

Utah State University

DigitalCommons@USU

---

All Graduate Theses and Dissertations

Graduate Studies

---

12-2018

## Sheaf Theory as a Foundation for Heterogeneous Data Fusion

Seyed M-H Mansourbeigi  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>



Part of the [Computer Sciences Commons](#)

---

### Recommended Citation

Mansourbeigi, Seyed M-H, "Sheaf Theory as a Foundation for Heterogeneous Data Fusion" (2018). *All Graduate Theses and Dissertations*. 7363.  
<https://digitalcommons.usu.edu/etd/7363>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



SHEAF THEORY AS A FOUNDATION FOR HETEROGENEOUS DATA FUSION

by

Seyed M-H Mansourbeigi

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

---

Heng-Da Cheng, Ph.D.  
Major Professor

---

Minghui Jiang, Ph.D.  
Committee Member

---

Haitao Wang, Ph.D.  
Committee Member

---

Zhaohu Nie, Ph.D.  
Committee Member

---

Jia Zhao, Ph.D.  
Committee Member

---

Laurens H. Smith, Ph.D.  
Interim Vice President for Research and  
Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2018

Copyright © Seyed M-H Mansourbeigi 2018

All Rights Reserved

## ABSTRACT

Sheaf Theory as a Foundation for Heterogeneous Data Fusion

by

Seyed M-H Mansourbeigi, Doctor of Philosophy

Utah State University, 2018

Major Professor: Heng-Da Cheng, Ph.D.  
Department: Computer Science

This dissertation proposes an effective geometric and topological approach in computational science for the study, analysis, and fusion of temporal and spatial heterogeneous data obtained from multiple sources, where the schema, availability and quality vary.

The approach provides tools for translating heterogeneous data into common language to enable data fusion. The utilization of this methodology studies the behavior of the system based on the failure in data exchange, detection of noise in the system and recognition of the redundant or complimentary sensors.

This method consists of objects, namely simplices that are attached to make a simplicial complex. Data sources are represented by the 0-dimensional simplices and interactions among two and more sensors are represented by higher dimensional simplices. Analysis of data, encoding and translating heterogeneous data into common language, is modeled by stalks. The fusion of data extracted from multiple sensors is modeled by a sheaf.

Homology groups help the interpretation of the behavior of the system based on its potentiality to exchange data. This interpretation helps to detect possible voids in data exchange.

Applications of the constructed methodology are brought into practice via two case studies: one from wildfire threat monitoring and the other from the air traffic monitoring.

A comparison between the sheaf theory methodology and the alternative methods is described to present another proof for the validity of the sheaf theory method. It is seen that the sheaf theory method has less computational complexity in both space and time.

(117 pages)

## PUBLIC ABSTRACT

## Sheaf Theory as a Foundation for Heterogeneous Data Fusion

Seyed M-H Mansourbeigi

A major impediment to scientific progress in many fields is the inability to make sense of the huge amounts of data that have been collected via experiment or computer simulation. This dissertation provides tools to visualize, represent, and analyze the collection of sensors and data all at once in a single combinatorial geometric object. Encoding and translating heterogeneous data into common language are modeled by supporting objects. In this methodology, the behavior of the system based on the detection of noise in the system, possible failure in data exchange and recognition of the redundant or complimentary sensors are studied via some related geometric objects.

Applications of the constructed methodology are described by two case studies: one from wildfire threat monitoring and the other from air traffic monitoring. Both cases are distributed (spatial and temporal) information systems. The systems deal with temporal and spatial fusion of heterogeneous data obtained from multiple sources, where the schema, availability and quality vary. The behavior of both systems is explained thoroughly in terms of the detection of the failure in the systems and the recognition of the redundant and complimentary sensors.

A comparison between the methodology in this dissertation and the alternative methods is described to further verify the validity of the sheaf theory method. It is seen that the method has less computational complexity in both space and time.

To my parents, my brother, my wife and my son

## ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Dr. Heng-Da Cheng, for his strong support during my Ph.D. studies. I would never have been able to complete this dissertation without his guidance. I am very grateful to my committee members, Dr. Minghui Jiang, Dr. Haitao Wang, Dr. Zhaohu Nie, and Dr. Jia Zhao, for their great comments and advice which improved the quality of this research. I would also like to thank my parents, my brother, my wife and my son for their invaluable support, understanding, and love. I would also like to thank Genie Hanson, Vicki Anderson and Cora Price for their advice and help that eased any troubles during the time I worked at the CS department and also special thanks to Erika Beckstrand from school of graduate studies. I give special thanks to my friends, and colleagues for their encouragement, moral support, and patience as I worked my way from the initial proposal writing to this final document.

Finally, I am grateful to all advisors and friends at the Computer Science, Electrical Engineering, Physics, and Mathematics departments at Utah State University. Their help, support and friendship are really helpful during my PhD program.

Seyed M-H Mansourbeigi



## CONTENTS

	Page
ABSTRACT.....	iii
PUBLIC ABSTRACT .....	v
ACKNOWLEDGMENTS .....	vii
LIST OF TABLES.....	xi
LIST OF FIGURES .....	xii
CHAPTER	
1 INTRODUCTION .....	1
1.1 The Motivation: "swimming in sensors and drowning in data".....	1
1.2 The Method .....	2
1.3 Why Topological Approach.....	3
1.4 In What Cases is the Topological Approach Better? .....	7
1.4.1 Simplicial Complex Model vs. Graph Theory.....	7
1.4.2 Advantages of the Sheaf Theory Approach.....	9
1.5 Advantages of the Alternative Approach .....	9
1.6 Prior Work History.....	10
1.7 The Research Contribution in this Dissertation .....	10
1.8 Outline.....	12
2 SIMPLICIAL COMPLEXES, HOMOLOGY AND DISCRETIZING DATA.....	13
2.1 Simplicial Complex.....	13
2.1.1 Simplicial Complex as a Poset .....	15
2.1.2 Constructing a Simplicial Complex from a Topological Space .....	17
2.1.3 Alexandroff's Definition .....	17
2.2 Simplicial Homology .....	18
2.2.1 Computation of Homology Groups Algorithm .....	24
2.2.2 The Python Program for Homology Computation .....	25
2.3 Simplicial Complex Beyond the Graph Structure for Data Representation.....	28
2.4 Summary .....	30
3 SHEAVES, DATA FUSION, COHOMOLOGY AND DATA ANALYSIS .....	31
3.1 Cellular Sheaves of Vector Spaces .....	31

3.2	Sheaf Cohomology.....	35
3.3	Pseudocodes for Computation of Cellular Sheaf .....	37
3.3.1	Step 1 (part 1): Find the Vector Space of 0-Simplices .....	38
3.3.2	Step 1 (part 2) Assignment of Stalks .....	38
3.3.3	Step 2: Restriction Maps.....	39
3.4	Mathematical Foundation for Sheaf Cohomology and Data Analysis.....	40
3.5	Summary .....	43
4	APPLICATIONS .....	44
4.1	Part 1: Example of Wildfire Threat Monitoring .....	44
4.1.1	The Construction of the Simplicial Complex .....	46
4.1.2	Homology calculation at time $t = t_0$ .....	46
4.1.3	The Sheaf Construction .....	50
4.1.4	Sheaf Cohomology Calculation at Time $t = t_0$ .....	51
4.1.5	Time Changes from $t=t_0$ to $t= t_1$ .....	57
4.1.6	The Sheaf Construction .....	59
4.1.7	Discussion.....	60
4.2	Part 2: Example of Air Traffic Monitoring .....	61
4.2.1	The Construction of the Simplicial Complex .....	63
4.2.2	The Sheaf Construction .....	64
4.2.3	Homology and Sheaf Cohomology .....	65
4.2.4	Time Changes from $t=t_0$ to $t= t_1$ .....	66
4.2.5	Feedback from the Example.....	69
4.3	Summary .....	69
5	ALTERNATIVE SOLUTION.....	70
5.1	Solving the Fire Monitoring with Alternative Tools.....	70
5.1.1	The Pseudocode to Confirm the Fire in Each Cell .....	72
5.2	Summary .....	75
6	NOISE AND INCONSISTENCY .....	76
6.1	Consistency Radius .....	76
6.2	How to Achieve the Desirable Consistency Radius.....	79
6.3	Summary .....	80

7 CONCLUSION AND FUTURE WORK .....	81
7.1 The Feedback Process .....	81
7.2 Summary and Future Work .....	81
REFERENCES .....	82
APPENDICES .....	88
VITAE.....	98

## LIST OF TABLES

Table		Page
2.1	The Co-authorship.....	29
4.1	Sensors and Duplication Numbers ( $t = t_0$ ) for Wildfire Monitoring.....	45
4.2	The Heterogeneous Data ( $t = t_0$ ) for Wildfire Monitoring.....	45
4.3	Sensors and Duplication Numbers, Wildfire Monitoring ( $t = t_1$ ).....	58
4.4	The Heterogeneous Data for Wildfire Monitoring ( $t = t_1$ ).....	58
4.5	Sensors and Duplication Numbers for Air Traffic Monitoring ( $t = t_0$ ).....	63
4.6	The Heterogeneous Data for Air Traffic Monitoring ( $t = t_0$ ).....	63
4.7	Sensors and Duplication Numbers for Air Traffic Monitoring ( $t = t_1$ ).....	66
4.8	The Heterogeneous Data for Air Traffic Monitoring ( $t = t_1$ ).....	66
5.1	Signals from Satellite Cameras, $\text{dim} = H \times G$ , complexity = $O(n^2)$ .....	71
5.2	Signals from CO2 Detectors, $\text{dim} = H \times G$ , complexity = $O(n^2)$ .....	71
5.3	Signals from IR Detectors, $\text{dim} = H \times G$ , complexity = $O(n^2)$ .....	72
5.4	Signals from Flame Detectors, $\text{dim} = H \times G$ , complexity = $O(n^2)$ .....	72
5.5	Existence of Fire Based on Sensors Measurements Time $t = t_0$ .....	73
5.6	Existence of Fire Based on Reporting from Fire Department Time $t = t_0$ .....	74

