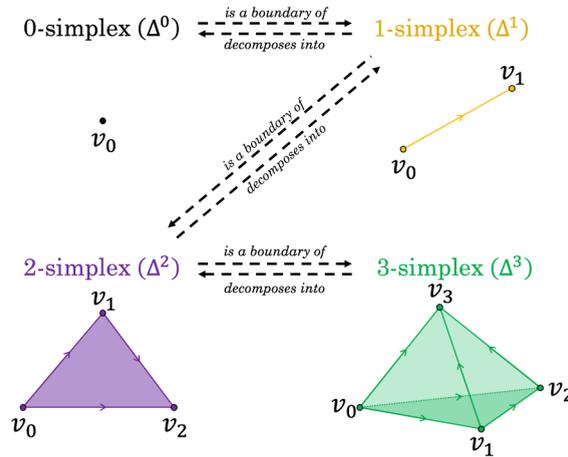


Fig. 4. Illustration of a 0-simplex, 1-simplex, 2-simplex, and 3-simplex. The edges are marked according to standard orientation conventions.

$(i, j) \in \{0, \dots, k\} \times \{0, \dots, k\}$ .

Another intuitive relationship between a  $k$ - and a  $(k \pm 1)$ -simplex for  $k \geq 1$  is the notion of *boundary decompositions*. Referring back to Fig. 4,  $k$ -simplices for  $k \geq 1$  compose the boundary of the  $(k + 1)$ -simplex, and can be decomposed into its own boundary formed from  $(k - 1)$ -simplices. This is a notion that we will formalize and exploit in our search for holes in our data.

We map our algebraic simplicial complex  $\{\mathcal{V}, \mathcal{K}\}$  to a topological simplicial complex. Let  $X_{NAS}$ , referred to as the *NAS simplicial complex*, denote the *topological space* that is the simplicial complex created from the data in Table 1. We provide an illustration of  $X_{NAS}$  in Fig. 5. An issue of notation must also be addressed here as well, since we can view  $k$ -simplices and simplicial complexes as algebraic or topological objects. A  $k$ -simplex can be denoted as  $\Delta^k$  or  $\sigma_{0, \dots, k}$  with the usual ordering  $0 < \dots < k$ ; the former carries a topological interpretation whereas the latter carries an algebraic one. For the purposes of introducing algebraic topology and

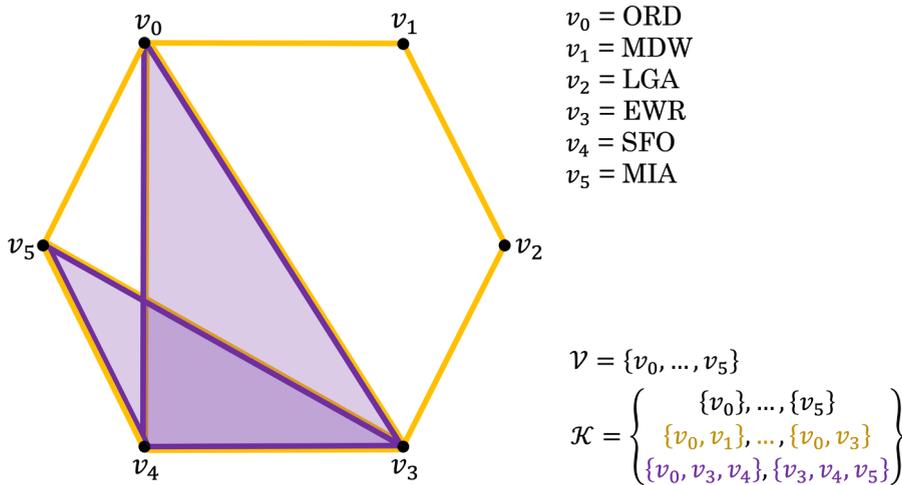


Fig. 5. Our simplicial complex  $X_{NAS}$  visualized using 0-, 1-, and 2-simplices. It is the topological counterpart of our algebraic structure  $\{\mathcal{V}, \mathcal{K}\}$  derived from the NAS data in Table 1.

homology to understand the basics of TDA/PH, we will use both notations depending on the context.

In the case of our simplicial complex  $X_{NAS}$  in Fig. 5, each black 0-simplex  $\sigma_i$  corresponds to the element  $\{v_i\} \in \mathcal{K}$ ; each orange 1-simplex  $\sigma_{i,j}$ ,  $i < j$  (recall standard orientation conventions) corresponds to the element  $\{v_i, v_j\} \in \mathcal{K}$ ; and each purple, filled-in 2-simplex  $\sigma_{i,j,k}$  with  $i < j < k$  corresponds to the element  $\{v_i, v_j, v_k\} \in \mathcal{K}$ . For example, the purple triangle denoting the 2-simplex  $\sigma_{0,3,4}$  formed by  $\{v_0, v_3, v_4\} \in \mathcal{K}$  encodes the shared feature amongst ORD, EWR, and SFO of being UAL hubs. Furthermore, we can define a substructure on  $X_{NAS}$  called a  $k$ -skeleton, typically denoted as  $X^k$ . A  $k$ -skeleton is the collection of all  $k$ -simplices within a larger simplicial complex.

#### 4.2. Algebra with simplicial complexes: chains and chain groups

We will now introduce a linear-algebraic abstraction that allows us to perform algebra on  $X_{NAS}$  with surprisingly intuitive visualizations. First, we formalize the idea of *chains* within simplicial complexes such as  $X_{NAS}$  with the definition of *chain groups*:

**Definition 4.** Given a simplicial complex  $\{\mathcal{V}, \mathcal{K}\}$ , the 1<sup>st</sup> chain group  $C_1$  is the group:

$$C_1 = \left\{ \sum a_n \sigma_{i,j} \mid a_n \in \mathbb{Z}, \sigma_{i,j} \in X^1 \text{ derived from } \{\mathcal{V}, \mathcal{K}\} \right\}. \tag{3}$$

Elements of  $C_1$  are known as *1-chains*. Note that from the definition of the 1<sup>st</sup> chain group  $C_1$ , 1-chains are simply linear combinations of 1-simplices  $\sigma_{i,j}$  with scalar integer coefficients  $a_n \in \mathbb{Z}$ . The notion of chain groups can be extended to all dimensions naturally:

$$C_k = \left\{ \sum a_n \sigma_{0, \dots, k} \mid a_n \in \mathbb{Z}, \sigma_{0, \dots, k} \in X^k \text{ derived from } \{\mathcal{V}, \mathcal{K}\} \right\}. \tag{4}$$

Analogously, elements of  $C_k$  are known as *k-chains*. Note that the summation in the definitions for the 1<sup>st</sup> and  $k^{\text{th}}$  chain group is formally the operator  $+_{C_k}: C_k \times C_k \rightarrow C_k$  that sums together two  $k$ -simplices. For example, for the 1<sup>st</sup> chain group, we have that  $\sigma_{i,j} + {}_{C_1} \sigma_{k,l}$  results in the new 1-simplex  $\sigma_{i,j} + \sigma_{k,l} \in C_1$ . We omit the subscript on the summation operator when the chain group we are working over is obvious. Furthermore, recalling the standard orientation on a  $k$ -simplex, any  $k$ -chains with negative coefficients correspond to the same  $k$ -chain but with reversed entry-exit orders.

While the definition for  $C_k$  may be reminiscent of a vector space spanned by  $k$ -simplices, it is actually a more generalized algebraic structure known as a *group*.

#### 4.3. Paths and holes

We can now visualize 1-chains in  $C_1$  in the context of the simplicial complex  $X_{NAS}$ . If we were to pick the 1-chain  $\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{3,5} - \sigma_{4,5} \in C_1$ , paying attention to the signs of the coefficients of each 1-simplex  $\sigma_{i,j}$ , we see that this is the 1-chain consisting of 1-simplices in the standard orientation connecting  $v_0, v_1, v_2, v_3$ , and  $v_5$  plus the 1-simplex connecting  $v_5$  to  $v_4$  with its orientation reversed by the negative coefficient on  $\sigma_{4,5}$ . Besides being a 1-chain, it is also a *path* consisting of 1-simplices, as the orientations on each 1-simplex agree consistently. Formally, we can define a path of  $k$ -simplices as:

**Definition 5.** A path consisting of  $k$ -simplices allows for a traversal with the same number of entrances and exits of each  $(k - 1)$ -simplex along the  $k$ -chain, with the exception of the initiating and terminating  $(k - 1)$ -simplexes. The initiating  $(k - 1)$ -simplex allows for only one exit, and the terminating  $(k - 1)$ -simplex allows for only one entrance.

Intuitively, we note a special type of path that colloquially-speaking does not have an end or a beginning. Formally, a *closed* path is defined as follows:

**Definition 6.** A closed path consisting of  $k$ -simplices allows for a traversal with the same number of entrances and exits of each  $(k - 1)$ -simplex along the  $k$ -chain.

This 1-chain path  $\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{3,5} - \sigma_{4,5} \in C_1$  can be visualized in Fig. 6 (left). While all paths consisting of 1-simplices are also 1-chains, not all 1-chains are paths. For example, pick two negative integers  $a_1$  and  $a_2$ , then  $a_1 \sigma_{0,5} + a_2 \sigma_{2,3}$  is a valid 1-chain in our NAS simplicial complex (Fig. 6 (right)), but *not* a valid path as the 1-simplices are disconnected.

The notion of paths can also be extended into higher dimensions using the same rules as in the case of 1-chains. In the NAS simplicial complex, there are two 2-simplices given by  $\sigma_{0,3,4}$  and  $\sigma_{3,4,5}$  representing UAL hub features shared by ORD, EWR and SFO, and departure delay features shared by EWR, SFO and MIA, respectively. The 2<sup>nd</sup> chain group  $C_2$  for the NAS simplicial complex is spanned by these two 2-simplices. The orientations do not match on the two 2-simplices – this can be visualized via Fig. 7. Thus,  $\sigma_{3,4,5} + \sigma_{0,3,4}$  is a valid 2-chain, but not a valid path consisting of 2-simplicies. Modifying the sign on, for example, the 2-simplex  $\sigma_{0,3,4}$ , gives us a different 2-chain that is a valid path (Fig. 8).

We now have the necessary background to formalize the process of finding topological holes within the simplicial complex representing our NAS data set. In the context of TDA/PH, we use the following definition to distinguish between holes and other features in our simplicial complex:

**Definition 7.** A  $k$ -dimensional hole (also known as a *void* or *cavity*) is a valid closed path composed of  $k$ -simplices within a simplicial complex. Furthermore, this valid closed path cannot be the *boundary* of some  $(k + 1)$ -simplex.

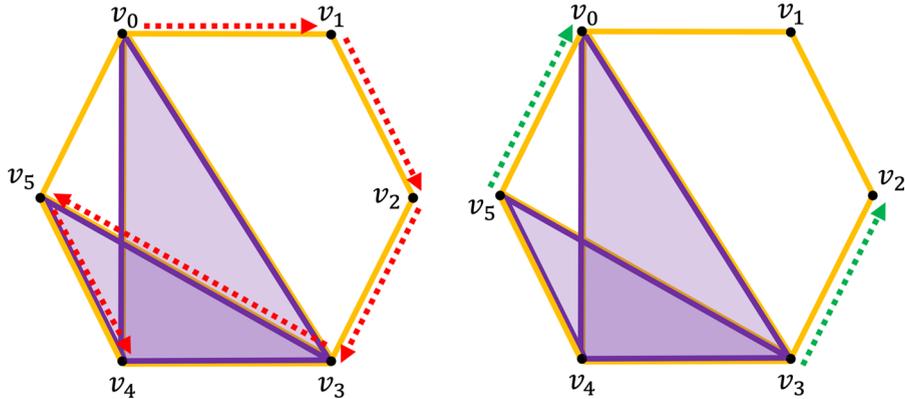


Fig. 6. Two 1-chains in the NAS simplicial complex, where the one on the left also forms a path, but the one on the right does not.

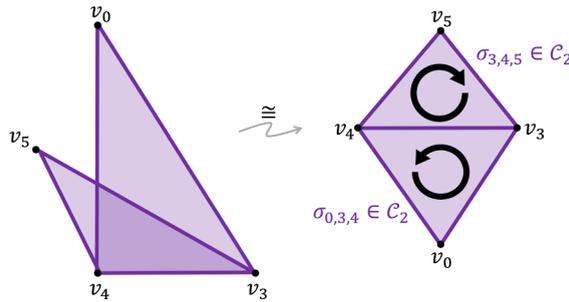


Fig. 7. The 2-chain  $\sigma_{3,4,5} + \sigma_{0,3,4} \in C_2$  does not form a path in our NAS simplicial complex.

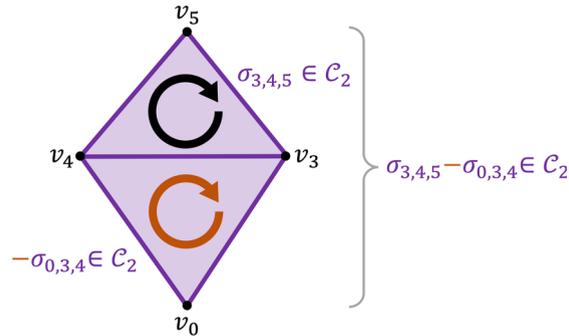


Fig. 8. The 2-chain  $\sigma_{3,4,5} - \sigma_{0,3,4} \in C_2$  forms a path in our NAS simplicial complex.

Suppose that we examine the following two paths composed of 1-simplices in our NAS simplicial complex:  $\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} - \sigma_{0,3} \in C_1$  and  $\sigma_{0,4} - \sigma_{3,4} - \sigma_{0,3} \in C_1$ . Both 1-chains are valid paths; they are indistinguishable until they are visualized on our NAS simplicial complex (Fig. 9).

The brown path in Fig. 9 depicting the second 1-chain does not qualify as a 1-dimensional hole, because it is the boundary of the 2-simplex encoding for the UAL hub feature. On the other hand, the cyan path surrounds a 1-dimensional hole. We can verify that there are two 1-dimensional holes within this data set (the current presentation of  $X_{NAS}$  yields the answer of two 1-dimensional holes more intuitively in the equivalent presentation shown in Fig. 10). However, it would be difficult to visually locate all 1-dimensional holes in a larger or richer data set, and impossible to do so for higher-dimensional holes. Given the definition for a  $k$ -dimensional hole, we now formalize the notion of a *boundary* and of *boundary operators*. These notions are needed to understand the basics of *simplicial homology*, an algebraic-topological construct that rigorously keeps track of holes across a given topological space.

#### 4.4. Boundaries and boundary operators

From our example in Fig. 9, we note the need to define precisely a notion of boundaries as it relates to simplicial complexes. Instead of providing a definition of a boundary, it will be more beneficial to instead introduce the concept of a *boundary operator* that maps a  $k$ -simplex to its boundary. There are computational and methodological reasons for why we use boundary operators.