## LIST OF FIGURES

Figure		Page
1.1	Reference data on the length (in centimeters) and weight (in grams) for Atlantic Ocean rockfish of several sizes (regression line).....	4
1.2	Scientific datasets are becoming more dynamic, requiring new mathematical techniques on par with the invention of calculus.....	4
1.3	In the point set cluster the k-median objective (left) minimizes the sum of distances from points to their representative data points. The k-means objective (right) minimizes the average of the squared Euclidean distances of all points within a cluster .....	5
1.4	Schematics of magnetically confined plasmas in (a) tokamaks; and (b) stellarator configurations. In the tokamak, the rotational transform of a helical magnetic field is formed by a toroidal field generated by external coils together with a poloidal field generated by the plasma current. In the stellarator, the twisting field is produced entirely by external non-axisymmetric coils .....	6
1.5	In patient and genotype networks each node represents a single or a group of patients with the significant similarity based on their clinical features. The edge connected with nodes indicates the nodes have shared patients. The red color represents the enrichment for patients with females, and blue color represents the enrichment for males .....	6
1.6	The geometric realization of a dataset by a simplicial complex. The blue balls are of a fixed scale-parameter (radius $r$ ). Two points are connected if they are within $r$ of each other. Connections between more than two points create higher dimensional simplices.....	7
1.7	Road map towards the creation of the modeling: From set theory to topology, homology and sheaf cohomology.....	11
2.1	A 3-simplex as the polytope convex hull of three geometrically independent points $a_0, a_1, a_2$ . The simplices are represented by their vertices.....	14
2.2	Simplices of dimensions zero, one and two (top). A simplicial complex of dimension two (left) and a collection of simplices (right) which do not comprise a simplicial complex.....	14
2.3	A 2-dimensional simplicial complex (left). The poset representation of the simplicial complex (right).....	16

2.4	The upper set $U$ represents an open set in the Alexandroff topology for the simplicial complex.....	16
2.5	A cover $\mathcal{O} = \{a, b, c, d, e, f\}$ of 6 sets with labels for each cover set (left) and its nerve complex (right).....	17
2.6	The simplicial complexes $K_1$ with two components and one hole (left) and $K_2$ with one component and no holes (right).....	20
2.7	The simplicial complex model for the co-authorship. Vertices represent the authors. The edges and triangle represent the multiple co-authorship.....	29
3.1	An example of a simplicial complex (left), the associated sheaf $F$ (middle and right). Inclusions of the faces are shown by upward arrows.....	32
3.2	A network of academic institutions that might share information about a student (left), and a sheaf representing associated information about a single student (right).....	33
4.1	Simplicial complex model with oriented simplices for the wildfire threat monitoring at time $t=t_0$ .....	46
4.2	Sheaf, stalks and restriction maps associated with the simplicial complex for the wildfire threat monitoring example at time $t= t_0$ .....	51
4.3	Simplicial complex model with oriented simplices for the wildfire threat monitoring at time $t=t_1$ .....	58
4.4	Sheaf, stalks and restriction maps associated with the simplicial complex for the wildfire threat monitoring for time $t=t_1$ .....	60
4.5	Air route and traffic control centers in the United States and its territories....	62
4.6	An example of air traffic monitoring system including air traffic control tower, air route traffic control center, and terminal radar approach control...	62
4.7	Simplicial complex model with oriented simplices for the air traffic monitoring at time $t= t_0$ .....	64
4.8	Sheaf, stalks and restriction maps associated with the simplicial complex for the air traffic monitoring example at time $t= t_0$ .....	65
4.9	Simplicial complex model with oriented simplices for the air traffic monitoring at time $t= t_1$ .....	67

4.10	Sheaf, stalks and restriction maps associated with the simplicial complex for the air traffic monitoring example at time $t = t_1$ .....	68
6.1	Sheaf of vector spaces on the partial order set associated with the example of wildfire threat monitoring system. The diagram commutes.....	77
6.2	Relating to the example of wildfire threat monitoring a global section is an assignment that is consistent with restrictions.....	78
6.3	Relating to the example of wildfire threat monitoring due to noise some assignments are not consistent. They are partially consistent.....	78
6.4	Consistency radius is the maximum distance between the value in a stalk and the values propagated along the restrictions.....	79
7.1	The feedback process.....	81

## CHAPTER 1

### INTRODUCTION

#### **1.1 The Motivation: "swimming in sensors and drowning in data"**

Data integration is the combination of technical processes to combine data from multiple sources into meaningful and valuable information. In this dissertation, the meaning of "shape" is construed as the way to think about data, with the shape of data being what carries the meaning. The objective of this dissertation is to study the shape of data. The combination of algebraic topology and sheaf theory is necessary in a quantitative study of "shape." The concept of topology is based on the fact that data has shape and the shape matters.

The framework for heterogeneous data integration should accurately represent the locally valid datasets in which the data types vary. It should also provide a common canonical language for heterogeneous datasets and multiple source interactions. There are classes of methods that study the characteristic of diversity in data types. For example, the Bayesian method is based on the data obtained from a probability distribution of specific parameter values [1]. In statistical methods for topological data analysis, it is assumed that a sample of data is drawn randomly from some distribution [2]. However, these methods tend to rely on the homogeneity of information sources to obtain strong theoretical results. Sheaf theory extends the reach of these methods by explaining that the most robust aspects of networks tend to be topological in nature. The theory provides the means for detecting topological features and, therefore, identifies relationships between information sources that present hazards to Bayesian reasoning [3]. Moreover, in many situations data often have a specific shape that escapes the reach of methods to provide required information.



The sheaf theory extends the robustness aspects of heterogeneous data integration by reasoning about the topological nature of data and rigorously extracts features of interest from heterogeneous data resources.

Another major issue in the operation and maintenance of sensor collections with various types is their high cost. The sheaf theory approach that is utilized in this dissertation can detect easily which type and what number of sensors are redundant and which sensors can be decommissioned in order to reduce the cost of operation and maintenance.

## 1.2 The Method

Geometry and topology are natural tools for analyzing massive amounts of data. The connection between topology and large amounts of data offers huge opportunities, as well as challenges, to big data communities. A survey on bringing together state-of-the-art research results on geometrical and topological methods for big data is shown in [4].

This dissertation presents a conceptual technique that addresses the problem of modeling and reasoning about temporal and spatial fusion of heterogeneous data from multiple sources, in which the schema, availability, quantity, and quality vary. The main idea is to present more predictive methods to study heterogeneous data using the topological data analysis approach.

Topological data analysis (TDA) is a collection of powerful tools that can quantify the shape and structure in data in order to answer questions from the data domain. It is done by representing some aspects of data in a simplified topological structure. An investigation

towards a representation of some aspects of the shape of data in a simplified form for study is shown in [5].

In this dissertation, the topological data analysis techniques are borrowed from algebraic topology and algebraic geometry. The topology approach reflects interactions among data sources, and the sheaf theoretic approach reflects integration of heterogeneous data types. Sheaf theory is a new tool for topological data analysis to track data. It is a way of attaching data to a topological space to manage heterogeneous data with various quality, quantity, schema, and availability. A sheaf may be regarded as a system of observations on a topological space, in which consistent local observations (sections) can be uniquely pasted together to provide a global observation (section).

### 1.3 Why Topological Approach

The topology and sheaf theory approach is a solid, powerful theoretical foundation to the analysis of datasets that are complex, high-dimensional, heterogeneous, incomplete, and noisy. Extracting such information is in general challenging. To explain how extracting information from datasets is related to the concept of "shape," the following examples are provided. More examples can be found in [6] and [7].

Data have the shape of a line as shown in Figure 1.1. In this example a straight line fits the given data quite well (linear regression). The figure illustrates how some variables are related to other variables (prediction). It gives the qualitative information that the weight-variable varies directly with the length-variable, and it helps to predict one of

the variables with reasonable accuracy if the value of the other variable is known. The idea is that the shape of data as a line allows the user to extract useful information from it.

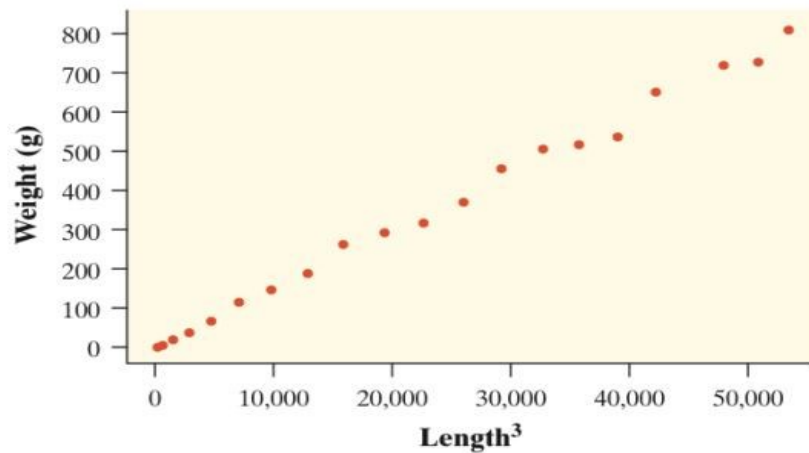


Figure 1.1. Reference data on the length (in centimeters) and weight (in grams) for Atlantic Ocean rockfish of several sizes (regression line) [8].

Data do not always cooperate and fit along a line. Consider the following example. The shape of data in Figure 1.2 is like the capital letter “Y.”

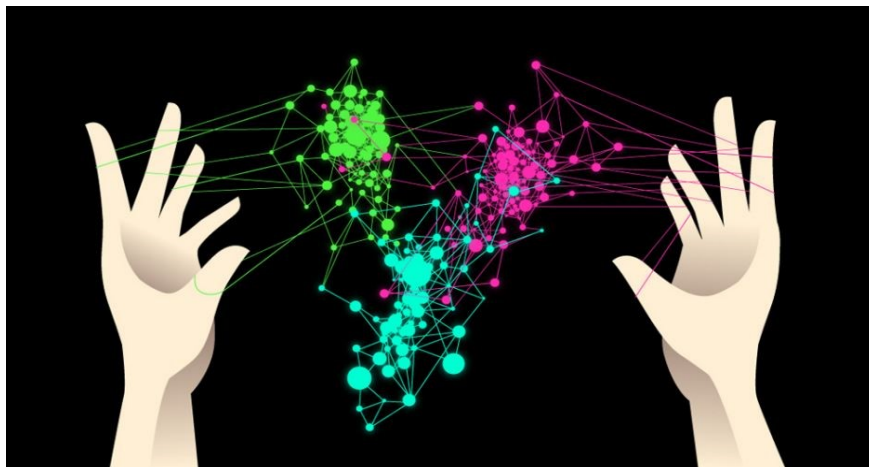


Figure 1.2. Scientific datasets are becoming more dynamic, requiring new mathematical techniques on par with the invention of calculus [9].

The problem is that there are an infinite variety of different possible shapes, a large number of which occur in real datasets. There are analytic ways to deal with these shapes of data. Data may be cut into pieces and each cluster can be dealt with separately. Figure 1.3 shows clusters of data.

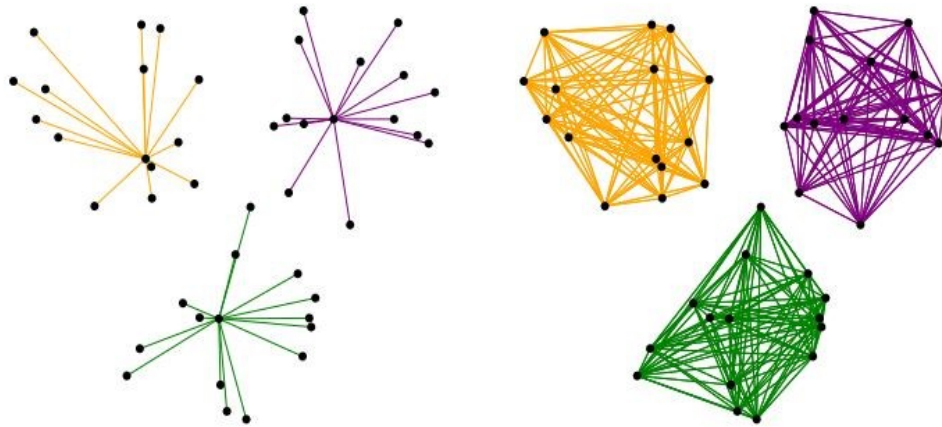


Figure 1.3. In the point set cluster the k-median objective (left) minimizes the sum of distances from points to their representative data points. The k-means objective (right) minimizes the average of the squared Euclidean distances of all points within a cluster [10].

At certain times, data must be dealt with as a whole. The idea is to produce representations of data and to show all data at once. What happens when data representation is neither linear nor cluster? It can have any shape. As an example in magnetic configurations for a toroidal plasma confinement system, the plasmas are confined by a magnetic field. An equilibrium between the plasma pressure and the magnetic forces creates the configuration shown in Figure 1.4.

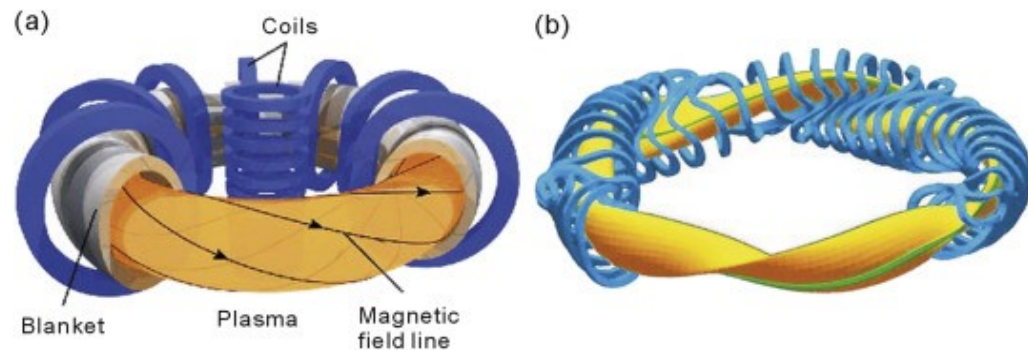


Figure 1.4. Schematics of magnetically confined plasmas in (a) tokamaks; and (b) stellarator configurations. In the tokamak, the rotational transform of a helical magnetic field is formed by a toroidal field generated by external coils together with a poloidal field generated by the plasma current. In the stellarator, the twisting field is produced entirely by external non-axisymmetric coils [11].

Sometimes data are more complex. See Figures 1.5 and 1.6 as examples of complex data.

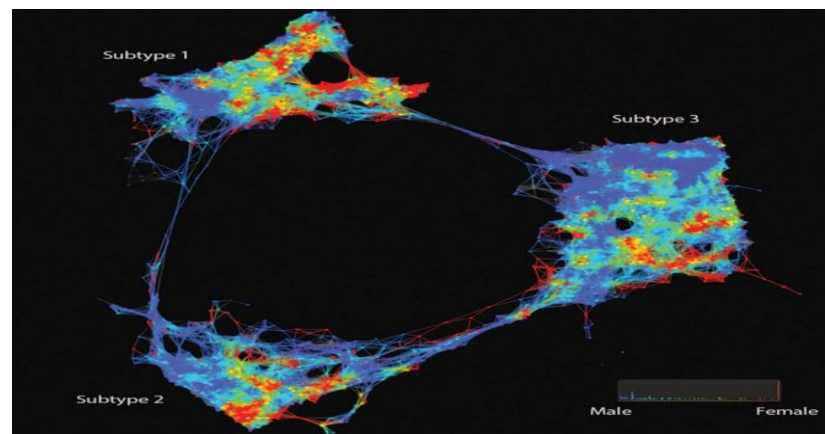


Figure 1.5. In patient and genotype networks each node represents a single or a group of patients with the significant similarity based on their clinical features. The edge connected with nodes indicates the nodes have shared patients. The red color represents the enrichment for patients with females, and blue color represents the enrichment for males [12].

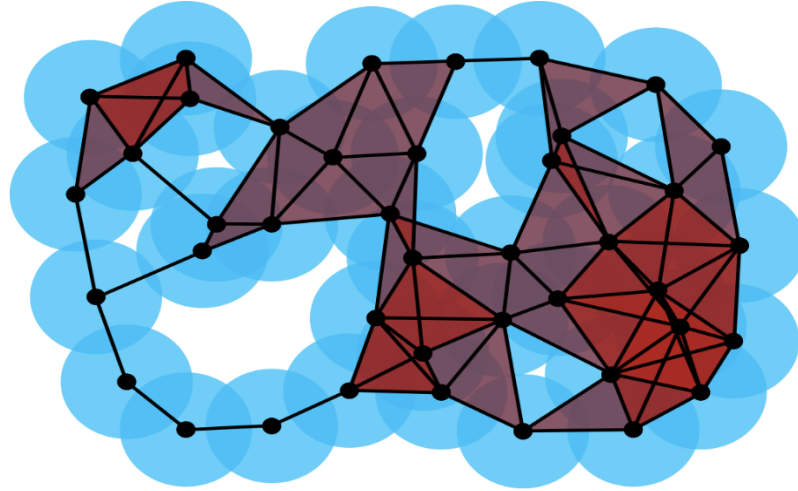


Figure 1.6. The geometric realization of a dataset by a simplicial complex. The blue balls are of a fixed scale-parameter (radius  $r$ ). Two points are connected if they are within  $r$  of each other. Connections between more than two points create higher dimensional simplices [13].

Methods are required to deal with complex data to visualize and describe a high-dimensional data shape. The above examples support the idea that data have shape and that shape matters. More examples of the applications of topological methods to study complex high-dimensional datasets by extracting shapes (patterns) and obtaining insights about them are shown in the list of references [14] and [15].

## 1.4 In What Cases is the Topological Approach Better?

### 1.4.1 Simplicial Complex Model vs. Graph Theory

The conventional method of handling data and describing a dataset is to build a graph in which the vertex set is the collection of points in data space, and each point is possibly a collection of data. Two vertices are connected by an edge. In fact, a

combinatorial theory of interactions between at most two datasets can be constructed using only graph theory (an example is the graph-based data fusion in [16]).

What is the problem with a graph model? There are cases involving data sources that encompass more than two interactions. To deal with these cases, one must apply combinatorial topology, a higher-dimensional version of graph theory. One approach will be a combinatorial model in which all possible interactions between multiple sources are captured using topological notions. In fact simplicial complexes are possible generalizations of graph-theoretic modeling, as shown in Figure 1.6.

There are methods to construct a simplicial complex from a graph. According to [17], topological framework enables the multifaceted approach. An application of algebraic topology and simplicial complex modeling for characterizing interactions between multiple sources obtained from opinion space of a group of individuals can be found in [18].

The cluster analysis method works with a set of subjects as statistical data units described by a set of homogeneous (of the same type) variables. The technique concerns exploratory multivariate data analysis for finding a clustering structure on a dataset [19]. The key idea is to represent all possible data at some time as a single, static, combinatorial geometric object, called a simplicial complex. It is done by providing methods which produce combinatorial representations of the data. There are many sources of high-dimensional data that are inherently structured, but the structure is difficult to conceptualize. In this dissertation, the motivation is to organize, associate, and connect multidimensional data to qualitatively understand the global content.

### **1.4.2 Advantages of the Sheaf Theory Approach**

When the type and the number of sensors increase, there is a need to develop systems to establish situational awareness of events based on multiple real-time information feeds. A sensor is an instrument that generates a quantified signal to a generic information process and returns a stream of observations, either direct measurements, derived measurements, or the output of an analytic process [20].

When translation of heterogeneous data into common language is required, data fusion techniques are extensively employed in multi-sensor environments with the aim of fusing and aggregating data obtained from different sensors. Modeling consistency between observations and encoding the interactions among heterogeneous information sources to integrate data requires a stronger tool. In this situation, the sheaf theory approach is the viable solution. A review of data fusion techniques may be found in [21], [22] and [23]. In short, sheaves are used to analyze dissimilar data types.

### **1.5 Advantages of the Alternative Approach**

When an event is reported by single-type smart sensors, the alternative approach potentially gives a shorter solution. In this case, the measurements of the event detected by multiple sensors are homogeneous; as a consequence, the event is reported based on the measurement that is compared with a threshold. The alternative approach is better when there are no heterogeneous data and no complex problems.