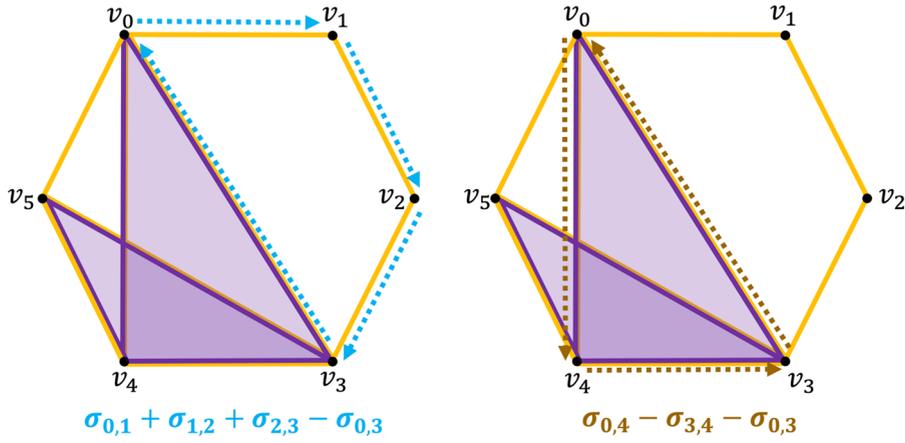


Fig. 9. Visualization of the two valid paths  $\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} - \sigma_{0,3} \in C_1$  and  $\sigma_{0,4} - \sigma_{3,4} - \sigma_{0,3} \in C_1$ . Intuitively, only the cyan path on the left surrounds a 1-dimensional hole.

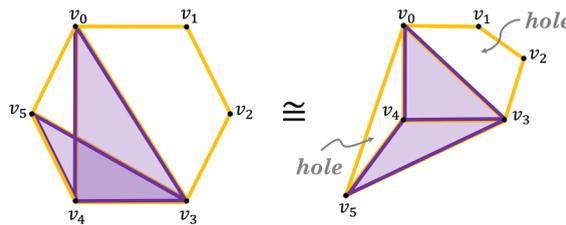


Fig. 10.  $X_{NAS}$  re-drawn to emphasize the existence of two 1-dimensional holes.

Computationally, boundary operators admit matrix representations (Section 4.5), allowing for computer algebra software manipulations. Methodologically, they are important in the definition of *homology groups* in Section 4.6. The boundary of a  $k$ -simplex and the boundary operator is defined as follows:

**Definition 8.** The *boundary operator*, denoted as  $\partial : \Delta^k \rightarrow \Delta^{k-1}$  with  $k > 0$ , takes a  $k$ -simplex  $\Delta^k$  and decomposes it into alternating sums of  $(k - 1)$ -simplices. This alternating sum of  $(k - 1)$ -simplices is the *boundary* of the original  $k$ -simplex. The boundary operator maps explicitly by (5).

$$\left\{ v_0, \dots, v_k \right\} \mapsto \partial \left( \left\{ v_0, \dots, v_k \right\} \right) = \sum_{i=0}^k (-1)^i \left\{ v_0, \dots, \hat{v}_i, \dots, v_k \right\}. \tag{5}$$

The notation  $\hat{v}_i$  or  $\hat{i}$  (the  $v$  is sometimes omitted when it is clear that we are referring to vertices) indicates that vertex  $v_i$  has been *deleted*.

The boundary operator  $\partial : \Delta^k \rightarrow \Delta^{k-1}$  formalizes the decomposition relations in Fig. 4. We illustrate a straightforward application of the boundary operator on a 3-simplex in Fig. 11. A quick note regarding labeling conventions: the vertex labels on a given  $k$ -simplex can always be relabeled to match the canonical  $v_0, \dots, v_i, \dots, v_k$  labeling scheme used in the definition of a boundary operator, as long as *orientation is preserved*. For example, in our NAS simplicial complex, the vertices of the 2-simplex  $\sigma_{0,3,4}$  can be relabeled as  $\sigma_{0,1,2}$ , and then the boundary operator can be explicitly applied to the relabeled 2-simplex. For consistency, the vertices on the resultant boundary of 1-simplices can be “unlabeled” according to the original 2-simplex vertex labels.

Two important properties of the boundary operator is that (1) the boundary operator is a *linear* operator, and (2) the composition  $\partial_{k-1} \circ \partial_k$  *always maps to nullity*. Both properties are surprisingly intuitive – the former indicates that taking the boundary of multiple  $k$ -simplices should give their overall boundary, and the latter indicates that the boundary of a boundary does not exist. While both properties can be proven generally by using the explicit definition of boundary operators given in (5), we will demonstrate the intuitions behind both properties via examples on  $X_{NAS}$ . Recall the 2-chain in  $X_{NAS}$  from Fig. 8; the expected result of applying the boundary operator on this 2-chain is the boundary consisting of the vertices  $v_0, v_4, v_5$ , and  $v_3$ . From the linearity of boundary operators, we have:

$$\partial_2(\sigma_{3,4,5} - \sigma_{0,3,4}) = \partial_2(\sigma_{3,4,5}) - \partial_2(\sigma_{0,3,4}). \tag{6}$$

The boundary operator  $\partial_k$  written with an index of  $k = 2$  denotes that this is the boundary operator that sends a 2-simplex to its boundary consisting of 1-simplices. Recall that we can relabel vertices in order to directly use the explicit definition of the boundary operator given in (5). Specifically, for  $\sigma_{3,4,5}$ , we relabel  $\{3, 4, 5\} \mapsto \{0, 1, 2\}$ , and for  $\sigma_{0,3,4}$ , we relabel  $\{0, 3, 4\} \mapsto \{0', 1', 2'\}$ . Continuing with the boundary calculation, we have:

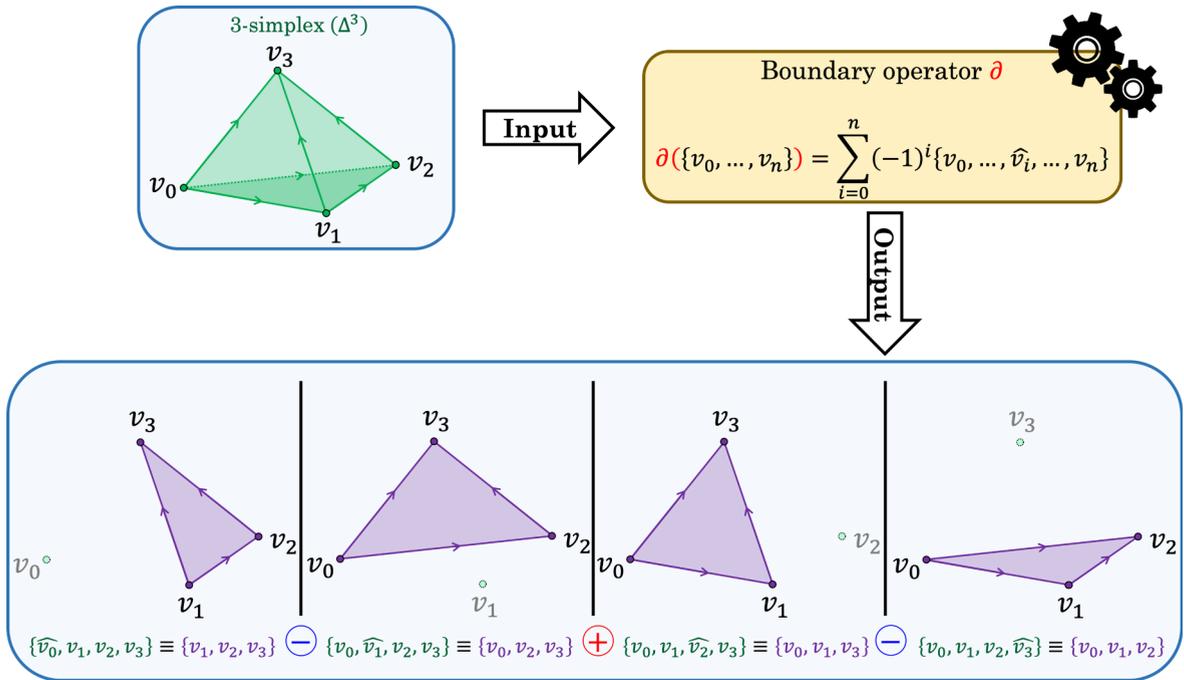


Fig. 11. An application of the boundary operator on a 3-simplex.

$$\begin{aligned}
 \partial_2(\sigma_{3,4,5}) - \partial_2(\sigma_{0,3,4}) &= \underbrace{((-1)^0\sigma_{1,2} + (-1)^1\sigma_{0,2} + (-1)^2\sigma_{0,1})}_{\text{relabelled vertex labeling for } \sigma_{3,4,5}} - \underbrace{((-1)^0\sigma_{1,2'} + (-1)^1\sigma_{0,2'} + (-1)^2\sigma_{0,1'})}_{\text{relabelled vertex labeling for } \sigma_{0,3,4}} \\
 &= (\sigma_{1,2} - \sigma_{0,2} + \sigma_{0,1}) - (\sigma_{1,2'} - \sigma_{0,2'} + \sigma_{0,1'}) \\
 &= \underbrace{(\sigma_{4,5} - \sigma_{3,5} + \sigma_{3,4})}_{\text{"un-labeled" vertices}} - \underbrace{(\sigma_{3,4} - \sigma_{0,4} + \sigma_{0,3})}_{\text{"un-labeled" vertices}} \\
 &= \sigma_{4,5} - \sigma_{3,5} + \sigma_{0,4} - \sigma_{0,3} \\
 &= \sigma_{0,4} + \sigma_{4,5} - \sigma_{3,5} - \sigma_{0,3} \in C_1.
 \end{aligned} \tag{7}$$

“”The resultant 1-chain  $\sigma_{0,4} + \sigma_{4,5} - \sigma_{3,5} - \sigma_{0,3} \in C_1$  is visualized as the green path in Fig. 12. To demonstrate the second property of boundary operators, we apply the next boundary operator  $\partial_1$  to the resultant 1-chain  $\sigma_{0,4} + \sigma_{4,5} - \sigma_{3,5} - \sigma_{0,3} \in C_1$ :

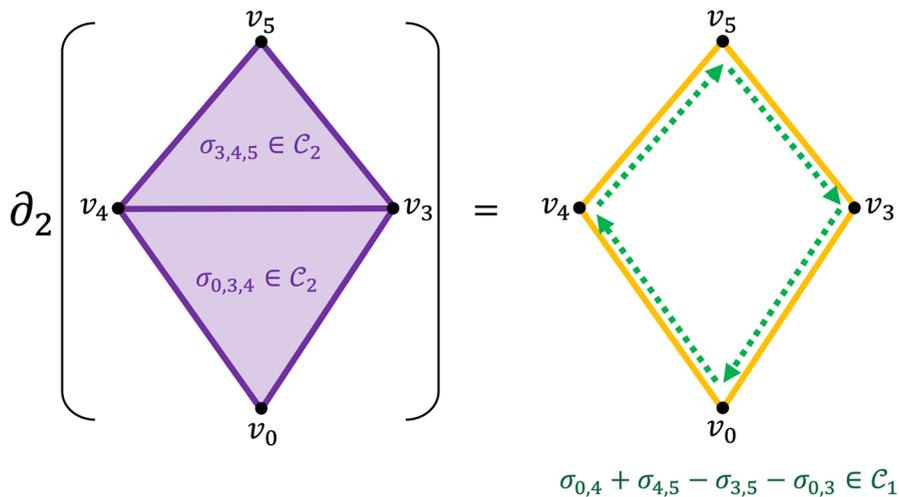


Fig. 12. Linearity of boundary operators demonstrated with the computation of a 2-chain’s boundary.

$$\begin{aligned}
 \partial_1(\partial_2(\sigma_{3,4,5} - \sigma_{0,3,4})) &= \partial_1(\sigma_{0,4} + \sigma_{4,5} - \sigma_{3,5} - \sigma_{0,3}) \\
 &= (\sigma_4 - \sigma_0) + (\sigma_5 - \sigma_4) - (\sigma_5 - \sigma_3) - (\sigma_3 - \sigma_0) \\
 &= 0.
 \end{aligned} \tag{8}$$

Thus, compositions of consecutive boundary operators map to nullity, i.e.,  $\partial_{k-1} \circ \partial_k: C_k \rightarrow \emptyset$ , formalizing the intuition that the *boundary of a boundary* does not exist.

#### 4.5. Matrix representation of boundary operators

While our illustrations of simplicial complexes and boundary operators are useful for purposes of introducing the algebraic-topological foundations of TDA, software implementations typically store these constructs as matrices (Fasy et al., 2014; Bauer et al., 2017; Henselman et al., 2018). Any linear operator  $\mathfrak{T}: V \rightarrow W$  can be represented by a matrix with dimensions related to the dimensions of the domain and codomain of  $\mathfrak{T}$ . More specifically, if  $M_{\mathfrak{T}}$  is the matrix representation of the linear operator  $\mathfrak{T}$ , then it will have dimensions  $\dim(W) \times \dim(V)$ . We can explicitly write down the matrix representation  $D_2$  of the boundary operator  $\partial_2: C_2 \rightarrow C_1$  that we applied to the example in Fig. 12 (colored text is used to annotate the columns and

$$\text{rows): } D_2 = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 1 & 1 \\ 0 & 1 \\ 0 & -1 \end{pmatrix} \begin{matrix} \sigma_{0,3} \\ \sigma_{0,4} \\ \sigma_{3,4} \\ \sigma_{4,5} \\ \sigma_{3,5} \end{matrix} .$$

The domain of  $\partial_2$  is  $C_2$  spanned by the two 2-simplices  $\sigma_{0,3,4}$  and  $\sigma_{3,4,5}$ , so we have that  $\dim(C_2) = 2$ , and hence  $D_2$  must have two columns. The codomain is the 1<sup>st</sup> chain group  $C_1$  associated to the NAS simplicial subcomplex in Fig. 12, spanned by five 1-simplices. Thus,  $D_2$  must have five rows. The (\*) column of  $D_2$  constructs the 2-simplex  $\sigma_{0,3,4}$ ; since the standard orientation on  $\sigma_{0,3,4}$  is counterclockwise, the coefficient on  $\sigma_{0,3}$  and  $\sigma_{3,4}$  is positive as it follows the orientation, and the coefficient on  $\sigma_{0,4}$  is negative since it is reversed with respect to the orientation. By the same construction, the (\*\*) column of  $D_2$  represents the other 2-simplex  $\sigma_{3,4,5}$ .