## 1.6 Prior Work History

The sheaf theory was developed in mathematics to study the relationships between local and global phenomena, and has been applied in algebraic geometry, differential geometry, analysis, and even logic. A broad class of presheaf models was proposed for a general calculus by Cattani and Winskel [24]. They studied presheaf models for concurrent computation. Application of sheaf theory in computer science has a long historical track. The basic technique towards the adoption of a topological view of data structures was applied to the derivation of pattern matching algorithm [25]. They applied the sheaf theory to characterize the extension of the occurrence relation. As a foundation for the behavior of concurrent processes Ehrich, Goguen and Sernadas [26] applied the sheaf model. Goguen [27] utilized concepts from the category theory and modeled objects by sheaves. The motivation in this dissertation is inspired by recent applications of sheaf theory in computer science and software engineering. These applications can be found in [28] for distributed systems and in [29] for understanding the behaviors of the networks. In this dissertation, sheaves are representations for the behavior of the sensors. Moreover the data structure is represented by the simplicial complex topological model.

## 1.7 The Research Contribution in this Dissertation

The research for this dissertation yields an explanation of the topological data analysis modeling technique together with an illustration of data integration from multiple sources that differ in terms of their schemas, granularity, and quality. For instance, an example of a wildfire detection application that gathers heterogeneous data from a

designated area is explained. In it, the area is covered by different types of sensors for measuring temperature, intensity, fire size, and smoke. The sensors are online or offline at different times and locations dynamically or are permanently disabled in some cases. This modeling technique is used to capture essential characteristics of the wildfire application and to answer questions such as:

- a) Do the sensors provide sufficient information to track a real fire, even when some of the sensors may go offline?
- b) Which types of sensors are redundant or complimentary?
- c) Is there any failure in data exchange in the spatial or temporal dimension?

Both approaches are used to answer the questions. Basically, the two approaches (algebraic topology and algebraic geometry) include creations of the data structures and algorithms for computation of homology and sheaf cohomology. Homology interprets the temporal and spatial shape of data interaction and cohomology interprets the data analysis. The road map for the two approaches is shown in Figure 1.7.

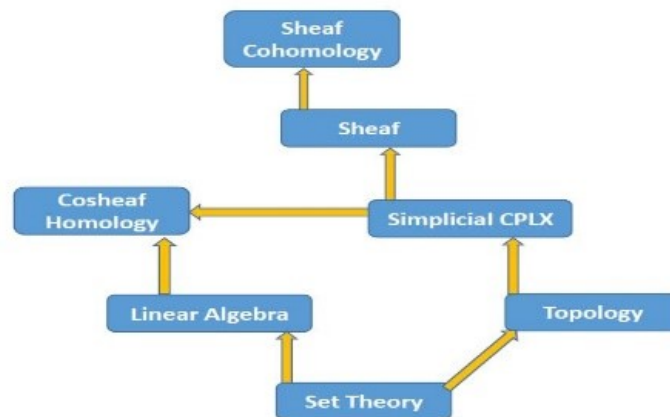


Figure 1.7. Road map towards the creation of the modeling: From set theory to topology, homology and sheaf cohomology.

## 1.8 Outline

This dissertation is organized as follows based on the main contribution of the research:

Chapters 1, and 2 introduce the topological approach to identify and study the system through the shape of data and data sources.

Chapter 3 describes the mathematical foundation by presenting the required definitions to bring the information of the system into a mathematical language. The validity of the method is verified by the main theorem that brings about a necessary and sufficient condition for a sensor to be significant.

Chapter 4 is dedicated to the applications of the methodology that has been constructed in the previous chapters to the two case studies: wildfire and air-traffic monitoring.

Chapter 5 presents the comparison between the sheaf theory methodology and the alternative methods.

Chapter 6 studies the case in the presence of noise in the system that results in the sheafification of the system to be disturbed.

Chapter 7 proposes opportunities for future work.

## CHAPTER 2

### SIMPLICIAL COMPLEXES, HOMOLOGY AND DISCRETIZING DATA

This chapter defines the concept of simplicial complex and continues with a comprehensive explanation of the constructions of simplicial complex modeling for data analysis. The topological approach modeling is applied to reflect interactions among data sources. Essentially, it includes creation of the data structures and algorithms for computation of homology in temporal and spatial shape of data interaction.

#### 2.1 Simplicial Complex

The following definitions are extracted from [30] and [31].

**Definition 2.1** A set of  $n$  points in Euclidean space ( $\mathbb{R}^k$ ) is geometrically independent if the points do not belong to any  $(n-2)$ -dimensional hyperplane.

**Definition 2.2** An  $n$ -simplex is the closed polytope convex hull of  $(n+1)$  geometrically independent ordered set of points. An  $n$ -dimensional simplex is denoted by  $S^n$ . A 0-simplex  $S^0$  is a vertex, a 1-simplex  $S^1$  is an edge, a 2-simplex  $S^2$  is a triangle, and so forth. A  $d$ -simplex  $S^d$  is a proper face of a  $t$ -simplex  $S^t$  if  $d < t$  and each vertex of  $S^d$  is a vertex of  $S^t$ . Consequently  $S^t$  is called a proper coface of  $S^d$ . For simplicity the  $n$ -simplex  $S^n$  with  $(n+1)$  vertex points  $\{a_0, a_1, a_2, \dots, a_n\}$  is denoted by  $S^n = a_0 a_1 a_2 \dots a_n$ . It is shown in Figure 2.1.

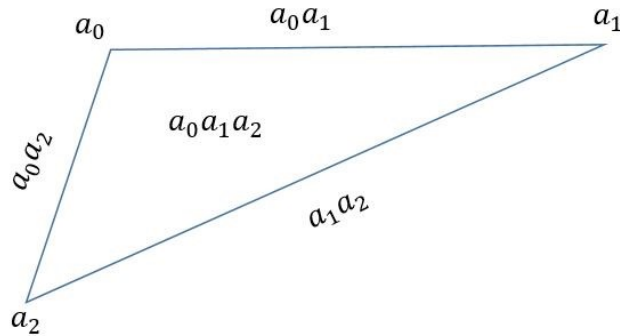


Figure 2.1. A 3-simplex as the polytope convex hull of three geometrically independent points  $a_0, a_1, a_2$ . The simplices are represented by their vertices.

Definition 2.3 A simplicial complex  $\mathbf{K}$  is a set of simplices satisfying the following conditions:

- 1- Any face of a simplex in  $\mathbf{K}$  also belongs to  $\mathbf{K}$ .
- 2- The intersection of any two simplices in  $\mathbf{K}$  is either empty or is another simplex.

The dimension of a simplicial complex is the maximum of the dimensions of its simplices.

See Figure 2.2.

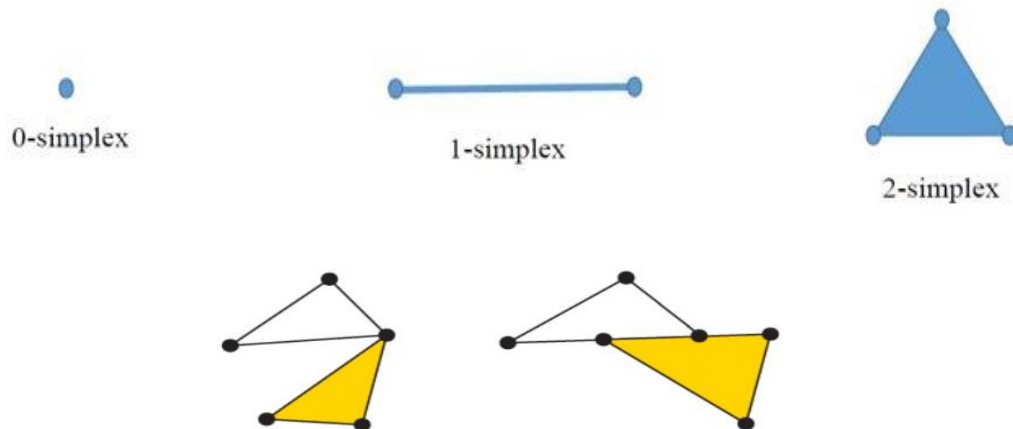


Figure 2.2. Simplices of dimensions zero, one and two (top). A simplicial complex of dimension two (left) and a collection of simplices (right) which do not comprise a simplicial complex [32].

**Definition 2.4** For the two simplices  $S^q$  and  $S^{q+1}$ , with  $\dim S^{q+1} = \dim S^q + 1$ , the incidence number which is denoted by  $[S^{q+1}:S^q]$  is defined to be 0 (if  $S^q$  is not a face of  $S^{q+1}$ ) and  $(-1)^n$  (if by deleting the  $n^{\text{th}}$  vertex of the simplex  $S^{q+1}$ , the simplex  $S^q$  is obtained). In short if  $b = S^{q+1}$  and  $a = S^q$ ,

$$[b : a] = \begin{cases} 0 & \text{if } a \text{ is not a face of } b \\ (-1)^n & \text{if you delete the } n^{\text{th}} \text{ vertex of } b \text{ to get } a \end{cases} \quad (1)$$

For example if  $S = a_0 a_1 a_2$  and  $T = a_1 a_5$  and  $U = a_1 a_2$  then:

$[S : T] = [a_0 a_1 a_2 : a_1 a_5] = 0$  and  $[S : U] = [a_0 a_1 a_2 : a_1 a_2] = (-1)^0 = 1$ . Similarly  $[a_0 a_1 a_2 : a_0 a_2] = (-1)^1 = -1$ .

As shown in the next two subsections, simplicial complexes inherit extra algebraic structures. The structures will be important in the data analysis in the coming chapters.

### 2.1.1 Simplicial Complex as a Poset

A relation, " $\leq$ ", is a partial order on a set  $S$  if it has reflexive property ( $a \leq a$  for all  $a$  in  $S$ ), antisymmetric property ( $a \leq b$  and  $b \leq a$  implies  $a = b$ ), and transitive property ( $a \leq b$  and  $b \leq c$  implies  $a \leq c$ ).

A partially ordered set (a poset) is a set together with a partial order on it. A simplicial complex carries a poset structure, in which the elements of the poset are simplices and the partial order is obtained by the face/coface relationship. This relationship is denoted by  $<$ . If  $S$  is a face of  $C$  then write  $S < C$ . See Figure 2.3.