Recall that the 2-chain we applied the boundary operator to was  $\sigma_{3,4,5} - \sigma_{0,3,4} \in C_2$ . We can rewrite this in a vectorized form

$$\vec{\sigma}: \vec{\sigma} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \begin{matrix} \sigma_{0,3,4} (*) \\ \sigma_{3,4,5} (**) \end{matrix} .$$

Note that the row order on  $\vec{\sigma}$  matches the column order on  $D_2$ . This is important because the vector representing the 2-chain  $\sigma_{3,4,5} - \sigma_{0,3,4}$  lives in the domain of the boundary operator being represented by  $D_2$ . We can now apply the boundary operator on our 2-chain as a matrix left-multiplication, yielding the 1-chain path illustrated in Fig. 12 (right). All simplicial complexes, subcomplexes,  $k$ -simplices, and boundary operators can be stored in a convenient matrix form, thus enabling

$$\text{computation. } D_2 \vec{\sigma} = \begin{pmatrix} -1 \\ 1 \\ 0 \\ 1 \\ -1 \end{pmatrix} \begin{matrix} \sigma_{0,3} \\ \sigma_{0,4} \\ \sigma_{3,4} \\ \sigma_{4,5} \\ \sigma_{3,5} \end{matrix} .$$

We can now count the number of holes within the NAS simplicial complex, creating the mathematical construct that reflects the intuition found in Fig. 10. In Section 4.6, we introduce two new sets that can be naturally defined on our NAS simplicial complex via boundary operators. Informally, one set contains candidates that could be holes, with some false positives (non-holes) thrown in. The other set contains the identity of the false positives. The natural next step would be to “divide out” the latter from the former until we are only left with the holes, which is the idea behind *simplicial homology*.

#### 4.6. Simplicial homology

Recall, from Fig. 9, some of the differences between the two paths: One path surrounds a 1-dimensional hole, whereas the other path does not. The second path failed to surround a hole because it is the boundary of a higher simplex. We impose two conditions to keep our examples simple: (1) We restrict ourselves to looking for 1-dimensional holes within  $X_{NAS}$ , and (2) we consider chain groups that admit coefficients from the set of integers modulo 2 (that is, from the set  $\mathbb{Z}_2 \cong \{0, 1\}$ ). The second restriction greatly simplifies the  $k$ -chains we will examine; instead of integer coefficients in our  $k$ -chain, all even integers are mapped to 0 and all odd integers are mapped to 1. Surprisingly, this choice of coefficients to work over does not affect TDA/PH. Specifically, this is due to the *universal coefficient theorem*; while this theorem is integral to algebraic topology, a full understanding is not needed for TDA/PH, and we refer readers to any standard algebraic topology textbook (e.g. (Hatcher, 2002)) for a much more in-depth overview.

Under the previous  $\mathbb{Z}$  coefficients, a cancellation would occur if we had  $\sigma_{i,j} - \sigma_{i,j} = 0$ . Now, under  $\mathbb{Z}_2$  coefficients, we have  $\sigma_{i,j} + \sigma_{i,j} = 2\sigma_{i,j} \equiv 0 \pmod 2$ . Note that under  $\mathbb{Z}_2$  coefficients, orientations do not matter since any odd combinations of  $k$ -simplices map to 1, and any even combinations of  $k$ -simplices map to 0. This simplification – and more generally the ability to flexibly pick

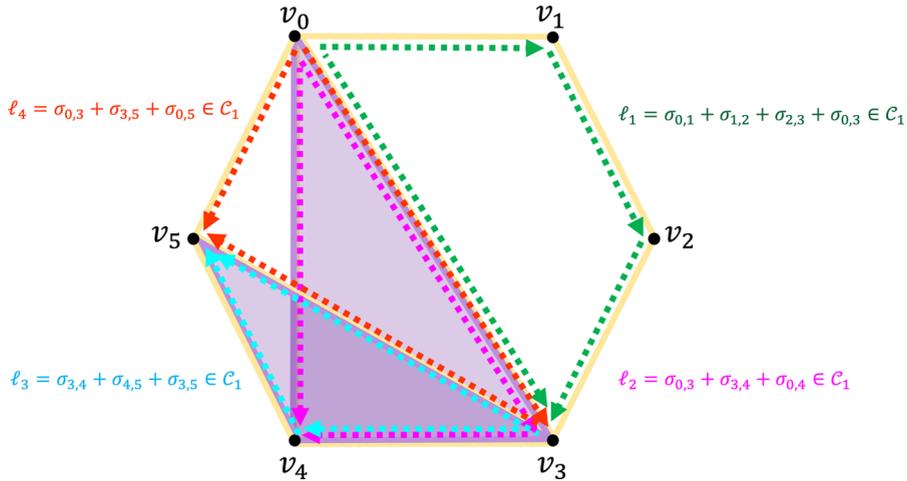


Fig. 13. The four 1-cycles that exist in our NAS simplicial complex.

coefficients to work over – has additional computational advantages (Fugacci et al., 2014; Bhattacharya et al., 2015).

4.6.1. Cycles and boundaries

Following the aforementioned line of thought regarding constructing a set of all possible hole candidates, we go back to our redrawn NAS simplicial complex in Fig. 10. There are four possible candidates  $\ell_1$ ,  $\ell_2$ ,  $\ell_3$ , and  $\ell_4$  as presented in Fig. 13. While it may be tempting to call these four candidates closed paths, note that the orientations on the 1-simplices do not line up, although they are certainly closed loops. The formal terminology for these objects  $\ell_1$ ,  $\ell_2$ ,  $\ell_3$ , and  $\ell_4$  are 1-cycles, with  $k$ -cycles being higher-dimensional equivalents:

**Definition 9.** A  $k$ -cycle is any combination of  $k$ -simplices that form a valid closed path with a finite number of orientation exchanges.

The next observation is that any other 1-cycles can be found in the form  $\sum_i a_i \ell_i$  with  $a_i \in \mathbb{Z}_2$ . For example, to form the 1-cycle that goes around the entire NAS simplicial complex (Fig. 14), we can examine the linear combination  $\ell_1 + \ell_3 + \ell_4$ :

$$\begin{aligned}
 \ell_1 + \ell_3 + \ell_4 &= (\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{0,3}) + (\sigma_{3,4} + \sigma_{4,5} + \sigma_{3,5}) + (\sigma_{0,3} + \sigma_{3,5} + \sigma_{0,5}) \\
 &= \sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{3,4} + \sigma_{4,5} + \sigma_{0,5} + 2\sigma_{0,3} + 2\sigma_{3,5} \\
 &\equiv \sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{3,4} + \sigma_{4,5} + \sigma_{0,5}.
 \end{aligned}
 \tag{9}$$

Note that since we are working under  $\mathbb{Z}_2$  coefficients, the terms  $2\sigma_{0,3}$  and  $2\sigma_{3,5}$  vanish because the coefficients are even. The resultant 1-cycle  $\sigma_{0,1} + \sigma_{1,2} + \sigma_{2,3} + \sigma_{3,4} + \sigma_{4,5} + \sigma_{0,5}$  is depicted in blue on the left-hand side of Fig. 14. We can now define the following space that contains all possible hole candidates:

**Definition 10.** The space of 1-cycles, denoted  $\mathcal{Z}_1$ , for our NAS simplicial complex is defined as follows:

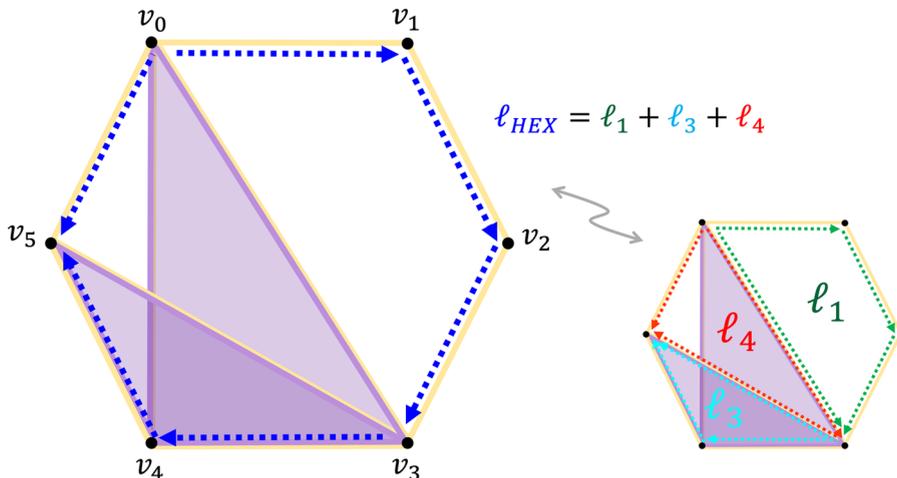


Fig. 14. Creating the 1-cycle  $\ell_{HEX}$  through linear combinations of  $\ell_1$ ,  $\ell_2$ ,  $\ell_3$ , and  $\ell_4$  under  $\mathbb{Z}_2$  coefficients.

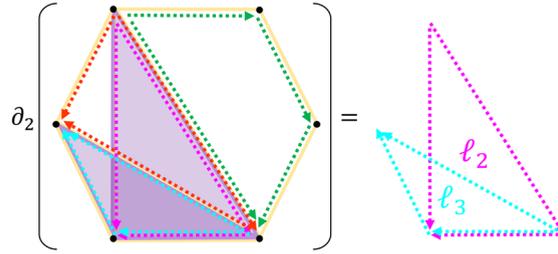


Fig. 15. The result of applying the boundary operator  $\partial_2: C_2 \rightarrow C_1$  to our NAS simplicial complex.

$$\mathcal{Z}_1 = \{a_1\ell_1 + a_2\ell_2 + a_3\ell_3 + a_4\ell_4 \mid a_i \in \mathbb{Z}_2, \forall i\}. \tag{10}$$

Since the boundary operator  $\partial_1$  would map all four 1-cycles that span  $\mathcal{Z}_1$  to triviality (the boundary of a boundary does not exist), we arrive at the following general definition of  $\mathcal{Z}_1$ :

**Definition 11.** The space of 1-cycles, denoted by  $\mathcal{Z}_1$ , is defined as follows:

$$\mathcal{Z}_1 = \ker(\partial_1). \tag{11}$$

Analogously, the space of  $k$ -cycles, denoted by  $\mathcal{Z}_k$ , is defined as follows:

$$\mathcal{Z}_k = \ker(\partial_k). \tag{12}$$

Recall that the kernel of some linear operator  $\mathfrak{T}: V \rightarrow W$  is the subset of  $V$  that contains all elements  $v \in V$  such that  $\mathfrak{T}(v) = 0 \in W$ . In this case, since  $\partial_1(\ell_i) = 0$  for all 1-cycles that span  $\mathcal{Z}_1$ , the two sets  $\ker(\partial_1)$  and  $\mathcal{Z}_1$  coincide.

If  $\mathcal{Z}_1$  contains all possible hole candidates, then we need to find an analogous space containing all false positives that cannot possibly be holes. The appearance of boundary operators in the definition of  $\mathcal{Z}_1$  can help us identify these non-holes, by applying the boundary operator  $\partial_2$  to  $X_{NAS}$ . We illustrate such an operation in Fig. 15.

Since the boundary operator  $\partial_2$  only affects 2-chains in  $C_2$ , it sends those 2-chains to their boundaries composing of 1-chains (more specifically, 1-cycles) in  $C_1$ . These are the elements that we want to exclude from  $\mathcal{Z}_1$ . We can now give this space of false positives a name, and a definition that hinges on the boundary operator  $\partial_2$ :

**Definition 12.** The space of 1-boundaries, denoted by  $\mathcal{B}_1$ , is defined as follows:

$$\mathcal{B}_1 = \text{im}(\partial_2). \tag{13}$$

Analogously, the space of  $k$ -boundaries, denoted by  $\mathcal{B}_k$ , is defined as follows:

$$\mathcal{B}_k = \text{im}(\partial_{k+1}). \tag{14}$$

The notation  $\text{im}(\partial_2)$  denotes the image of the boundary operator, which is a subset of 1-cycles in  $\mathcal{Z}_1 \subset C_1$ . Note that since all 1-boundaries must be 1-cycles, but not all 1-cycles are 1-boundaries (the ones that are not are exactly the 1-dimensional holes we are looking for), we must have that  $\mathcal{B}_1 \subseteq \mathcal{Z}_1$ . Analogously, we can always construct the space of  $k$ -cycles  $\mathcal{Z}_k$  and the space of  $k$ -boundaries  $\mathcal{B}_k$  for  $k \geq 0$ , and the intuition behind the relation  $\mathcal{B}_k \subseteq \mathcal{Z}_k$  remains the same. Furthermore, the spaces of  $k$ -cycles and  $k$ -boundaries are actually groups as well.

#### 4.6.2. Homology groups and Betti numbers

Homology groups arise as the formal results of “dividing out” the imposters (the  $k$ -boundaries) from the possible candidates (the  $k$ -cycles). Note that since  $\mathcal{Z}_k$  and  $\mathcal{B}_k$  are sets (more formally, groups) of all possible  $k$ -cycles and  $k$ -boundaries formed from linear combinations of a finite number of cycles (these cycles can be thought of as being analogous to basis for vector spaces in linear algebra), the notion of quotients carries through in this context by acting on the cycles themselves. In other words, shared cycles in the basis for both  $\mathcal{Z}_k$  and  $\mathcal{B}_k$  can be divided out if the appropriate quotient is taken. Since  $\mathcal{B}_k \subseteq \mathcal{Z}_k$ , only one sensible quotient can be taken, and this quotient actually forms another group:

**Definition 13.** The  $k^{\text{th}}$  homology group for a topological space  $X$ , denoted as  $\mathcal{H}_k(X)$ , is the group obtained from the following quotient:

$$\mathcal{H}_k(X) = \mathcal{Z}_k / \mathcal{B}_k, \tag{15}$$

or alternatively,

$$\mathcal{H}_k(X) = \ker(\partial_k) / \text{im}(\partial_{k+1}). \tag{16}$$

The dimension of  $\mathcal{H}_k(X)$  is exactly equal to the number of  $k$ -dimensional holes in  $X$ . This quantity is known as the  $k^{\text{th}}$  Betti number, denoted by  $\beta_k$ .

Thus, for our NAS simplicial complex, we can compute the 1<sup>st</sup> homology group as follows:

$$\mathcal{H}_1(X_{NAS}) = \mathcal{Z}_1/\mathcal{B}_1 = \langle \ell_1, \ell_2, \ell_3, \ell_4 \rangle_{\mathbb{Z}_2} / \langle \ell_2, \ell_3 \rangle_{\mathbb{Z}_2} \cong \langle \ell_1, \ell_4 \rangle_{\mathbb{Z}_2}. \quad (17)$$

The brackets notation is a shorthand for indicating that the space is spanned by linear combinations of elements within the brackets with coefficients from  $\mathbb{Z}_2$ . Finally, the number of 1-dimensional holes, or alternatively, the 1<sup>st</sup> Betti number  $\beta_1$  is equal to  $\dim(\mathcal{H}_1(X_{NAS})) = 2$ . Note that from the definition of what the 0<sup>th</sup> homology group should be, the 0<sup>th</sup> Betti number  $\beta_0$  measures the number of *connected components* in the simplicial complex. Thus, for our NAS simplicial complex, we have that  $\beta_0 = 1$ .

#### 4.7. Topological data analysis and persistent homology (TDA/PH)

The notions from algebraic topology discussed above serve as the starting foundations for TDA/PH methods. In general, TDA focuses on transforming a given data set into complex-like structures such as  $X_{NAS}$ . Other more nuanced constructions of complexes from data can also be used, as we will see in Sections 6 and 7. After transforming the data into a simplicial complex-like structure, some TDA tools focus on running summary statistics on the structure and its various substructures. Other TDA tools explore the global “shape” of the data trying to find topologically significant features, such as the  $k$ -dimensional holes. Another possible approach is through persistent homology (PH); the key idea here is to dynamically generate successions of simplicial complex-like structures and to look for *persistent topological features*.

We now illustrate ways to bridge the gap between the *abstract results* stemming from TDA/PH to *actual aviation applications* by reviewing three studies that leverage a topological analysis of aviation data (Section 5). To further emphasize the potential insights gained from examining aviation data from a topological perspective, we present in Section 6 a computational case study examining the topology of surface operations at five major US airports using so-called *nerve complexes*, a sub-type of the simplicial complex we introduced in Section 4.1. Thereafter, in Section 7 we discuss the results of our case study, as well as the range of operational and managerial insights derivable from applying TDA/PH to aviation data. Section 7 also explicitly maps topological features such as  $k$ -dimensional holes to physical processes and phenomena in the aviation domain. We conclude our work with numerous potential avenues of future research utilizing TDA/PH in aviation (Section 8), followed by a summary of our work and contributions (Section 9).

### 5. Review of applications of TDA/PH to aviation

With an understanding of important algebraic-topological concepts such as simplicial complexes, homology groups, and Betti numbers, we review three applications of TDA/PH methodologies to various problem settings in aviation (Li and Ryerson, 2018; Zhou et al., 2018; Cho and Yoon, 2018). This is followed by a computational case study investigating the relationship between simplicial complexes and airport surface assets in Section 6. We provide an in-depth discussion on the operational and managerial implications of topological features (e.g. holes) studied in Li and Ryerson (2018), Zhou et al. (2018), and Cho and Yoon (2018) as well as the computational case study in Section 7.

#### 5.1. Detection of anomalous trajectories

The trajectories of aircraft executing airborne holding patterns and missed approach procedures can be classified as *anomalous trajectories* within the terminal arrival airspace. Although rare under nominal conditions, it is important to be able to better detect and characterize these anomalous trajectories to improve terminal area operations. Prior work uses PH to explore the topology of the filtrations of simplicial complexes that emerge from examining an aircraft’s trajectory as a planar latitude-longitude point cloud (Li and Ryerson, 2018). Specifically, the topological features of interest are 1-dimensional holes that signify anomalous trajectories due to airborne holding.

The simplicial complex considered in Li and Ryerson (2018) is an *Alpha complex*; Alpha complexes are a sub-type of simplicial complexes that are formed from a different geometric perspective (Edelsbrunner and Harer, 2009). Specifically, Alpha complexes are closely related to  $\alpha$ -shapes, which we define and analyze in-depth for the study reviewed in Section 5.3. The authors construct the Alpha complexes for each aircraft trajectory and ascertain if the aircraft flew an anomalous trajectory (Fasy et al., 2014; Li and Ryerson, 2018). The results of examining persistent topological features in the case of an aircraft that experienced airborne holding prior to Instrument Landing System (ILS) establishment is shown in Fig. 16. The 1-dimensional hole indicative of airborne holding is shown as a red-highlighted loop in Fig. 16 (right). Correspondingly, the Alpha complex persistence diagram shows the birth and death of several holes – the most significant hole with the longest survival time is detected and shown in the persistence diagram as the red triangle farthest in distance from the diagonal. This persistent topological feature indicates cyclic behavior within the trajectory data set, correctly identifying this aircraft as having encountered airborne holding.

#### 5.2. Characterizing network robustness via homology

An algebraic-topological approach is coupled with a graph-theoretic approach in Zhou et al. (2018) to characterize robustness in the context of random and targeted node removal-type attacks in highly connected and complex networks. Given a network represented by a canonical graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  and a probability  $q$  of node removal during an attack, the more traditional graph-theoretic measures computed in Zhou et al. (2018) include the network connectivity  $\gamma(\mathcal{G})$ , and a fraction  $\delta(Q)$  that indicates the fraction of nodes within so-called giant components after the removal of  $Q = q|\mathcal{V}|$  nodes due to an attack.  $\delta(Q)$  provides a measure of the giant

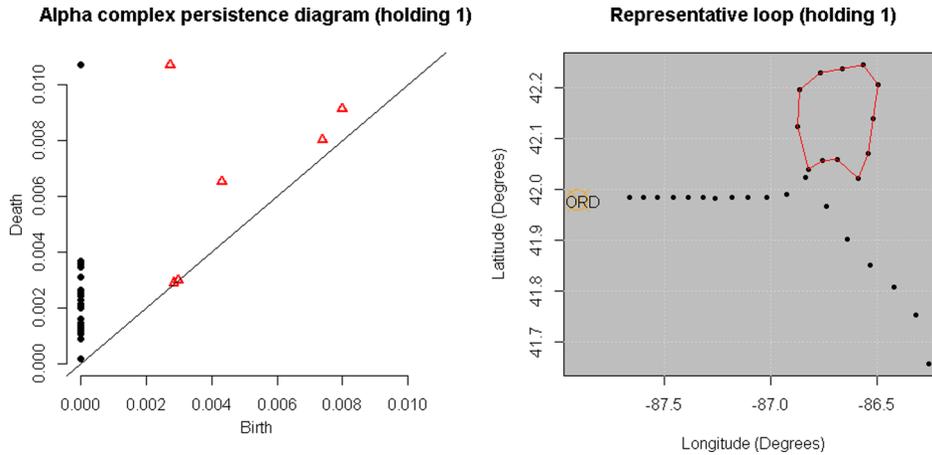


Fig. 16. Alpha complex persistence diagram (left) and representative loop in latitude-longitude space (right) for the base leg and final approach trajectory of a flight that experienced in-air holding prior to establishment on the ILS (Li and Ryerson, 2018).

component size relative to the whole network, and is used by the authors to compute a robustness  $R$ -index. These metrics are evaluated using (18).

$$\gamma(\mathcal{G}) = \frac{2|\mathcal{E}|}{|\mathcal{V}|^2 - |\mathcal{V}|}; \quad R(\mathcal{G}) = \frac{1}{|\mathcal{V}|} \sum_{Q=1}^{|\mathcal{V}|} \delta(Q). \tag{18}$$

The algebraic-topological measure that is used in Zhou et al. (2018) to characterize the topology of the network under node-removal attacks is a ratio  $h_k$  of the number of  $k$ -dimensional holes to the number of connected components in the network. This ratio  $h_k$  leverages homology to succinctly measure the impact of removing a node on the intrinsic topology of the network. The scenario shown in Fig. 17 represents the delay network experienced by American Airlines hubs during a period of bad weather across the NAS. The 1-simplices encode pairwise departure delays, and the 2-simplex encodes Miles-In-Trial restrictions affecting three airports. The delay topology on the right has two 1-dimensional holes and one connected component, resulting in a ratio of  $h_1 = 2$ . Suppose that an “attack” results in the node representing John F. Kennedy International Airport (JFK) being removed; this may be due to an airline-specific technical issue that requires a ground stop. The delay topology changes dramatically, as captured by the change in the ratio to  $h_1 = 1$ , due to the fact that there are now two connected components.

5.3.  $\alpha$ -shapes and geofencing for unmanned aerial vehicles (UAVs)

We review work by Cho and Yoon (2018) that presents an application of topological methods to UAV geofencing applications. The current expansion of UAV technology is exemplified both in the increasing affordability and accessibility of small UAVs for personal use as well as the growing interest on the part of companies in harnessing the ability of UAVs for commercial and industrial

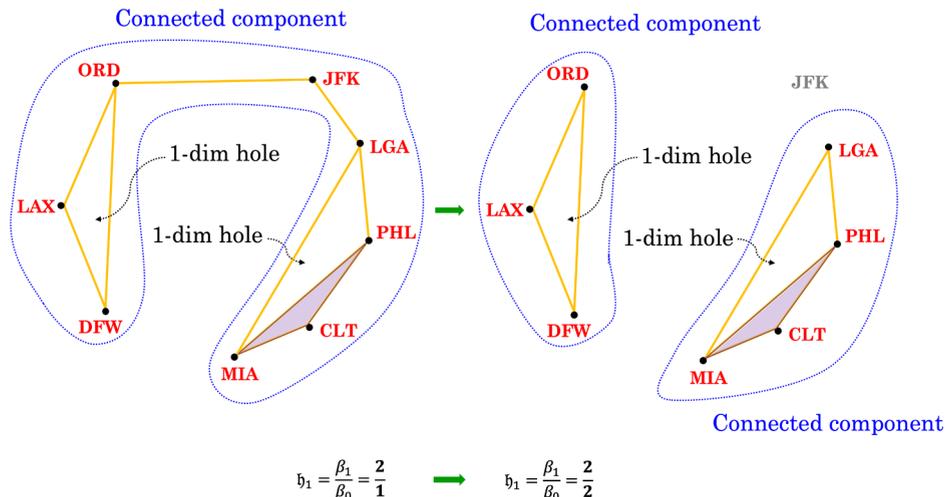
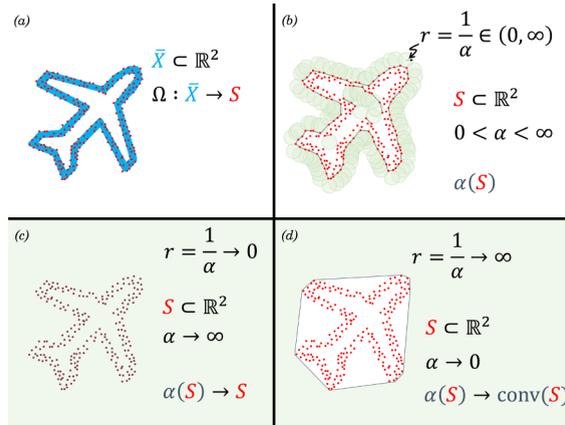


Fig. 17. Hypothetical delay topologies at AAL hubs before and after a node-removal “attack”.



**Fig. 18.** (a) Sampled set of red points  $S \subset \mathbb{R}^2$  from  $\bar{X}$  space; (b)  $\alpha$ -shape for nonzero, finite radius  $r = 1/\alpha$ ; (c)  $\alpha$ -shape degenerates to set of points  $S$  as  $r$  vanishes ( $\alpha \rightarrow \infty$ ); (d)  $\alpha$ -shape degenerates to convex hull  $\text{conv}(S)$  as  $r \rightarrow \infty$  ( $\alpha \rightarrow 0$ ).

operations. The natural progression of this technology is the development of low-altitude, high-density urban airspace used by both guided and autonomous UAVs. While the structure of higher altitude airspace occupied by general and commercial aviation is relatively rigid with strict separation minima and traffic procedures, such characterizations for the future UAV-centric airspace are active areas of research (Sunil et al., 2018).

Research concerning unmanned aircraft system traffic management contain settings and data sets ripe for topological analysis; we review one prominent example presented in Cho and Yoon (2018). The authors of Cho and Yoon (2018) investigate the structure and capacity of a given urban airspace in the context of UAV usage by using topological *keep-in* and *keep-out* geofences constructed via  $\alpha$ -shapes. In Fig. 18 we give an illustration of the  $\alpha$ -shape construction used in Cho and Yoon (2018). For ease of visualization, we present this illustrative example in  $\mathbb{R}^2$ , but the construction generalizes to higher dimensions. Suppose we sample  $\bar{X} \subset \mathbb{R}^2$  via sampling function  $\Omega: \bar{X} \rightarrow S$ , obtaining a set of data points  $S$  (Fig. 18(a)). We can construct a closed disk  $D_r \cup \partial D_r$ , where the radius depends inversely on the  $\alpha$  parameter and  $\partial D_r$  is the boundary.

We first examine the construction of the  $\alpha$ -shape (denoted as  $\alpha(S)$ ) when  $\alpha \in (0, \infty)$ . As shown in Fig. 18(b), we construct an edge of  $\alpha(S)$  if it is possible to construct a closed disk  $D_r \cup \partial D_r$  such that two members of  $S$  lie on  $\partial D_r$ , and  $S \cap D_r = \emptyset$ . Note that in Fig. 18(b), one could select the appropriate  $\alpha$  parameter such that a rough outline of  $\bar{X}$  is recovered by  $\alpha(S)$ . Two degenerate cases as  $\alpha$  tends to  $\infty$  and 0 are illustrated in Fig. 18(c) and (d), respectively. Intuitively, as  $\alpha \rightarrow \infty$ , the radii of all the possible disks  $D_r \cup \partial D_r$  we could construct vanishes, and we cover the entirety of  $\mathbb{R}^2 \setminus S$ . Hence, the case where  $\alpha \rightarrow \infty$  we have that  $\alpha(S)$  degenerates to  $S$ . On the other hand, if  $\alpha \rightarrow 0$  we have that the radius of each closed disk grows to infinity, wherein each closed disk – losing its curvature as  $r \rightarrow \infty$  – becomes indistinguishable from a closed half-plane. In trying to construct  $\alpha(S)$  with these unwieldy closed half-planes, we end up with the convex hull  $\text{conv}(S)$  as the  $\alpha$ -shape of  $S$  (Fig. 18(d)).

The authors of Cho and Yoon (2018) use the construction of  $\alpha$ -shapes given above, albeit in  $\mathbb{R}^3$ , to construct three-dimensional keep-in geofences. They found that these keep-in spaces via  $\alpha$ -shapes provide an upper-bound on the amount of usable airspace, whereas keep-out geofences give more conservative results. Keep-out geofences were not modeled via  $\alpha$ -shapes since their goal was to provide a uniform airspace buffer between UAVs and static obstacles such as buildings and other infrastructures. This produces a natural primal-dual relationship between the two geofences, the trade-offs of which are explored in Cho and Yoon (2018) using both simulated urban environments as well as actual terrain data from a heavily built-up district of Seoul, South Korea. The method of  $\alpha$ -shapes for constructing keep-in geofences presented in Cho and Yoon (2018) is a pertinent example of harnessing computational topology in aviation research. Furthermore, the authors of Cho and Yoon (2018) note in their conclusion that TDA/PH could be used in the future to assess the connectivities and continuities of usable urban airspace data sets extracted from their keep-in and keep-out partitioning.

## 6. Case study: nerve topologies of airport configurations

The setting of networks naturally appears when examining the airport surface. The active runways and ramp areas can be seen as sinks and sources where aircraft originate and depart from, traveling on paths given by taxiways and inactive runways. Current research in modeling, optimizing, and controlling airport surface operations focuses on graph-based approaches to represent the system of runways and taxiways at an airport (Khadilkar and Balakrishnan, 2014; Guépet et al., 2017). However, this graph representation is constrained to pairwise relationships, unlike the aforementioned ability of simplicial complexes to represent higher-order interactions. An interesting application of TDA/PH may be to examine topological features within the *nerve complex* corresponding to the prevailing configuration at an airport. We introduce the notion of *covers* for topological spaces in Section 6.1, which will allow us to introduce nerve complexes as a sub-type of simplicial complexes. We then present our computational case study investigating the nerve complexes of airports within the US NAS.