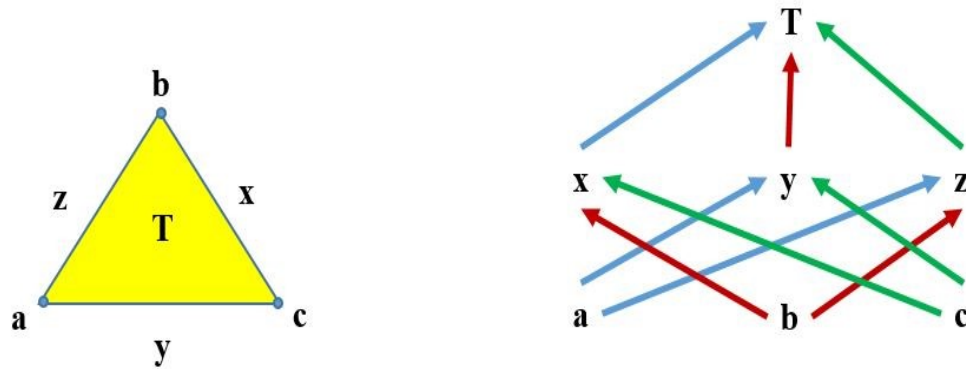


Figure 2.3. A 2-dimensional simplicial complex (left). The poset representation of the simplicial complex (right).

A topology (Alexandroff topology [33]) is associated with the poset of faces in the simplicial complex  $\mathbf{K}$ . The open sets in this topology are defined by the upper sets in the following way:

In the simplicial complex  $\mathbf{K}$ , a subset  $U \subseteq \mathbf{K}$  is open if and only if it satisfies the following condition:

For the two simplices  $S$  and  $C$  in  $\mathbf{K}$ , if  $S \in U$  and  $S < C$  then  $C \in U$ . See Figure 2.4.

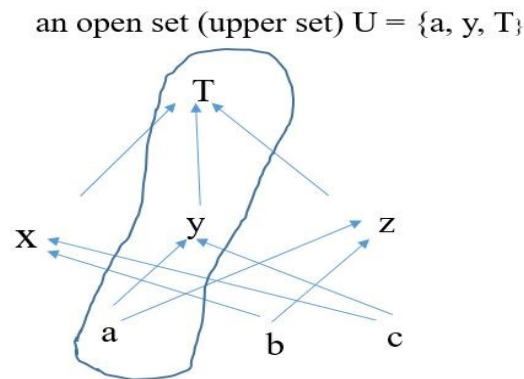


Figure 2.4. The upper set  $U$  represents an open set in the Alexandroff topology for the simplicial complex.

### 2.1.2 Constructing a Simplicial Complex from a Topological Space

Pavel Alexandroff [33] introduced the construction of a simplicial complex from the open covering of a topological space. All topological spaces in this dissertation are considered to be compact (they have finite open covers).

### 2.1.3 Alexandroff's Definition

Suppose  $\mathcal{G} = \{U_i ; i \in I\}$  is an open cover of the topological space  $X$ . The nerve complex  $N(\mathcal{G})$  of this open cover is constructed as follows:

The vertices (0-simplices) are the elements of the open cover. The intersection of the  $n$ -number of elements of the open cover represents a  $(n-1)$ -dimensional simplex (if nonempty), see Figure 2.5.

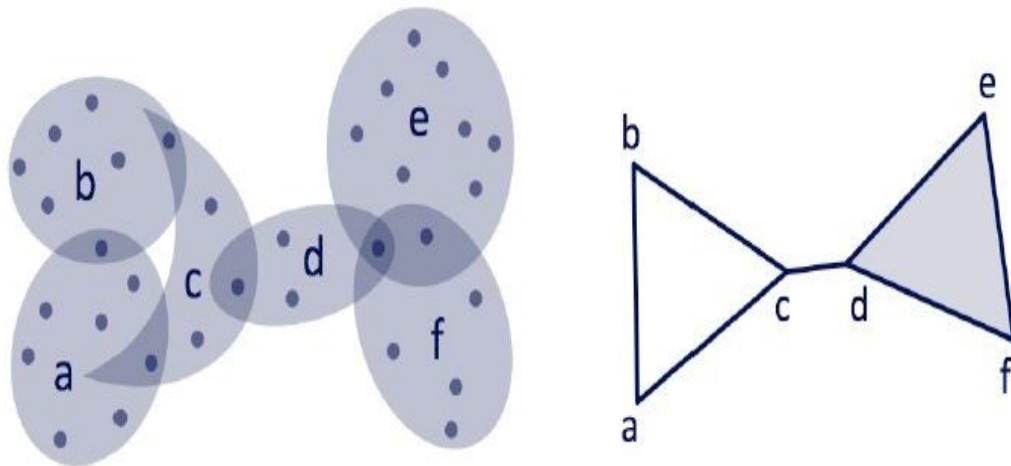


Figure 2.5. A cover  $\mathcal{G} = \{a, b, c, d, e, f\}$  of 6 sets with labels for each cover set (left) and its nerve complex (right) [34] .



The nerve complex is an appropriate approach for the construction of a simplicial complex from a dataset. In fact the topological space and its associated nerve complex have the same “shape.” More precisely:

**Theorem 2.1 [30] (Corollary 4G.3)** If  $\mathcal{O}$  is an open cover of a compact topological space  $X$  such that every nonempty intersection of finitely many sets in  $\mathcal{O}$  is contractible (contains no voids), then  $X$  is homotopy equivalent to the nerve  $\mathbf{N}(\mathcal{O})$ .

Throughout this dissertation, all simplicial complexes are considered to be finite (have finite number of simplices).

## 2.2 Simplicial Homology

The definitions and the theorems in this subsection are taken from [31] and [30]. Furthermore, all simplicial complexes are oriented (i.e., an order is assigned to their vertex sets).

**Definition 2.5** For the oriented simplicial complex  $\mathbf{K}$  and for a non-negative integer  $n$ , the  $n$ -chain real vector space  $C_n(\mathbf{K})$  is defined to be the formal sum of the  $n$ -simplices in  $\mathbf{K}$  with coefficients in  $\mathbb{R}$ . For simplicity when the simplicial complex  $\mathbf{K}$  is fixed, the notation  $C_n$  is applied.

**Remark 2.1** In a simplicial complex the  $n$ -chain vector space  $C_n$  is isomorphic to the direct sum of the copies of  $\mathbb{R}$  over the set of all  $n$  dimensional simplices.

**Definition 2.6** For each non-negative integer  $n$ , the linear boundary operator

$$C_{n+1} \xrightarrow{d_{n+1}} C_n \quad (2)$$

is defined on an element by

$$d_{n+1}(b) = \sum_{a < b} [b:a]a; \quad b \in C_{n+1} \quad (3)$$

and is extended linearly to the entire space  $C_{n+1}$ .

**Theorem 2.2** For each non-negative integer  $n$ , the composition of two consecutive operators is trivial (i.e.  $d_n \circ d_{n+1} = 0$ ). Therefore, the following chain complex is constructed:

$$C_{n+1} \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} C_{n-1} \xrightarrow{d_{n-1}} \dots \dots \dots \xrightarrow{d_2} C_1 \xrightarrow{d_1} C_0 \xrightarrow{d_0} 0 \quad (4)$$

From the above-referenced theorem the subgroup relation  $Img \, d_{n+1} \subseteq Ker \, d_n$  is concluded and therefore, an equivalent relation is defined as follows. Any real vector space is an abelian group.

**Definition 2.7** Two elements  $c_1$  and  $c_2 \in Ker \, d_n$  are homologous if and only if  $c_1 - c_2 \in Img \, d_{n+1}$ .

It can be seen that the homologous relation is an equivalent relation. The equivalent classes make a group (homology group of the simplicial complex).

**Definition 2.8** The formula for computation of the  $p$ -dimensional homology group is as follows:

$$H_p = \frac{\text{Ker } d_p}{\text{Img } d_{p+1}} \quad (5)$$

Remark 2.2 The interpretation of the p-dimensional homology group  $H_p$  is as follows:

1 -  $H_0$  represents the number of connected components. If the simplicial complex has n combinatorial connected components, then the 0-homology group  $H_0$  is the direct sum of n-copies of  $\mathbb{R}$ .

2 -  $H_1$  represents the number of the one dimensional holes.

3 -  $H_n$  ( $n > 1$ ) represents number of the voids (n-dimensional holes).

The application of this interpretation for the coverage and hole-detection in sensor networks is shown in [35], [36], and [37].

The following two examples demonstrate the computation and interpretation of the homology groups, Figure 2.6.

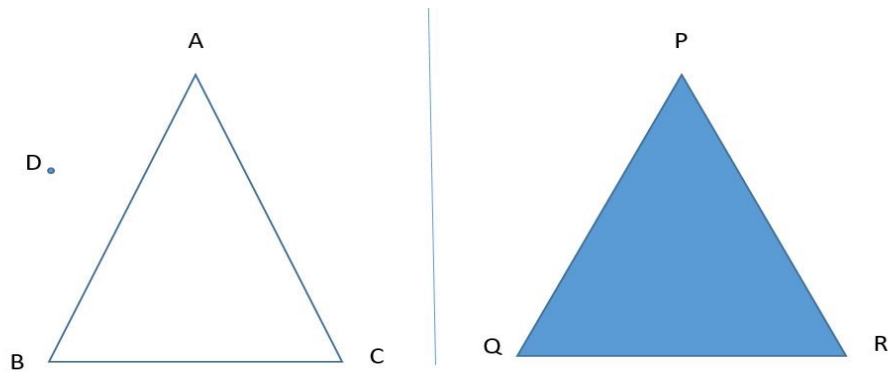


Figure 2.6. The simplicial complexes  $K_1$  with two components and one hole (left) and  $K_2$  with one component and no holes (right).