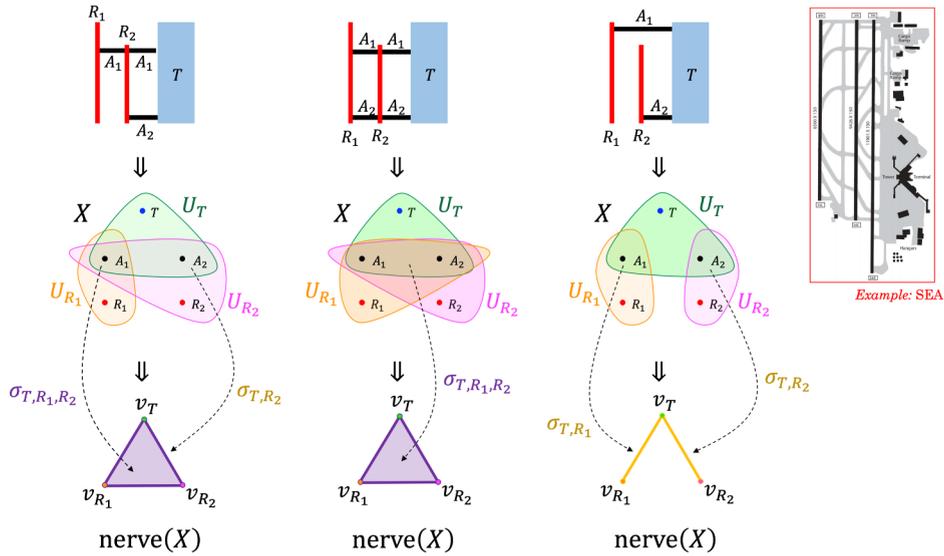


Fig. 19. Underlying nerve topology of an airport configuration: An example of an airport with parallel runways (e.g. Seattle-Tacoma International Airport).

6.1. Covers and nerves

Depending on prevailing wind conditions and operational demand, airports tend to operate a specific *runway configuration* that assigns some subset of available runways as arrival and departure runways. Furthermore, depending on the runway configuration, there will be specific taxiway flow patterns to guide aircraft to and from the runways and ramp areas. Given a simplified example airport such as the ones presented in Figs. 19 and 20, we can define the set of ramp areas as  $\mathcal{T} = \{T\}$ , the set of active taxiways as  $\mathcal{A} = \{A_1, A_2\}$ , and the set of active runways as  $\mathcal{R} = \{R_1, R_2\}$ . Given an airport and its current runway-taxiway configuration, we can construct a simple topological space  $X$  that contains  $|\mathcal{T}| + |\mathcal{A}| + |\mathcal{R}|$  point elements, each corresponding one-to-one with a specific ramp area, taxiway, or runway.

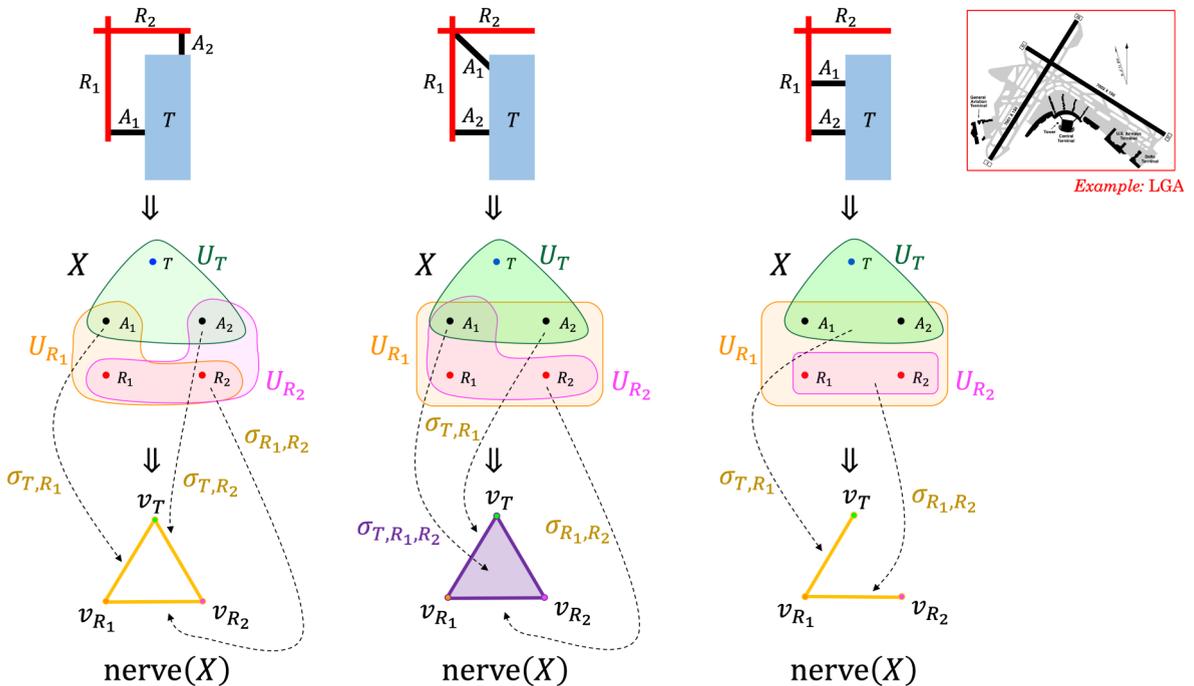


Fig. 20. Underlying nerve topology of an airport configuration: An example of an airport with two perpendicular intersecting runways is shown here (e.g. New York-LaGuardia Airport).

In order to construct the nerve complex corresponding to the runway-taxiway configuration at a specific airport, we first construct a cover for the topological space  $X$  associated with the runway-taxiway configuration at an airport. Let  $\mathcal{I} = \{T, R_1, R_2\}$  be an index set; we now define a topological cover as follows:

**Definition 14.** The cover of a topological space  $X$  is a collection of subsets  $U_i \in \mathcal{U}, \forall i \in \mathcal{I}$  such that their union forms  $X$ .

We associate each  $U_i$  with a ramp area in  $\mathcal{T}$  or runway in  $\mathcal{R}$ . The point elements in  $X$  is included within a particular  $U_i$  if and only if the airport surface structure associated with that point element is directly reachable from the ramp area or runway associated with the index  $i \in \mathcal{I}$  for that particular  $U_i \in \mathcal{U}$ . If we look at the leftmost configuration for the airport shown in Fig. 19, taxiways  $A_1$  and  $A_2$  are directly reachable from ramp area  $T$ , so the green-colored subset  $U_T$  encompasses three point elements in  $X$  representing  $T, A_1$ , and  $A_2$ . The other subsets  $U_{R_1}$  and  $U_{R_2}$  are constructed analogously. We now give the formal definition of the nerve complex of a topological space  $X$  using the language of covers:

**Definition 15.** The nerve complex of a topological space  $X$ , denoted by  $\text{nerve}(X)$ , corresponding to a valid cover of  $X$  given by  $\mathcal{U}$  is the collection of  $k$ -simplices mapped to by nonempty intersections of  $U_i \in \mathcal{U}$ , where  $k$  indicates the number of  $U_i$  subsets that intersect. The indices of the intersecting subsets map to the vertices of the  $k$ -simplices.

We can explicitly use the definition of nerve complexes to construct the nerve complex for airport surface assets. Using the same leftmost configuration in Fig. 19 as an example, there are two non-empty intersections given by  $U_T \cap U_{R_1} \cap U_{R_2} = \{A_1\}$  and  $U_T \cap U_{R_2} = \{A_2\}$ . The first nonempty intersection of three subsets map to the 3-simplex  $\sigma_{T,R_1,R_2}$ , and the second non-empty intersection of two subsets map redundantly to the 2-simplex  $\sigma_{T,R_2}$ . Thus,  $\text{nerve}(X)$  in this example is exactly the 3-simplex  $\sigma_{T,R_1,R_2}$ . All nerve complexes at the bottom of Figs. 19 and 20 are constructed in the same manner.

Through examining nerve complexes, we can already begin to connect various topological features to operational realities. Although the taxiways in use are different between the leftmost and center configurations in Fig. 19, their underlying nerve complexes are identical. The rightmost configuration in Fig. 19 shows a non-trivial change in the nerve complex; a degeneration from a 3-simplex to a 2-chain consisting of 2-simplices occurs when taxiway  $A_1$  serves only runway  $R_1$ , and taxiway  $A_2$  services only runway  $R_2$ . In an airport with intersecting runways, such as the one presented in Fig. 20, a similar configuration where taxiway  $A_1$  services only runway  $R_1$  and taxiway  $A_2$  services only runway  $R_2$  results in an underlying nerve complex that is a 2-chain with a detectable 1-dimensional hole. Loosely speaking, these nerve complexes trade local information that a graph-theoretic representation contains in order to look at the underlying topology associated with a particular airport runway-taxiway configuration.

## 6.2. Constructing nerve complexes of airports from real operational data

In order to construct and analyze nerve complexes for our airports of interest in this computational case study, we first collect the necessary runway configuration data and declared capacities from the Aviation System Performance Metrics (ASPM) database maintained by the US Federal Aviation Administration (Section 6.2.1). Since active taxiways are not included in ASPM, we develop a framework for inferring active taxiway flow patterns via air traffic control audio (Section 6.2.2). We then construct and interpret the corresponding nerve complexes for our case study airports (Sections 6.2.3 and 6.2.4).

### 6.2.1. Runway configurations and declared capacities

The five case study airports are the five highest-ranked US airports in terms of total enplanement statistics for 2017 (Federal Aviation Administration, 2016); in order, these airports are Hartsfield-Jackson Atlanta International Airport (ATL), Los Angeles International Airport (LAX), Chicago O'Hare International Airport (ORD), Dallas/Fort Worth International Airport (DFW), and Denver International Airport (DEN). We retrieve the runway configurations and declared capacities data for these airports for November 2018 via ASPM (Federal Aviation Administration, 2018a). For each airport, the maximum total declared capacity – the Airport Arrival Rate (AAR) plus the Airport Departure Rate (ADR) – and the minimum total declared capacity is located and retrieved, along with the accompanying runway configurations. Since ASPM records hourly data, we will refer to the maximum total declared capacity and its runway configuration as the “high-capacity scenario” of the airport, and the minimum total declared capacity and its runway configuration as the “low-capacity scenario”.

### 6.2.2. Inferring active taxiway configurations via air traffic control audio

Since taxiway configurations and active taxiways are not recorded in ASPM, in order to accurately re-create the active taxiways during high- and low-capacity scenarios, we retrieve archived audio recordings of ground controller frequencies at the five case study airports (LiveATC, 2018). Since these ground controller audio archives are given in 30-min intervals, for each capacity scenario, the first 30-min segment was used to retrieve the list of active taxiways used by arriving and departing aircraft. arrivals and departures. Repeating this process for all five case study airports and combining information regarding active taxiways with the runway configurations provided via ASPM, we are able to construct airport runway-taxiway configuration diagrams for all five case study airports. An example of these airport runway-taxiway configuration diagrams for ORD during its high- and low-capacity scenarios can be found in the left panels of Figs. 21 and 22, respectively. Terminal and ramp areas are given in blue, taxiways in gray, arrival runways in red, and departure runways in green.

### 6.2.3. Constructing nerve complexes for our case study airports

Recall the definitions of covers and nerve complexes associated with a given topological space provided in Section 6.1. Given the

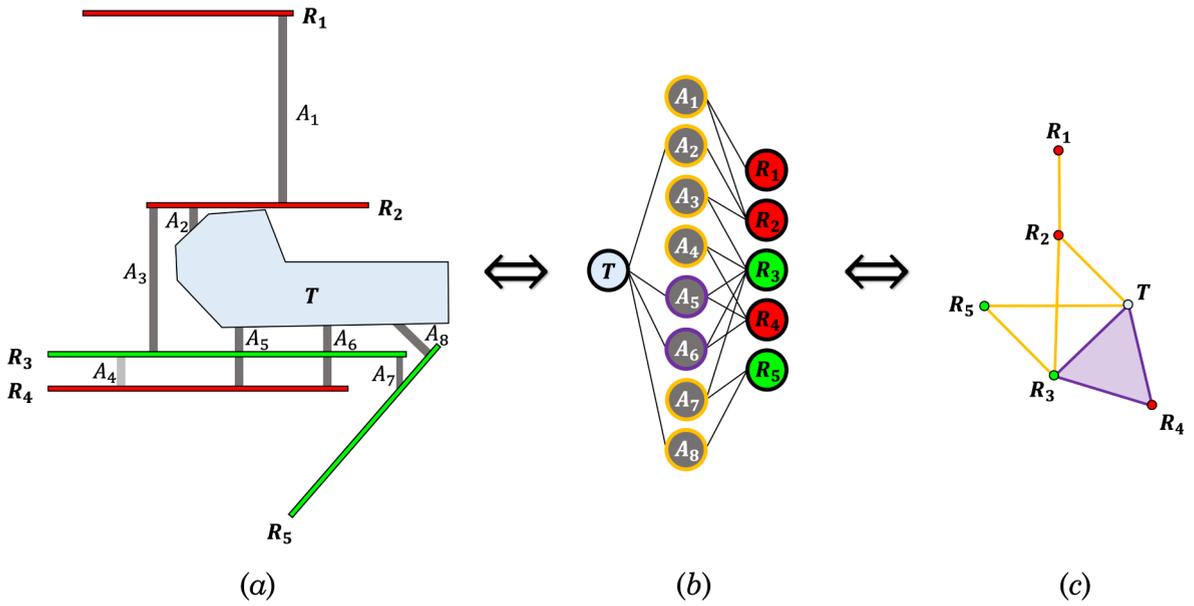


Fig. 21. (a) Complete airport runway-taxiway configuration diagram for ORD high-capacity scenario, (b) the associated intersection graph, and (c) the associated nerve complex.

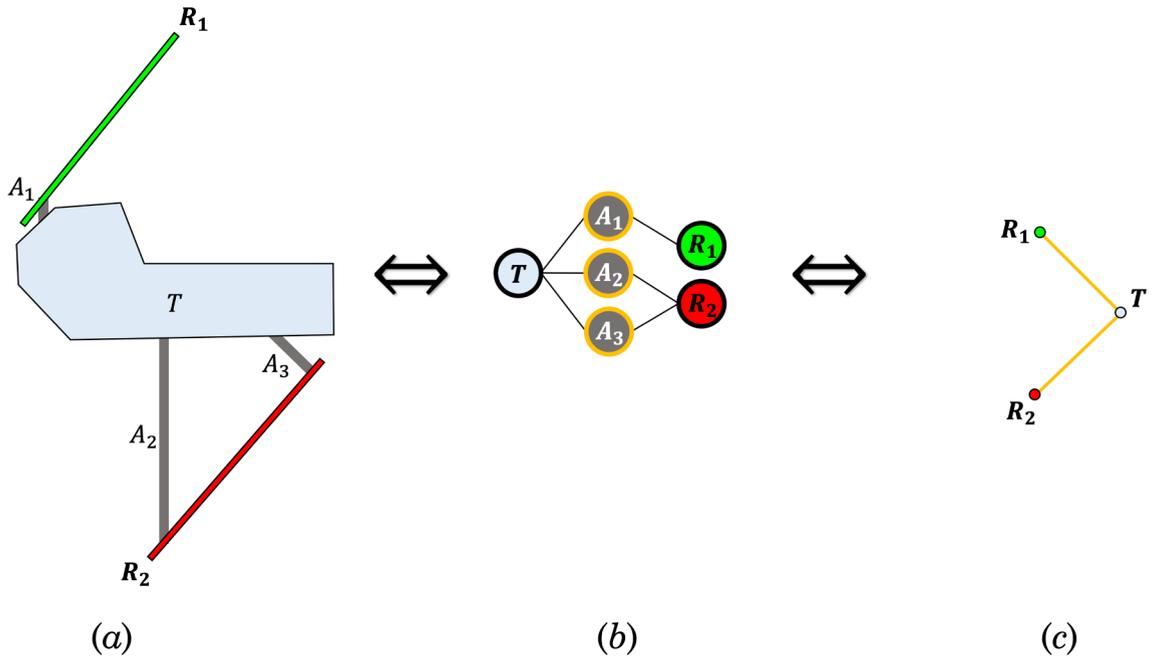


Fig. 22. (a) Complete airport runway-taxiway configuration diagram for ORD low-capacity scenario, (b) the associated intersection graph, and (c) the associated nerve complex.

runway-taxiway configuration diagram at an airport, we can construct the associated nerve complex. Using ORD during its high-capacity scenario as the guiding example, we first attach proxy labels  $R_1, \dots, R_N$  to each runway, and labels  $A_1, \dots, A_M$  to each taxiway. The ramp area is labeled as  $T$ . From the airport runway-taxiway configuration diagram for ORD in its high-capacity scenario (Fig. 21(a)), we construct the associated intersection graph (Fig. 21(b)). The intersection graph contains a node for each active taxiway and runway asset, as well as a node representing the ramp area. For illustrative purposes, the nodes in the intersection graph are colored in accordance to the airport runway-taxiway configuration diagram. Edges are drawn between nodes in the intersection graph representing a direct connection between the ramp and runway nodes to a taxiway node. For example, taxiway  $A_1$  connects  $R_1$  and  $R_2$ ; thus, an edge  $\{A_1, R_1\}$  is drawn along with an edge  $\{A_1, R_2\}$ .

The intersection graph allows us to easily write down the various non-empty intersections of covers needed to construct the

corresponding nerve complex. The number of edges attached to each taxiway node is exactly the number of covers that intersect in a non-empty manner, and the nodes that are connected to that taxiway node by the edges are the indices of the intersecting covers. For example, the  $A_1$  taxiway node has two edges connecting it to runways  $R_1$  and  $R_2$ . Thus, this is a non-empty intersection of two covers, i.e.  $U_{R_1} \cap U_{R_2} = \{A_1\} \neq \emptyset$ . Recall from Section 6.1 that this corresponds to the 1-simplex  $\sigma_{R_1, R_2}$ , which can be found in Fig. 21(c) as the orange-colored edge (1-simplex) connecting nodes (0-simplices)  $R_1$  and  $R_2$ . Analogously, note that both  $A_5$  and  $A_6$  taxiway nodes are connected to  $T$ ,  $R_3$ , and  $R_4$ . Thus, these are both non-empty intersections of two covers:

$$\bigcap_{i \in \{T, R_3, R_4\}} U_i = \{A_5, A_6\} \neq \emptyset. \tag{19}$$

These intersections both correspond to the 2-simplex  $\sigma_{T, R_3, R_4}$ , found in Fig. 21(c) as the purple-shaded triangle (2-simplex) connecting nodes  $T$ ,  $R_3$ , and  $R_4$ . Every non-empty intersection in the intersection graph can be mapped to corresponding  $k$ -simplices in the nerve complex representation. A compression of information can be seen in translating from the complete airport runway-taxiway configuration diagram to the nerve complex representation: The complete airport runway-taxiway configuration diagram conveys geographic and geometric information as well as information regarding connectivities, whereas only topologically non-trivial connectivities are preserved in the nerve complex representation.

As a comparison, we give the complete airport runway-taxiway configuration diagram, intersection graph, and nerve complex representation for ORD during its low-capacity scenario as well in Fig. 22. Note the drastic changes in the intersection graph as well as the corresponding nerve complex representation. This particular instance was exacerbated by an ongoing winter storm impacting ORD, causing many taxiways to be impassable due to snow drifts. In particular, the loss of the 2-simplex within the nerve complex represents a large loss in airport surface connectivity.

6.2.4. Loss of maximal simplex indicate large changes in airfield topology

For each of the five case study airports, we construct the associated intersection graphs for the full airport runway-taxiway

**Table 2**  
Summary of computational results from the airport surface nerve complex case study.

Airport	GMT Date (m/d/y)	Declared rates	Runway configurations	0-simplices	1-simplices	2-simplices	3-simplices
ATL HIGH	1400Z 11/11/2018	AAR: 132 ADR: 118	Arr: 8L, 9R, 10 Dep: 8R, 9L	T, 8L, 9R, 10, 8R, 9L	{T, 8L}, {8L, 8R}, {T, 8R}, {T, 9L}, {9L, 9R}, {T, 9R}, {9R, 10}	{T, 8L, 8R}, {T, 9L, 9R}	$\emptyset$
ATL LOW	1200Z 11/5/2018	AAR: 72 ADR: 80	Arr: 8L, 9R, 10 Dep: 8R, 9L	T, 8L, 9R, 10, 8R, 9L	{T, 8L}, {8L, 8R}, {T, 8R}, {T, 9L}, {9L, 9R}, {9R, 10}	{T, 8L, 8R}	$\emptyset$
LAX HIGH	1400Z 11/4/2018	AAR: 74 ADR: 74	Arr: 24R, 25L Dep: 24L, 25R	T, 24R, 25L, 24L, 25R	{24R, 24L}, {24L, 25R}, {24L, T}, {T, 25R}, {T, 25L}, {25R, 25L}	{T, 25R, 25L}	$\emptyset$
LAX LOW	0900Z 11/8/2018	AAR: 29 ADR: 29	Arr: 6R Dep: 25R	T, 6R, 25R	{T, 6R}, {6R, 25R}, {25R, T}	$\emptyset$	$\emptyset$
ORD HIGH	0500Z 11/1/2018	AAR: 114 ADR: 114	Arr: 27L, 27R, 28C Dep: 22L, 28R	T, 27L, 27R, 28C, 22L, 28R	{27R, 27L}, {27L, T}, {27L, 28R}, {22L, T}, {22L, 28R}, {28R, T}, {28R, 28C}, {T, 28C}	{T, 28R, 28C}	$\emptyset$
ORD LOW	1200Z 11/26/2018	AAR: 16 ADR: 20	Arr: 4R Dep: 4L	T, 4R, 4L	{T, 4L}, {T, 4R}	$\emptyset$	$\emptyset$
DFW HIGH	1200Z 11/10/2018	AAR: 84 ADR: 100	Arr: 31R, 35R, 36L Dep: 31L, 35L, 36R	T, 31R, 35R, 36L, 31L, 35L, 36R	{31L, 36L}, {36L, T}, {T, 31L}, {31L, 36R}, {T, 36R}, {36L, 36R}, {35L, 31R}, {31R, 35R}, {35R, 35L}, {35L, T}, {31R, T}, {35R, T}	{T, 31L, 36L}, {T, 31L, 36R}, {31L, 36L, 36R}, {T, 36L, 36R}, {T, 35L, 31R}, {T, 35L, 35R}, {T, 31R, 35R}, {35L, 31R, 35R}	{T, 31L, 36L, 36R}, {T, 35L, 31R, 35R}
DFW LOW	0200Z 11/23/2018	AAR: 60 ADR: 80	Arr: 17L, 18R Dep: 17R, 18L	T, 17L, 18R, 17R, 18L	{T, 18R}, {18R, 18L}, {T, 18L}, {T, 17R}, {17R, 17L}, {T, 17L}	{T, 18R, 18L}, {T, 17R, 17L}	$\emptyset$
DEN HIGH	2300Z 11/8/2018	AAR: 152 ADR: 114	Arr: 16L, 16R, 35L, 35R Dep: 8, 25, 34L, 34R	T, 16L, 16R, 35L, 35R, 8, 25, 34L, 34R	{T, 16R/34L}, {T, 25}, {T, 35L}, {T, 35R}, {T, 34R/16L}, {T, 8}, {34R/16L, 8}	{T, 8, 34R/16L}	$\emptyset$
DEN LOW	0600Z 11/24/2018	AAR: 32 ADR: 32	Arr: 34R, 35L, 35R Dep: 34L	T, 34R, 35L, 35R, 34L	{T, 34L}, {T, 34R}, {T, 35L}, {T, 35R}	$\emptyset$	$\emptyset$

configuration diagram, as well as the corresponding nerve complex representations. We list the date, declared capacities, runway configurations, and  $k$ -simplex records for each of the case study airports under high- and low-capacity scenarios in Table 2. Note that some nerve complex representations do not have higher-order  $k$ -simplices beyond 1-simplices – in these cases, “ $\emptyset$ ” is recorded to indicate that there were no 3- or 4-fold intersections of covers to map to a 2- or 3-simplex.

From Table 2, we see that for our five case study airports, the loss of the *maximal* simplices can be an indication that there was a large loss in surface connectivity during low-capacity scenarios. A maximal simplex within a simplicial complex – in this case, the nerve complex – is the “largest” existing  $k$ -simplex, where  $k$  is maximized across all existing  $k$ -simplices. For example, the maximal simplex in the ORD high-capacity scenario would be the 2-simplex given by  $\sigma_{T,28R,28C}$ , whereas the maximal simplex in the DFW high-capacity scenario would be a 3-simplex given by  $\sigma_{T,31L,36L,36R}$  or  $\sigma_{T,35L,31R,35R}$ . We see that in the case of every case study airport with the exception of ATL, the maximal simplex that exists in the high-capacity scenario disappears in the low-capacity scenario. The loss of a 2-simplex means the loss of a taxiway node that facilitated up to three connections between ramp and runway assets, and the loss of a 3-simplex means the loss of a taxiway node that facilitated up to four connections. In Section 7 we discuss the operational and managerial implications of this case study, along with the other three reviewed studies, all of which leverage topological methods to examine aviation data sets.

## 7. Potential operational and managerial insights

The reviewed past studies and our own computational case study in Sections 5 and 6, respectively, bridge the gap between theory (Section 4) and aviation applications. We now discuss in-depth the connections between the results from Sections 5 and 6 to aviation, specifically in terms of operational and managerial insights. Furthermore, we explicitly map holes and other abstract topological features to concrete physical processes occurring within the specific aviation domain that TDA/PH was utilized in. In doing so, we hope to emphasize the widespread applicability and flexibility of these topological methods, as well as highlight the fact that TDA/PH serve to complement and augment existing data analysis methods, not succeed them. This discussion will lead into a presentation of future research avenues for TDA/PH within the aviation domain in Section 8.

### 7.1. Topological features map to trajectory characteristics

Understanding terminal airspace constraints and situations that may result in excessive in-air holding patterns and missed approaches is critical for air traffic flow managers as well as airspace planners. The application of TDA/PH in detecting individual cyclic trajectories indicative of holding patterns and missed approaches reviewed in Section 5.1 could provide decision support insight regarding terminal airspace conditions that may result in an increased number of these anomalous trajectories. Data resulting from using TDA/PH to locate anomalous trajectories is of importance to airspace planners and airport managers with more long-term, strategic goals as well. For example, when airports plan for capital projects such as the extension or construction of a new runway, the terminal airspace also needs to be readjusted to plan for new arrival and departure routes given the new surface infrastructure (Li and Ryerson, 2017). Any new arrival and departure routes within the terminal airspace should not interfere with sectors of the airspace generally reserved for in-air holdings and missed approach clearances. The topological approach in Section 5.1 useful for identifying such anomalous operations within the terminal airspace complement other methods such as trajectory clustering (Murça and Hansman, 2018) and learning anomalous trajectories via machine learning (Li et al., 2015), but with advantages such as a decreased reliance on parameter tuning and training. Anomalous cyclic trajectories with holes are natural objects identifiable via TDA/PH. Furthermore, there is an increased tolerance for noise in the data set, as noisy, *non-persistent* data can be filtered out via PH (Christ, 2014, 2017).

### 7.2. Topological features map to robustness in air transportation networks

Paying homage to the flexibility of TDA/PH and simplicial complex representation, the same algebraic-topological foundations of holes and connectivities can also be applied to a data set of airport nodes within the US NAS (Section 5.2), a completely different data context from a point cloud representation of aircraft trajectories (Section 5.1). The approach leveraged by the reviewed study in Section 5.2 combines traditional graph-theoretic measures with algebraic-topological measures based on the ratio of various Betti numbers, mapping both to metrics describing network robustness. Here, topological holes indicate dyadic relationships between airports with no higher-order connectivities, whereas a filled-in simplex (*i.e.* no holes) indicate non-trivial, higher-order relationships. The ability to include higher-order relationships gives a more nuanced and realistic network robustness model than graph-based approaches (Lordan et al., 2014; Wei et al., 2014). From a tactical, operational standpoint, if it is known *a priori* that a particular clique of airports often propagate delays between themselves, then traffic management initiatives and airspace flow programs could be better tailored to target such specific cliques and prevent a larger cascade of delays and cancellations. On the other hand, from a managerial and strategic standpoint, identification of such cliques within the air transportation network could lead to revised re-routing policies that locally increase network robustness against perturbations such as convective weather. An example of where such a managerial amendment could occur within the US NAS is at the level of “plays” within the “playbook” of Severe Weather Avoidance Plans (SWAPs) (Federal Aviation Administration, 2018b).