Computation of homology groups for the simplicial complex  $K_1$  with orientation AB, AC and BC:

$C_0$  = The  $\mathbb{R}$ -vector space generated by the 0-simplices A, B, C, D as the basis elements

$$= \{a_1 A + a_2 B + a_3 C + a_4 D : a_i \in \mathbb{R}\} \cong \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$C_1$  = The  $\mathbb{R}$ -vector space generated by the 1-simplices AB, AC, BC as basis elements

$$= \{b_1 AB + b_2 AC + b_3 BC : b_i \in \mathbb{R}\} \cong \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$$d_0 : C_0 \rightarrow 0$$

$$d_0 (a_1 A + a_2 B + a_3 C + a_4 D) = a_1 d_0 A + a_2 d_0 B + a_3 d_0 C + a_4 d_0 D = 0$$

(since the boundary of a vertex is zero). Consequently  $\text{Ker } d_0 = C_0 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$ .

$$d_1 : C_1 \rightarrow C_0$$

$$d_1 (b_1 AB + b_2 AC + b_3 BC) = b_1 d_1 AB + b_2 d_1 AC + b_3 d_1 BC = b_1 (B-A) + b_2 (C-A) + b_3 (C-B)$$

$$= (-b_1 - b_2) A + (b_1 - b_3) B + (b_2 + b_3) C.$$

To compute  $\text{Im } d_1$ , consider the following equation:

$$(-b_1 - b_2) A + (b_1 - b_3) B + (b_2 + b_3) C = a_1 A + a_2 B + a_3 C + a_4 D.$$

Compare the coefficients to obtain:

$$-b_1 - b_2 = a_1 ; b_1 - b_3 = a_2 ; b_2 + b_3 = a_3 \text{ and } a_4 = 0.$$

Sum up the above-referenced equations to get  $a_1 + a_2 + a_3 = 0$ . Thus, the degree of freedom is 2. Consequently:

$Img\ d_1 = \mathbb{R} \oplus \mathbb{R}$  and therefore  $H_0 = \frac{Ker\ d_0}{Img\ d_1} = \frac{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}}{\mathbb{R} \oplus \mathbb{R}} = \mathbb{R} \oplus \mathbb{R}$ , meaning that the simplicial complex  $K_1$  has two components.

To calculate the  $ker\ d_1$  consider the equality:

$$d_1 (b_1 AB + b_2 AC + b_3 BC) = (-b_1 - b_2) A + (b_1 - b_3) B + (b_2 + b_3) C = 0.$$

So, each coefficient must be zero (since A, B, C are the basis for the vector space  $C_0$ ),

$$-b_1 - b_2 = 0 ; b_1 - b_3 = 0 ; b_2 + b_3 = 0.$$

As a result  $b_1 = -b_2 = b_3$ . So the degree of freedom is 1 and  $Ker\ d_1 = \mathbb{R}$ . Consequently:

$$H_1 = \frac{Ker\ d_1}{Img\ d_2} = \frac{\mathbb{R}}{0} = \mathbb{R}, \text{ meaning that the simplicial complex } K_1 \text{ has one hole.}$$

Computation of homology groups for the simplicial complex  $K_2$  with orientation PQ, QR and RP and PQR:

$C_0$  = The  $\mathbb{R}$ -vector space generated by the 0-simplices P, Q, R as the basis elements.

$$= \{a_1 P + a_2 Q + a_3 R : a_i \in \mathbb{R}\} = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$C_1$  = The  $\mathbb{R}$ -vector space generated by the 1-simplices PQ, QR, RP as the basis elements.

$$= \{b_1 PQ + b_2 QR + b_3 RP : b_i \in \mathbb{R}\} = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$C_2$  = The  $\mathbb{R}$ -vector space generated by the only 2-simplex PQR as the basis element.

$$= \{e PQR : e \in \mathbb{R}\} = \mathbb{R}$$

$$d_0 : C_0 \rightarrow 0$$

$$d_0 (a_1 P + a_2 Q + a_3 R) = 0 \text{ (since the boundary of a vertex is zero).}$$

$$\text{Consequently } \text{Ker } d_0 = C_0 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}.$$

$$d_1 : C_1 \rightarrow C_0$$

$$d_1 (b_1 PQ + b_2 QR + b_3 RP) = b_1 (Q-P) + b_2 (R-Q) + b_3 (P-R) = (-b_1 + b_3) P + (b_1 - b_2) Q + (b_2 - b_3) R$$

$$= a_1 P + a_2 Q + a_3 R$$

Comparing the coefficients results in the equation  $a_1 + a_2 + a_3 = 0$ . Therefore, the degree of freedom is 2. Consequently:

$$\text{Img } d_1 = \mathbb{R} \oplus \mathbb{R} \text{ and therefore } H_0 = \frac{\text{Ker } d_0}{\text{Img } d_1} = \frac{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}}{\mathbb{R} \oplus \mathbb{R}} = \mathbb{R}, \text{ meaning that the}$$

simplicial complex  $K_2$  has one component.

To calculate  $\text{ker } d_1$  consider the equality:

$$d_1 (b_1 PQ + b_2 QR + b_3 RP) = b_1 (Q-P) + b_2 (R-Q) + b_3 (P-R) = (-b_1 + b_3) P + (b_1 - b_2) Q + (b_2 - b_3) R = 0.$$

As a result  $b_1 = b_2 = b_3$ . So the degree of freedom is 1 and  $\text{Ker } d_1 = \mathbb{R}$ .

$$d_2 : C_2 \rightarrow C_1$$

$$d_2 (e PQR) = e d_2 (PQR) = e (PQ + QR + RP) = e PQ + e QR + e RP$$

To compute  $\text{Img } d_2$ , consider the following equation:

$e \text{ PQ} + e \text{ QR} + e \text{ RP} = b_1 \text{ PQ} + b_2 \text{ QR} + b_3 \text{ RP}$  . Compare the coefficients to obtain:

$e = b_1 = b_2 = b_3$  . So the degree of freedom is 1 and  $\text{Img } d_2 = \mathbb{R}$ .

$H_1 = \frac{\text{Ker } d_1}{\text{Img } d_2} = \frac{\mathbb{R}}{\mathbb{R}} = 0$ , indicating that the simplicial complex  $K_2$  has no holes.

### 2.2.1 Computation of Homology Groups Algorithm

Consider the following chain complex extracted from a simplicial complex  $K$ :

$$C_{n+1} \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} C_{n-1} \xrightarrow{d_{n-1}} \dots \dots \dots \xrightarrow{d_2} C_1 \xrightarrow{d_1} C_0 \xrightarrow{d_0} 0 \quad (6)$$

To compute the  $n^{\text{th}}$ -homology groups for this chain complex, the following considerations are crucial:

1- The image of the operator  $d_{n+1}$  is inside the kernel of  $d_n$  ( $d_n \circ d_{n+1} = 0$ ).

So, to compute  $H_n = \frac{\text{Ker } d_n}{\text{Img } d_{n+1}}$ , one must look at the two sequential operators:

$$C_{n+1} \xrightarrow{d_{n+1}} C_n \xrightarrow{d_n} C_{n-1} \quad (7)$$

To simplify the identification of  $\text{Img } d_{n+1}$  inside the  $\text{Ker } d_n$ , rows and columns reduction is applied from the co-reduction homology algorithm formula from [38] and [39]. This algorithm is applied to the rows and columns of the matrices corresponding to the linear operators  $d_{n+1}$  and  $d_n$  to create as many zero rows and columns as possible.

2- The basis of the vector space  $C_n$  generates the rows of the matrix associated with the linear operator  $d_{n+1}$ . In the meantime, it generates the columns of the linear operator  $d_n$ .

The rows and columns reduction can reduce the matrices as simply as possible. Let's call the matrices in the new basis  $D_{n+1}$  and  $D_n$ .

3- Suppose the column reduction is applied by a matrix  $Q$  to the operator  $d_n$ . The inverse of the matrix  $Q$  ( $Q^{-1}$ ) is applied to the operator  $d_{n+1}$ , since

$$d_n \circ d_{n+1} = d_n \circ Q \circ Q^{-1} d_{n+1} = 0 \quad (8)$$

Set  $d_n \circ Q = D_n$  and  $Q^{-1} d_{n+1} = D_{n+1}$ .

4- Everything is in the place to conclude that:

$$\text{Ker } D_n = \text{the span of the zero columns of the matrix } D_n \quad (9)$$

$$\text{Img } D_{n+1} = \text{the span of the nonzero rows of the matrix } D_{n+1} \quad (10)$$

5- The  $n$  dimensional homology group is:

$$\begin{aligned} H_n &= \frac{\text{Ker } D_n}{\text{Img } D_{n+1}} = \frac{\text{the span of the zero columns of the matrix } D_n}{\text{the span of the nonzero rows of the matrix } D_{n+1}} \\ &= \text{the span of the quotient} \end{aligned} \quad (11)$$

### 2.2.2 The Python Program for Homology Computation

The Python Program for this subsection is from repository “GITHUB” [40].

Part 1: Auxiliary functions for doing the elementary operations on rows and columns on matrices. Everything is done in “numpy.”



```

1  import numpy
2
3  def rowSwap(A, i, j):
4      temp = numpy.copy(A[i, :])
5      A[i, :] = A[j, :]
6      A[j, :] = temp
7
8  def colSwap(A, i, j):
9      temp = numpy.copy(A[:, i])
10     A[:, i] = A[:, j]
11     A[:, j] = temp
12
13  def scaleCol(A, i, c):
14     A[:, i] *= c*numpy.ones(A.shape[0])
15
16  def scaleRow(A, i, c):
17     A[i, :] *= c*numpy.ones(A.shape[1])
18
19  def colCombine(A, addTo, scaleCol, scaleAmt):
20     A[:, addTo] += scaleAmt * A[:, scaleCol]
21
22  def rowCombine(A, addTo, scaleRow, scaleAmt):
23     A[addTo, :] += scaleAmt * A[scaleRow, :]

```

Part 2: The column reduction is applied by a matrix  $Q$  to the operator  $d_n$ . The inverse of the matrix  $Q$  (*i.e.*  $Q^{-1}$ ) is applied to the operator  $d_{n+1}$ , the algorithm is doing column reduction on one matrix and applying the corresponding row operations to the other.

```

1  def simultaneousReduce(A, B):
2      if A.shape[1] != B.shape[0]:
3          raise Exception("Matrices have the wrong shape.")
4
5      numRows, numCols = A.shape # col reduce A
6
7      i,j = 0,0
8      while True:
9          if i >= numRows or j >= numCols:
10             break
11
12         if A[i][j] == 0:
13             nonzeroCol = j
14             while nonzeroCol < numCols and A[i,nonzeroCol] == 0:
15                 nonzeroCol += 1
16
17         if nonzeroCol == numCols:
18             i += 1
19             continue
20
21         colSwap(A, j, nonzeroCol)
22         rowSwap(B, j, nonzeroCol)
23
24         pivot = A[i,j]
25         scaleCol(A, j, 1.0 / pivot)
26         scaleRow(B, j, 1.0 / pivot)
27
28         for otherCol in range(0, numCols):
29             if otherCol == j:
30                 continue
31             if A[i, otherCol] != 0:
32                 scaleAmt = -A[i, otherCol]
33                 colCombine(A, otherCol, j, scaleAmt)
34                 rowCombine(B, j, otherCol, -scaleAmt)
35
36         i += 1; j += 1
37
38     return A,B

```

Part 3: The actual algorithm to compute homology is counting pivots. Here are two pivot counting functions in numpy fashion.

```

1  def numPivotCols(A):
2      z = numpy.zeros(A.shape[0])
3      return [numpy.all(A[:, j] == z) for j in range(A.shape[1])].count(False)
4
5  def numPivotRows(A):
6      z = numpy.zeros(A.shape[1])
7      return [numpy.all(A[i, :] == z) for i in range(A.shape[0])].count(False)

```

Part 4: The final function is:

```

1  def bettiNumber(d_k, d_kplus1):
2      A, B = numpy.copy(d_k), numpy.copy(d_kplus1)
3      simultaneousReduce(A, B)
4
5      dimKChains = A.shape[1]
6      kernelDim = dimKChains - numPivotCols(A)
7      imageDim = numPivotRows(B)
8
9      return kernelDim - imageDim

```

### 2.3 Simplicial Complex Beyond the Graph Structure for Data Representation

For cases involving data sources that encompass more than two interactions, a combinatorial topology as a higher-dimensional version of graph theory is required. This mathematical model is provided by utilizing topological methods which produced simple representations of the data. Simplicial complex technique generalized the graph-theoretic modeling. The key idea is to represent all possible data at some time as a single, static, combinatorial geometric object. The following example illustrates the simplicial complex modeling for the multiple co-authorship interactions. Authors can have mutual papers, (see Table 2.1). The simplicial complex model consists of 0-simplices which corresponded to each individual author. Edges (1-simplices) corresponded to the papers that two authors published jointly. Similarly,  $d$ -simplices ( $d > 1$ ) represent  $(d + 1)$ -authors who have jointly published papers. Refer to Figure 2.7.

Table 2.1. The Co-authorship

Authors vs. Papers	Paper #1	Paper #2	Paper #3	Paper #4
Author A				
Author B				
Author C				
Author D				
Author E				

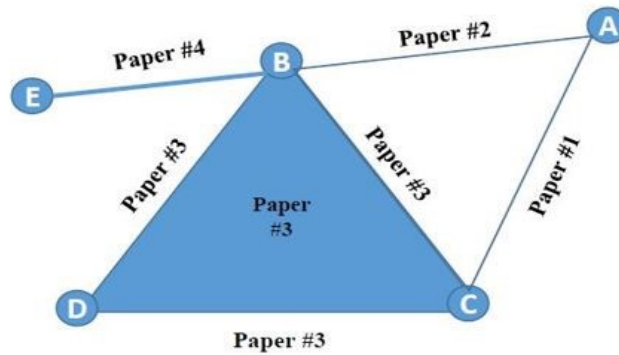


Figure 2.7. The simplicial complex model for the co-authorship. Vertices represent the authors. The edges and triangle represent the multiple co-authorship.

The following is the algorithm for authorship representation by simplicial complex.

- Add a vertex for each author
- Order all vertices according to their indices. Let  $V = \{a_0 a_1 a_2 a_3 \dots a_n\}$ , the order is  $a_0 < a_1 < a_2 < a_3 < \dots < a_n$ .
- For  $k = 1$  to  $p$  ( number of papers)
  - SUM  $\leftarrow 0$ ;
  - For  $j = 1$  to  $n$  ( number of authors)
    - If paper  $k$  has  $j$ -authors then SUM  $\leftarrow$  SUM + 1

```

    End if
    make a (SUM-1)-simplex on those related authors
  End for j
End for k

```

More sophisticated examples regarding the application of simplicial complexes to modeling the phenomena may be found in [41] and [42].

## 2.4 Summary

This chapter is devoted to the topological approach to identify and study the system through the shape of data and data sources. The topological modeling consists of objects, namely simplices that are attached to make a simplicial complex. This is the visualization and representation of the collection of data and their sources simultaneously in a single combinatorial geometric and topological object.

Data sources are represented by the 0-dimensional simplices and interactions among two and more sensors are represented by higher dimensional simplices.

Algebraic objects, namely, homology groups of the simplicial complex help in the interpretation of the behavior of the system based on its potential to exchange data. They could also detect possible failure in data exchange.

The computational formulas for homology groups are stated, and the algorithms for the computation of the groups are presented in details.

The advantage of simplicial complex modeling over the graph structures is explained in detail and is shown by an example.

## CHAPTER 3

### SHEAVES, DATA FUSION, COHOMOLOGY AND DATA ANALYSIS

In mathematics, when data are locally attached to open sets of a topological space, the sheaf theory is a tool to track the locally defined data. This chapter starts with an abstract definition of cellular sheaves and continues with comprehensive details regarding the computation of sheaf cohomology and its application in data analysis. The sheaf theory model analyzes heterogeneous data types by the integration of data collected from sensor clusters. The mathematical construction is the sheaf of vector spaces over a simplicial complex. Without being too complicated, the structure of vector spaces are strong enough for analyzing and integrating heterogeneous data and their redundancy. The foundations of sheaf theory that cover the algebraic geometer's schemes as well as the topological and analytic kinds can be found in [43].

#### 3.1 Cellular Sheaves of Vector Spaces

Cellular sheaves are mathematical structures that are built on simplicial complexes. In fact, a cellular sheaf is an assignment of data to each simplex in a simplicial complex together with the two pillars: first, it addresses the restrictions of data from a smaller simplex to the larger one and second, it deals with the information consistency in the overlap of two data sources. The categorical point of view for the definition of cellular sheaves may be found in [44]. A linear algebraic data presentation for the category of sheaves on simplicial complexes is obtained from [45]. The concept of sheaves in a categorical manner is obtained in [46]. The cellular sheaves point of view in this

dissertation is associated with the field of computer science, and the definitions are presented accordingly.

**Definition 3.1** Let  $\mathbf{K}$  be a simplicial complex. A cellular sheaf  $F$  of vector spaces over the simplicial complex  $\mathbf{K}$ , consists of the following two assignments. See Figure 3.1:

1. Assignment of a vector space  $F(S)$  to each simplex  $S$  in  $\mathbf{K}$ . The vector space is called the stalk of the simplex  $S$ . Each element of the vector space  $F(S)$  is called a local section at  $S$ .
2. Assignment of a linear map  $(S \rightarrow C) : F(S) \rightarrow F(C)$  for any two simplices  $S$  and  $C$  in  $\mathbf{K}$  with  $S < C$  ( $S$  a face of  $C$ ). This linear map is called the restriction map.

The assignments are such that the three simplices with the face relation  $S < C < D$  satisfy:

$$F(S \rightarrow C) \circ F(C \rightarrow D) = F(S \rightarrow D) \quad (12)$$

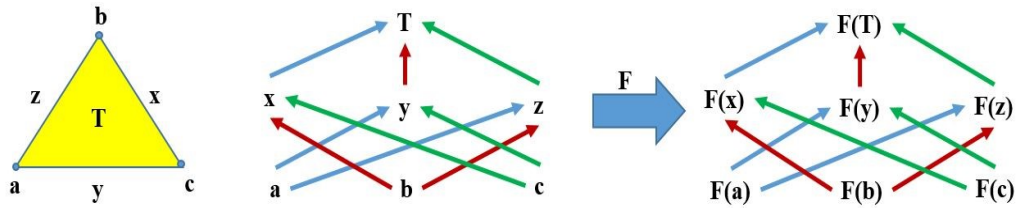


Figure 3.1. An example of a simplicial complex (left), the associated sheaf  $F$  (middle and right). Inclusions of the faces are shown by upward arrows.

**Definition 3.2** For a sheaf  $F$  on a simplicial complex  $\mathbf{K}$ , a global section is an assignment of values from each of the stalks that is consistent with the restrictions. More precisely, the local sections  $f(S_p) \in F(S_p)$  and  $f(S'_p) \in F(S'_p)$  can be glued together to make a

global section if and only if for any two  $p$ -simplices  $S_p$  and  $S'_p$  and any  $p + 1$ -simplex  $S_{p+1}$  with  $S_p, S'_p < S_{p+1}$ , the following equality satisfies:

$$F(S_p \rightarrow S_{p+1})(f(S_p)) = F(S'_p \rightarrow S_{p+1})(f(S'_p)) \quad (13)$$

In [47] Hubbard states, “It is fairly easy to understand what a sheaf is, especially after looking at a few examples. Understanding what they are good for is rather harder; indeed, without cohomology theory, they aren’t good for much.”

The following example from [48] gives an idea of representation of data in a cellular sheaf.

**Example 3.1** Consider a student who attends high school, an undergraduate institution, a graduate institution, and then is accepted in a postdoctoral position. Each school that the student attends maintains records of his grades. Each institution is represented as a vertex in a cell complex, as shown in Figure 3.2.