Drop in AAR (top) and ADR (bottom) – Top 30 Airports (Nov 2018)

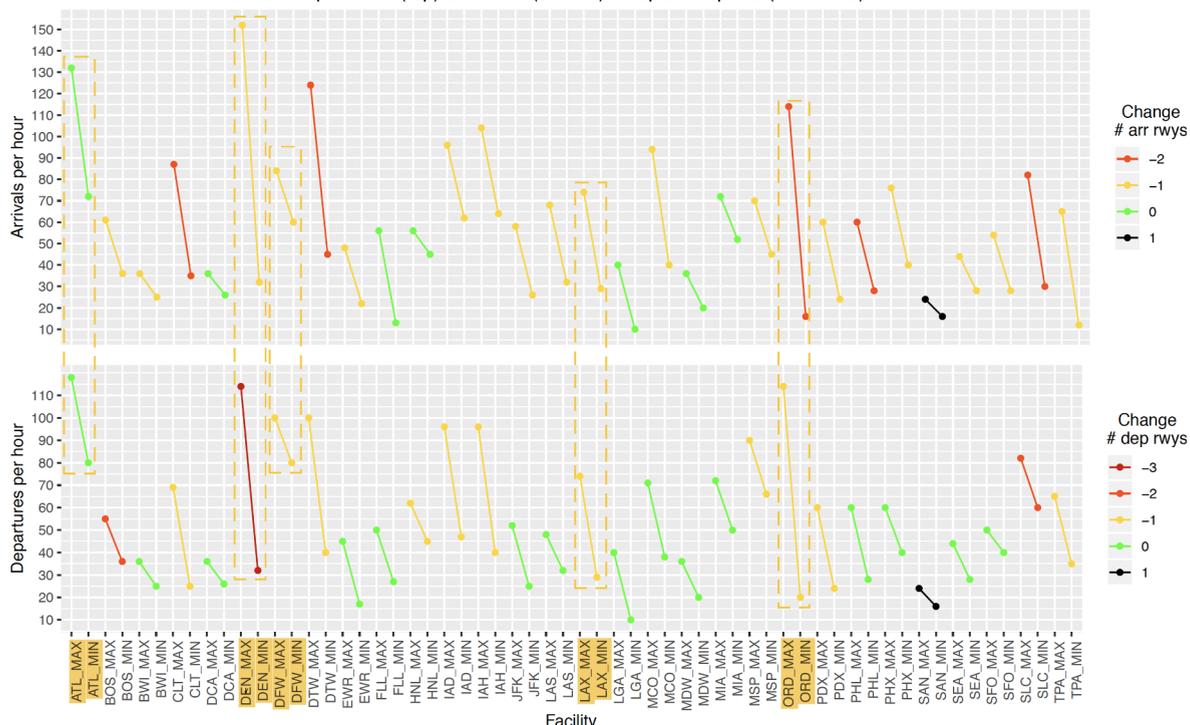


Fig. 23. Change in ADR and AAR at each airport between high- and low-capacity hours during November 2018, with case study airports highlighted. Runway asset losses are also noted.

### 7.3. Topological features map to airspace partitions

Elements of topology and geometric analysis are used to inform urban air mobility (UAM) airspace planning scenarios in the research reviewed in Section 5.3. The need to design corridors embedded within the low-altitude, high-density airspace that mitigate UAV-on-infrastructure collision risks is crucial towards the future adoption and expansion of UAM. The underlying data set in this study arises from urban features such as buildings and other public utilities infrastructures, with topological  $\alpha$ -shapes partitioning and demarcating safe versus unsafe UAV operating airspace regions. The airspace partitions and geofences procured via such a topological analysis could help inform future city planners and Municipal Planning Organizations (MPOs) regarding how to best policy the use of UAVs for commercial, personal, and governmental purposes within an urban environment. Designing and optimizing topologically efficient urban airspace corridors for UAV operations have the potential to lead to new avenues in logistics and traffic flow management research (Sunil et al., 2018).

### 7.4. Topological features map to airport surface characteristics

It is well-known that the arrival and departure capacities at an airport is largely a function of the service rate of the runway (de Neufville et al., 2013); the arrival and departure service rates of airport runways play an important role in many models ranging from simple deterministic time-based models to complex network queueing models (Badrinath et al., 2019). In Fig. 23, we plot separately the AAR and ADR for each airport, colored by the drop in the number of arrival and departure runways, respectively. From Fig. 23, we can see that while the number of runways in service does dictate the capacity to some degree, there are variations that cannot be explained by the number of runways alone. For example, ATL experiences a noticeable drop in both AAR and ADR, but the number of arrival and departure runways in use during the high- and low-capacity scenarios remained the same. On the other hand, airports such as DEN and ORD not only suffered large drops in capacities, but also the loss of several departure and arrival runways between high- and low-capacity scenarios.

Our computational case study in Section 6 depicts yet another aviation data context within which TDA/PH can be applied: Airport surface taxiway and runway reachability and connectivity. By representing the full airport runway-taxiway configuration as a simplicial complex, we were able to not only give a sparse representation that preserves the important connectivities, but also glean insight regarding how airport surface operations map to easier-to-understand topological shapes. In particular, while the AAR and ADR plot in Fig. 23 confirms the relationship between number of active runways and the airport capacity, the results from the computation case study (Table 2) suggest that one could simply keep track of the maximal simplex within the simplicial complex representation of the full airport runway-taxiway configuration and get a sense of how well-connected and reachable various airfield

assets are. Furthermore, a simplified representation such as the intersection graph or the nerve complex (Figs. 21 and 22) could be more helpful to airport surface managers and ground controllers making high-level decisions regarding taxiway flow patterns, as these representations retain only the important reachability indicators between airfield assets.

## 8. Future research directions

We have only begun to scratch the surface of applying topological methods such as TDA/PH to examine aviation data sets. Equipped with an understanding of the fundamental algebraic-topological notions that enable TDA/PH and a preview of how various topological features could map to pertinent features and characteristics previously hidden within aviation data, we dedicate this penultimate section to a sampling of proposals for future research avenues combining TDA/PH and aviation.

### 8.1. Topological persistence and embedded strategy complexes of airport surface assets

Our computational case study from Section 6 provides a framework for translating airport surface networks into nerve complexes amenable to various theorems and techniques found in algebraic topology and TDA/PH. For example, PH could be used to track topological features of an airport's surface that persist through a wide range of capacity scenarios. The identification of transient and persistent topological features may aid in formalizing the process of finding “hot spots” in airport surface operations and traffic flows. Furthermore, the topology of the nerve complexes themselves may be interesting, particularly if the nerve complexes are viewed as a *strategy complex* (Erdmann, 2010). Depending on the similarity between our airport surface nerve complexes and certain classes of strategy complexes, there are guarantees on which “strategies” are possible and which are not. In the aviation context, a strategy on a strategy complex could map to, for example, a pre-defined class of taxiway flows.

### 8.2. Homotopic and homological path planning in aviation

Another area in aviation data science that presents opportunities for applying TDA/PH is origin-destination trajectory generation. Given an origin-destination pair  $(o_i, d_i)$ , it can be thought of as points within a target space  $S \subset \mathbb{R}^2$ . One trajectory  $\tau_1: [0, \mathbb{T}] \rightarrow S$  travels from  $o_i$  to  $d_i$  via time parameter  $\lambda \in [0, \mathbb{T}]$  by sending  $\tau_1(0) = o_i$  and  $\tau_1(\mathbb{T}) = d_i$ . We can consider another trajectory  $\tau_2: [0, \mathbb{U}] \rightarrow S$  that also traverses from  $o_i$  to  $d_i$ , but via a different trek. Imagine if we can *continuously deform* trajectory  $\tau_1$  to  $\tau_2$  – if this is possible, then  $\tau_1$  is *fixed end point homotopic* to  $\tau_2$ , and there is a *fixed end point homotopy*  $\mathcal{F}_{1,2}: [0, \mathbb{T}] \times [0, \mathbb{U}] \rightarrow S$  such that any other trajectories from  $o_i$  to  $d_i$  created during the smooth deformation process from  $\tau_1$  to  $\tau_2$  is reachable by  $u \in [0, \mathbb{U}]$ , with  $\lambda \in [0, \mathbb{T}]$  traveling along this path. Explicitly,  $\mathcal{F}_{1,2}$  can be written as:

$$\mathcal{F}_{1,2}: \left[ 0, \mathbb{T} \right] \times \left[ 0, \mathbb{U} \right] \rightarrow S: \begin{cases} \mathcal{F}_{1,2}(\lambda, 0) = \tau_1(\lambda), & \mathcal{F}_{1,2}(0, u) = \tau_1(0) = \tau_2(0) \\ \mathcal{F}_{1,2}(\lambda, \mathbb{U}) = \tau_2(\lambda), & \mathcal{F}_{1,2}(\mathbb{T}, u) = \tau_1(\mathbb{T}) = \tau_2(\mathbb{T}) \end{cases} \quad (20)$$

Homotopy theory can become very complex and rich when considering exotic topologies, but for the purposes of aviation, the workspace is typically just the surface of the Earth, with various discontinuities that represent airspace features such as no-fly zones and convective weather which will affect the resultant homotopies. Vidosavljevic et al. (2017) presents one possible application of homotopy theory in aviation trajectory generation.

### 8.3. Topological trajectory and fuel flow analysis

Aviation trajectory data sets lend itself well to topological methods due to the high-dimensional and sparse nature of the data. Another active research area in aviation that may benefit from an exploration using TDA/PH is aircraft taxi-in and taxi-out fuel flow analysis, particularly using raw data from on-board flight data recorders (Chati and Balakrishnan, 2013). An aircraft taxiing to and from the gate at a busy airport may encounter runway and ramp queues, resulting in a repeated pattern of stop-and-go thrust inputs as the aircraft taxis behind another aircraft within the queue itself (Levine and Gao, 2007). This pattern of thrust inputs and corresponding fuel flow rates can be modeled periodically in time, and this periodicity can be mapped to cycles detectable through topological means. Similar to the previous study reviewed in Section 5.1, the taxiway network can be analyzed from this viewpoint, potentially identifying time periods when anomalous trajectories of fuel flow patterns appear. Such analysis may lead to the development of better optimization models and system control inputs to reduce taxi-in and taxi-out emissions and delays.

### 8.4. Aviation sensor coverage problems and higher-order network comparisons

Finally, Dłotko et al. (2012) demonstrated the use of TDA/PH in sensor networks and sensor coverage. Applications in aviation research include characterizing the expansion and coverage of Automatic Dependent Surveillance-Broadcast (ADS-B) and Multi-lateration (MLAT) capabilities, particularly in regions of the world with poor primary radar coverage. Zooming back out to the perspective of the entire air transportation network, TDA/PH – and in particular homological algebra – could be used to classify how similar one higher-order network is compared to another. A higher-order network in aviation could resemble something similar to Fig. 2, where multiple layers of simplicial complexes are used to represent different numerical and categorical data related to the NAS. This application stems from previous work that investigates using PH to establish approximations of network distances within

higher-order graphs (Huang and Ribeiro, 2017). Given the highly interconnected nature of the global aviation system coupled with a plethora of data regarding the edges and nodes of the network, it would be interesting to apply some of the results from Huang and Ribeiro (2017) to differentiate between various multi-layered networks in the aviation domain using topological features and homological algebra.

## 9. Summary and conclusion

In addressing the call for new data science methodologies to help tackle the oncoming big data era of aviation research, we first discussed the influx of aviation data from a variety of sources. This has led to a surge in data-driven aviation research. However, some aviation data sets are high-dimensional, highly-connected, and sparse. As classical data analytics may fail to uncover higher-order structures and relationships in these data sets, new data science tools that provide a more global perspective are needed.

As TDA/PH may be able to unearth global topological features within aviation data sets, we introduce TDA/PH and propose its use to further aviation data science research. These topological features could be translated into new benchmarks, models, and applications. To translate TDA/PH methods to air transportation, one must have both the necessary aviation domain knowledge and an understanding of how TDA/PH works; only then can one recognize what TDA/PH can and cannot reveal. In order to make the algebraic-topological foundations of TDA/PH more readily accessible to aviation researchers, we dedicated a large portion of this work to an introduction of the fundamental notions, definitions, and constructs behind TDA/PH (Section 4). We created an instructive example drawn from a sample data set containing aviation geography, airspace, and operational features. This sample data set is used throughout the manuscript, from constructing simplicial complexes out of data to introducing simplicial homology.

Equipped with an understanding of the theory underlying TDA/PH, we then bridge the gap between the abstractions provided by algebraic topology to actual aviation applications with potential managerial and operational insights. In particular, we motivate the applicability of TDA/PH in aviation data science research by reviewing past studies that incorporated a topological approach to examine aviation data sets (Section 5). The contrast in the types of data sets used in the past studies, ranging from aircraft trajectories to geography-based data for constructing UAV geofences, showcases the flexibility of TDA/PH. We add to this growing collection of studies with our own computational case study exploring nerve complex representations of airport surface networks (Section 6), where we showed that maximal simplices map to large losses in airport surface connectivity. We elaborate on this translation between topological features and aviation-specific features in Section 7, as well as describe managerial and operational insights derivable from these topological data science approaches. To close our foray into TDA/PH and aviation, we suggest in Section 8 an abundance of potential future research avenues for topological methods in aviation data science. We hope that the motivation for and introduction to TDA/PH we have provided stimulates an interest in examining aviation data sets through these topological lenses.

## Acknowledgements

This work was supported in part by NSF CPS1739505. M. Z. Li was also supported by an NSF Graduate Research Fellowship. We would also like to thank the anonymous reviewers whose comments, feedback, and suggestions greatly improved this manuscript.

## References

- Air Transport Action Group, 2018. Aviation: Benefits Beyond Borders. Tech. rep., Air Transport Action Group.
- Ayasdi, 2015. Why Topological Data Analysis Works. <https://www.ayasdi.com/blog/bigdata/why-topological-data-analysis-works/>.
- Badrinath, S., Li, M.Z., Balakrishnan, H., 2019. Integrated surface-airspace model of airport departures. *J. Guid. Control Dyn.* 1–15. <https://doi.org/10.2514/1.G003964>.
- Bauer, Ulrich, Kerber, Michael, Reininghaus, 2017. PHAT (Persistent Homology Algorithm Toolbox). <https://bitbucket.org/phant-code/phant>.
- Belkoura, S., Cook, A., Peña, J.M., Zanin, M., 2016. On the multi-dimensionality and sampling of air transport networks. *Transp. Res. Part E: Logist. Transp. Rev.* 94, 95–109. <https://doi.org/10.1016/j.tre.2016.07.013>. <http://www.sciencedirect.com/science/article/pii/S1366554516301946>.
- Bhattacharya, S., Lipsky, D., Ghrist, R., Kumar, V., 2013. Invariants for homology classes with application to optimal search and planning problem in robotics. *Ann. Math. Artif. Intell.* 67 (3), 251–281. <https://doi.org/10.1007/s10472-013-9357-7>.
- Bhattacharya, S., Ghrist, R., Kumar, V., 2015. Persistent homology for path planning in uncertain environments. *IEEE Trans. Robot.* 31 (3), 578–590. <https://doi.org/10.1109/TRO.2015.2412051>.
- Bhattacharya, S., Ghrist, R., Kumar, V., 2014. Multi-robot coverage and exploration on riemannian manifolds with boundaries. *Int. J. Robot. Res.* 33 (1), 113–137. <https://doi.org/10.1177/0278364913507324>. arXiv:<https://doi.org/10.1177/0278364913507324>.
- Burmester, G., Ma, H., Steinmetz, D., Hartmann, S., 2018. *Big Data and Data Analytics in Aviation*. Springer International Publishing, Cham, pp. 55–65.
- Carlsson, G., 2009. Topology and data. *Bull. Amer. Math. Soc.* 46, 255–308. <https://doi.org/10.1090/S0273-0979-09-01249-X>.
- Chati, Y.S., Balakrishnan, H., 2013. Aircraft engine performance study using flight data recorder archives. In: AIAA AVIATION Forum. American Institute of Aeronautics and Astronautics, <https://doi.org/10.2514/6.2013-4414>.
- Chatterjee, A., Flores, H., Sen, S., Hasan, K.S., Mani, A., 2017. Distributed location detection algorithms using IoT for commercial aviation. In: 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 126–131. <https://doi.org/10.1109/ICRCICN.2017.8234493>.
- Cho, J., Yoon, Y., 2018. How to assess the capacity of urban airspace: a topological approach using keep-in and keep-out geofence. *Transp. Res. Part C: Emerg. Technol.* 92, 137–149. <https://doi.org/10.1016/j.tre.2018.05.001>. <http://www.sciencedirect.com/science/article/pii/S0968090X18305850>.
- Comitz, P., Ayhan, S., Gerberick, G., Pesce, J., Bliensner, S., 2013. Predictive analytics with aviation big data. In: 2013 Integrated Communications, Navigation and Surveillance Conference (ICNS), pp. 1–35. <https://doi.org/10.1109/ICNSurv.2013.6548645>.
- Cook, A., Blom, H.A., Lillo, F., Mantegna, R.N., Micciché, S., Rivas, D., Vázquez, R., Zanin, M., 2015. Applying complexity science to air traffic management. *J. Air Transp. Manage.* 42, 149–158. <https://doi.org/10.1016/j.jairtraman.2014.09.011>. <http://www.sciencedirect.com/science/article/pii/S0969699714001331>.
- de Neufville, R., Odoni, A.R., Belobaba, P.P., Reynolds, T.G., 2013. *Airport Systems, Second Edition: Planning, Design and Management*, second ed. McGraw-Hill Education.
- Djokic, J., Lorenz, B., Fricke, H., 2010. Air traffic control complexity as workload driver. *Transp. Res. Part C: Emerg. Technol.* 18 (6), 930–936. <https://doi.org/10.1016/j.tre.2010.03.005>. special issue on Transportation Simulation Advances in Air Transportation Research.. <http://www.sciencedirect.com/science/article/10.1016/j.tre.2010.03.005>.

- pii/S0968090X10000318.
- Dłotko, P., Ghrist, R., Juda, M., Mrozek, M., 2012. Distributed computation of coverage in sensor networks by homological methods. *Appl. Algebra Eng. Commun. Comput.* 23 (1), 29–58. <https://doi.org/10.1007/s00200-012-0167-7>.
- Du, W.-B., Zhou, X.-L., Lordan, O., Wang, Z., Zhao, C., Zhu, Y.-B., 2016. Analysis of the chinese airline network as multi-layer networks. *Transp. Res. Part E: Logist. Transp. Rev.* 89, 108–116. <https://doi.org/10.1016/j.tre.2016.03.009>. <http://www.sciencedirect.com/science/article/pii/S1366554515300521>.
- Du, W.-B., Zhang, M.-Y., Zhang, Y., Cao, X.-B., Zhang, J., 2018. Delay causality network in air transport systems. *Transp. Res. Part E: Logist. Transp. Rev.* 118, 466–476. <https://doi.org/10.1016/j.tre.2018.08.014>. <http://www.sciencedirect.com/science/article/pii/S1366554518301042>.
- Durak, U., 2018. *Flight 4.0: The Changing Technology Landscape of Aeronautics*. Springer International Publishing, Cham, pp. 3–13.
- Edelsbrunner, H., Harer, J.L., 2009. *Computational Topology: An Introduction*. American Mathematical Society.
- Erdmann, M., 2010. On the topology of discrete strategies. *Int. J. Robot. Res.* 29 (7), 855–896. <https://doi.org/10.1177/0278364909354133>.
- Fasy, B.T., Kim, J., Lecci, F., Maria, C., Millman, D.L., Rouvreau, V., 2014. Introduction to the R package TDA. <https://cran.r-project.org/web/packages/TDA/vignettes/article.pdf>.
- Federal Aviation Administration, 2016. Passenger boarding (enplanement) and all-cargo data for U.S. airports (5 2016). [https://www.faa.gov/airports/planning\\_capacity/passenger\\_allcargo\\_stats/passenger/f](https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/f).
- Federal Aviation Administration, 2018. Aviation system performance metrics (aspm) (23 2018). <https://aspm.faa.gov/apm/sys/main.asp>.
- Federal Aviation Administration, 2018. Air Traffic Control System Command Center – National Severe Weather Playbook. <https://www.fly.faa.gov/PLAYBOOK/pbindex.html>.
- FlightGlobal, 2019. Five Key Themes For Big Data In Aerospace. <https://www.flightglobal.com/news/articles/analysis-five-key-themes-for-big-data-in-aerospace-456869/>.
- Fugacci, U., Iuricich, F., Floriani, L.D., 2014. Efficient computation of simplicial homology through acyclic matching. In: 2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, 2014, pp. 587–593. <https://doi.org/10.1109/SYNASC.2014.84>.
- Ghrist, R., 2014. *Elementary Applied Topology*. CreateSpace.
- Ghrist, R., 2017. Homological algebra and data.
- Giusti, C., Ghrist, R., Bassett, D.S., 2016. Two's company, three (or more) is a simplex. *J. Comput. Neurosci.* 41 (1), 1–14. <https://doi.org/10.1007/s10827-016-0608-6>.
- Govindan, K., Cheng, T., Mishra, N., Shukla, N., 2018. Big data analytics and application for logistics and supply chain management. *Transp. Res. Part E: Logist. Transp. Rev.* 114, 343–349. <https://doi.org/10.1016/j.tre.2018.03.011>. <http://www.sciencedirect.com/science/article/pii/S1366554518302606>.
- Guépet, J., Briant, O., Gayon, J.-P., Acuna-Agost, R., 2017. Integration of aircraft ground movements and runway operations. *Transp. Res. Part E: Logist. Transp. Rev.* 104, 131–149. <https://doi.org/10.1016/j.tre.2017.05.002>. <http://www.sciencedirect.com/science/article/pii/S1366554516308237>.
- Hatcher, A., 2002. *Algebraic Topology*. Cambridge University Press.
- Henselman, Gregory, Eirene, 2018. <http://gregoryhenselman.org/eirene/index.html>.
- Hongyong, W., Ruiying, W., Yifei, Z., 2015. Analysis of topological characteristics in air traffic situation networks. *Proc. Inst. Mech. Eng. Part G: J. Aerosp. Eng.* 229 (13), 2497–2505. <https://doi.org/10.1177/0954410015578482>.
- Huang, W., Ribeiro, A., 2017. Persistent homology lower bounds on high-order network distances. *IEEE Trans. Signal Process.* 65 (2), 319–334. <https://doi.org/10.1109/TSP.2016.2620963>.
- Kasten, J., Hotz, I., Noack, B.R., Hege, H.-C., 2011. *On the Extraction of Long-living Features in Unsteady Fluid Flows*. Springer, Berlin, Heidelberg, pp. 115–126.
- Keller, P., Kreylos, O., Vanco, M., Hering-Bertram, M., Cowgill, E.S., Kellogg, L.H., Hamann, B., Hagen, H., 2011. *Extracting and Visualizing Structural Features in Environmental Point Cloud LiDaR Data Sets*. Springer, Berlin, Heidelberg, pp. 179–192.
- Khadilkar, H., Balakrishnan, H., 2014. Network congestion control of airport surface operations. *J. Guid. Control Dyn.* 37 (3), 933–940. <https://doi.org/10.2514/1.57850>.
- Kotegawa, T., Fry, D., DeLaurentis, D., Puchaty, E., 2014. Impact of service network topology on air transportation efficiency. *Transp. Res. Part C: Emerg. Technol.* 40, 231–250. <https://doi.org/10.1016/j.tre.2013.11.016>. <http://www.sciencedirect.com/science/article/pii/S0968090X13002441>.
- Levine, B.S., Gao, H.O., 2007. Aircraft taxi-out emissions at congested hub airports and implications for aviation emissions reduction in the united states. *Transp. Res. Rec.: J. Transp. Res. Board*. <https://trid.trb.org/view/802435>.
- Li, M.Z., Ryerson, M.S., 2017. A data-driven approach to modeling high-density terminal areas: a scenario analysis of the new Beijing, China airspace. *Chin. J. Aeronaut.* 30 (2), 538–553. <https://doi.org/10.1016/j.cja.2016.12.030>. <http://www.sciencedirect.com/science/article/pii/S1000936117300213>.
- Li, M.Z., Ryerson, M.S., 2018. Detection of individual anomalous arrival trajectories within the terminal airspace using persistent homology. In: 8th International Conference for Research in Air Transportation, 2018, pp. 4.
- Li, M.Z., Ryerson, M.S., 2019. Reviewing the data of aviation research data: diversity, availability, tractability, applicability, and sources. *J. Air Transp. Manage.* 75, 111–130. <https://doi.org/10.1016/j.jairtraman.2018.12.004>. <http://www.sciencedirect.com/science/article/pii/S096969971830379X>.
- Li, L., Das, S., John Hansman, R., Palacios, R., Srivastava, A.N., 2015. Analysis of flight data using clustering techniques for detecting abnormal operations. *J. Aerosp. Inf. Syst.* 12 (9), 587–598. <https://doi.org/10.2514/1.1010329>.
- LiveATC, 2018. LiveATC.net. <https://www.liveatc.net/>.
- Lordan, O., Sallan, J.M., Simo, P., Gonzalez-Prieto, D., 2014. Robustness of the air transport network. *Transp. Res. Part E: Logist. Transp. Rev.* 68, 155–163. <https://doi.org/10.1016/j.tre.2014.05.011>. <http://www.sciencedirect.com/science/article/pii/S1366554514000805>.
- Mott, J.H., Bullock, D.M., 2018. Estimation of aircraft operations at airports using mode-c signal strength information. *IEEE Trans. Intell. Transp. Syst.* 19 (3), 677–686. <https://doi.org/10.1109/TITS.2017.2700764>.
- Murça, M.C.R., Hansman, R.J., 2018. Identification, characterization, and prediction of traffic flow patterns in multi-airport systems. *IEEE Trans. Intell. Transp. Syst.* 1–14. <https://doi.org/10.1109/TITS.2018.2833452>.
- Oracle, 2012. Oracle Airline Data Model. <https://www.oracle.com/technetwork/database/options/airlines-data-model/airlines-data-model-bus-overview-1451727.pdf>.
- Ramanathan, R., Bar-Noy, A., Basu, P., Johnson, M., Ren, W., Swami, A., Zhao, Q., 2011. Beyond graphs: Capturing groups in networks. In: *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. <https://doi.org/10.1109/INFOCOMW.2011.5928935>.
- Rocha, L.E., 2017. Dynamics of air transport networks: a review from a complex systems perspective. *Chin. J. Aeronaut.* 30 (2), 469–478. <https://doi.org/10.1016/j.cja.2016.12.029>. <http://www.sciencedirect.com/science/article/pii/S1000936117300171>.
- Schaar, D., Sherry, L., 2010. Analysis of airport stakeholders. In: 2010 Integrated Communications, Navigation, and Surveillance Conference Proceedings, pp. J4-1–J4-17. <https://doi.org/10.1109/ICNSURV.2010.5503233>.
- Simić, T.K., Babić, O., 2015. Airport traffic complexity and environment efficiency metrics for evaluation of atm measures. *J. Air Transp. Manage.* 42, 260–271. <https://doi.org/10.1016/j.jairtraman.2014.11.008>. <http://www.sciencedirect.com/science/article/pii/S0969699714001471>.
- Sizemore, A.E., Phillips-Cremins, J., Ghrist, R., Bassett, D.S., 2018. The importance of the whole: topological data analysis for the network neuroscientist. <https://arxiv.org/abs/1806.05167>.
- Sunil, E., Ellerbroek, J., Hoekstra, J.M., Maas, J., 2018. Three-dimensional conflict count models for unstructured and layered airspace designs. *Transp. Res. Part C: Emerg. Technol.* 95, 295–319. <https://doi.org/10.1016/j.tre.2018.05.031>. <http://www.sciencedirect.com/science/article/pii/S0968090X1830771X>.
- Szymczak, A., 2011. *A Categorical Approach to Contour, Split and Join Trees with Application to Airway Segmentation*. Springer, Berlin, Heidelberg, pp. 205–216.
- Vidosavljević, A., Delahaye, D., Tosic, V., 2017. *Homotopy Route Generation Model for Robust Trajectory Planning*. Springer, Japan, Tokyo, pp. 69–88.
- Wei, P., Chen, L., Sun, D., 2014. Algebraic connectivity maximization of an air transportation network: the flight routes' addition/deletion problem. *Transp. Res. Part E: Logist. Transp. Rev.* 61, 13–27. <https://doi.org/10.1016/j.tre.2013.10.008>. <http://www.sciencedirect.com/science/article/pii/S1366554513001750>.
- Xia, K., 2018. Persistent homology analysis of ion aggregations and hydrogen-bonding networks. *Phys. Chem. Chem. Phys.* 20, 13448–13460. <https://doi.org/10.1039/C8CP01552J>.

- Yousefi, A., Zadeh, A.N., 2013. Dynamic allocation and benefit assessment of nextgen flow corridors. *Transp. Res. Part C: Emerg. Technol.* 33, 297–310. <https://doi.org/10.1016/j.trc.2012.04.016>. <http://www.sciencedirect.com/science/article/pii/S0968090X12000666>.
- Yu, B., Guo, Z., Asian, S., Wang, H., Chen, G., 2019. Flight delay prediction for commercial air transport: a deep learning approach. *Transp. Res. Part E: Logist. Transp. Rev.* 125, 203–221. <https://doi.org/10.1016/j.tre.2019.03.013>. <http://www.sciencedirect.com/science/article/pii/S1366554518311979>.
- Zanin, M., Lillo, F., 2013. Modelling the air transport with complex networks: a short review. *Eur. Phys. J. Spec. Top.* 215 (1), 5–21. <https://doi.org/10.1140/epjst/e2013-01711-9>.
- Zhou, A., Maletić, S., Zhao, Y., 2018. Robustness and percolation of holes in complex networks. *Phys. A: Stat. Mech. Appl.* 502, 459–468. <https://doi.org/10.1016/j.physa.2018.02.149>. <http://www.sciencedirect.com/science/article/pii/S0378437118302188>.
- Zhou, Y., Wang, J., Huang, G.Q., 2019. Efficiency and robustness of weighted air transport networks. *Transp. Res. Part E: Logist. Transp. Rev.* 122, 14–26. <https://doi.org/10.1016/j.tre.2018.11.008>. <http://www.sciencedirect.com/science/article/pii/S1366554518311165>.