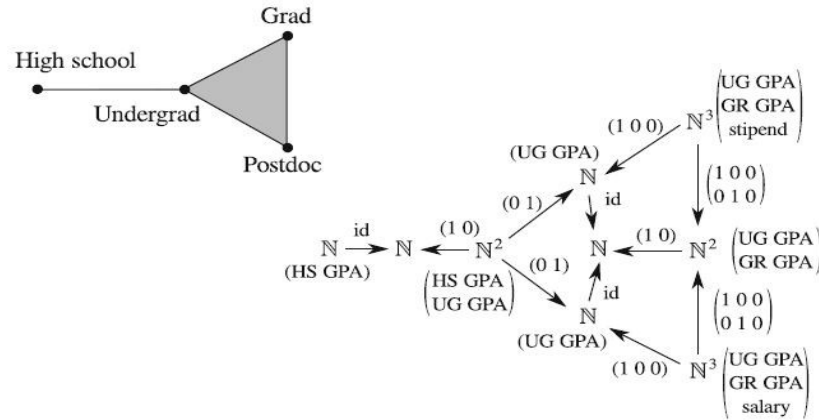


Figure 3.2. A network of academic institutions that might share information about a student (left), and a sheaf representing associated information about a single student (right) [48].



Every pair of institutions that shares a piece of information is represented as an edge between their respective vertices. A common piece of information that is shared among three institutions is represented as a 2-simplex. For instance, high schools typically only communicate with undergraduate institutions, therefore, no edges exit between a high school's vertex and any other institutions. Assume the following:

1. The high school only keeps a record of the high school GPA.
2. The undergraduate institution keeps records of both the high school and the undergraduate GPAs.
3. The graduate institution keeps records of the undergraduate and graduate GPAs, and any graduate stipend.
4. The postdoctoral institution keeps records of the undergraduate and graduate GPAs, and postdoctoral salary.
5. Stipend and salary information is not shared between institutions.
6. Grades are shared as appropriate and are consistent.

The assumptions lead to the sheaf structure shown on the right of Figure 3.2. Each piece of information is represented by a natural number (grades and salaries cannot be negative, and are rounded to the nearest whole number). In the sheaf structure, the stalk over each vertex contains the information held by each institution. Each edge of the complex contains the information shared by the two institutions. Each 2-simplex contains the common information among three institutions, which, in this example, is only the undergraduate GPA. Each restriction map is represented by a projection matrix that selects

the appropriate shared information. In particular, the restriction maps from the two postgraduate institutions share no any information regarding the student's pay.

Hereafter, “sheaf” means cellular sheaf of vector spaces (stalks are real vector spaces).

### 3.2 Sheaf Cohomology

Since all the topological spaces and accordingly all simplicial complexes under consideration in this dissertation are paracompact (every open cover had an open refinement that is locally finite), according to [49] (theorem 3.16), the sheaf cohomology on the simplicial complex  $\mathbf{K}$  is isomorphic to the Čech cohomology. For the detail on Čech cohomology see [50]. For more detailed definition of Čech cohomology, see [51] and [52]. This dissertation relies on sheaf cohomology based on the Čech cochains. The remainder of this section provides theoretical implementation about the concept of sheaf cohomology and its interpretation and application in computer science. All definitions related to the sheaf cohomology are given according to the above-mentioned isomorphism.

**Definition 3.3** Suppose  $F$  is a sheaf on a simplicial complex  $\mathbf{K}$ . The  $p$ -cochain group is defined to be the direct sum of stalks over all  $p$ -simplices  $S_p$  in  $\mathbf{K}$ :

$$C^p(\mathbf{K}; F) = \bigoplus_{S_p \in \mathbf{K}} F(S_p) = \bigoplus_{S_p \in \mathbf{K}} \text{Stalk}(S_p) \quad (14)$$

From now on when the simplicial complex  $\mathbf{K}$  and the associated sheaf  $F$  are known, the simplified notation  $C^p$  is applied instead of  $C^p(\mathbf{K}; F)$ .

**Definition 3.4** For each non-negative integer  $n$ , the linear coboundary operator

$$\mathcal{C}^n \xrightarrow{d^n} \mathcal{C}^{n+1} \quad (15)$$

is defined by

$$d^n(c)(S_{n+1}) = \sum_{S_n \in K} [S_{n+1} : S_n] F(S_n \rightarrow S_{n+1}) c(S_n) \quad (16)$$

for all  $c \in \mathcal{C}^n$  and  $S_{n+1} \in \mathbf{K}$ . The matrix form of the coboundary operator can be written as:

$$d^n = [ [S_{n+1} : S_n] (F(S_n \rightarrow S_{n+1})) ]_{S_n, S_{n+1} \in \mathbf{K}} \quad (17)$$

Theorem 3.1 [51] For each non-negative integer  $n$ , the composition of two consecutive coboundary operators is trivial, i.e.  $d^n \circ d^{n-1} = 0$ . Thus, for the  $(n+1)$ -dimensional complex  $\mathbf{K}$ , the following Čech cochain complex is constructed:

$$\mathcal{C}^0 \xrightarrow{d^0} \mathcal{C}^1 \xrightarrow{d^1} \mathcal{C}^2 \rightarrow \dots \rightarrow \mathcal{C}^n \xrightarrow{d^n} \mathcal{C}^{n+1} \quad (18)$$

From theorem 3.1, the subgroup relationship  $\text{Img } d_{p-1} \subseteq \text{Ker } d_p$  is concluded and, therefore, an equivalent relation is defined as follows.

Definition 3.5 [48] The cohomology of the sheaf  $F$  over the simplicial complex  $\mathbf{K}$ , is defined to be the homology of the previous chain complex. It is denoted by  $(C^\bullet(\mathbf{K}; F), d)$ .

More precisely the  $p$ -cohomology group is defined by:

$$H^p(\mathbf{K}; F) = \frac{\text{Ker } d^p}{\text{Img } d^{p-1}} \quad (19)$$

The algorithms to compute the cohomology groups may be found in [53] and [54]. The computation of cellular sheaf cohomology from the Morse theory technique is described in [55].

Theorem 3.2 ([48] theorem 4.3). The space of global sections of the sheaf  $F$  over the simplicial complex  $\mathbf{K}$  is isomorphic to the zero<sup>th</sup> cohomology  $H^0(\mathbf{K}; F)$ .

From the theorem 3.2, and also from chapter 3 of [30], the following modified interpretation of the zero<sup>th</sup>-cohomology group is given for the purpose of this dissertation.

From the fact that  $(\text{Im } d^{-1} = 0)$  and  $H^0(\mathbf{K}; F) = \frac{\text{Ker } d^0}{\text{Im } d^{-1}} = \text{Ker } d^0$ , the following interpretation about the zero<sup>th</sup>-cohomology group is obtained.

Suppose  $\{S_1, S_2, S_3, \dots, S_t\}$  is the set of vertices (0-simplices) in the simplicial complex  $\mathbf{K}$ . Also suppose  $\{F(S_1), F(S_2), F(S_3), \dots, F(S_t)\}$  is the set of their corresponding stalks.

An element  $f = (f(S_1), f(S_2), f(S_3), \dots, f(S_t)) \in \bigoplus_{i=1, \dots, t} F(S_i)$  is in the  $H^0(\mathbf{K}; F)$  if and only if for all  $i, j = 1, \dots, t$  and  $S_i, S_j < S_{ij}$  ( $S_{ij}$  is the edge between  $S_i$  and  $S_j$ ),

$$F(S_i \rightarrow S_{ij})f(S_i) = F(S_i \rightarrow S_{ij})f(S_j) \quad (20)$$

Meaning that  $f(S_i)$  and  $f(S_j)$  both can be extended to the 1-simplex  $S_{ij}$ .

### 3.3 Pseudocodes for Computation of Cellular Sheaf

This subsection is devoted to the construction of the cellular sheaf over the simplicial complex  $\mathbf{K}$ . It is done in two sequential steps: first the assignment of stalks to

each simplex in the simplicial complex, and second, the definition of the restriction maps between the stalks.

### 3.3.1 Step 1 (part 1): Find the Vector Space of 0-Simplices

The preprocessor is given as a table  $T$  with  $n$  rows and  $m$  columns, respectively, for the representation of sensors and representation of data types as vector spaces. The table has the property that for a fixed column  $j$ , the row elements  $T_{ij}$  of the table (if nonzero) are all assigned to the same vector space.

Make the table  $Q$  with one column and  $n$  rows, same rows with the same sensors representations as of table  $T$ , and initialize it to empty.

```

For i = 1 .....n (number of the rows)
    V = zero vector space (place holder)
    For j = 1 .....m (number of the columns representing data type)
        V = V  $\oplus$   $T_{ij}$ 
    End For j
     $Q_i$  = V ( representation for 0-simplex in row i) (direct sum of  $T_{ij}$ 's)
End For i

```

### 3.3.2 Step 1 (part 2) Assignment of Stalks

Finding the Vector Space for Ordered Set of  $p$ -Simplices for  $p > 0$ .

```

p-simplex[r] = 0; p = 1....w simplex and r = 1... a
w and a are dynamic
Vector-space[w*a] = 0
For i = 1 .....n (number of the rows in Q)
    p = 1;
    r = 1
    result_intersection = 0
    p-simplex-dimension = 0 (p > 0)
    b = 0
    For t = i+1 .....n (number of the rows in matrix Q)

```

```

For j = 1 .....m (number of the columns representing data type position)
  If (( Tij is non-zero ) && (Tij is non-zero)) //if 1
    Then { //then 1
      If ( b == 0) //if 2
        Then { //then 2
          b = j
          result_intersection = result_intersection  $\oplus$  Tij
          p-simplex-dimension = +1
          p = p-simplex-dimension
          p-simplex[r] = +1
          Vector-space[p-simplex[r]] = result_intersection
        } //then 2
      Else { //else 2
        If ( b == j) //if 3
          Then { //then 3
            result_intersection = result_intersection  $\oplus$  Tij
            p-simplex-dimension = +1
            p = p-simplex-dimension
            p-simplex[r] = +1
            Vector-space[p-simplex[r]] = result_intersection
          } // then 3
        Else { //else 3
          b = j
          r = +1
          result_intersection = 0
          p-simplex-dimension = 0
          result_intersection = result_intersection  $\oplus$  Tij
          p-simplex-dimension = +1
          p-simplex[r] = +1
          Vector-space[p-simplex[r]] = result_intersection
        } // else 3
      } // if 3
    } //then 1 and if 1
  End for j
End For t
End For i

```

### 3.3.3 Step 2: Restriction Maps

i = 0

M: for each i-simplex and (i+1)-simplex

if (i-simplex and (i+1)-simplex is face-connected)  
 the restriction map is  
     number of rows from (i+1)-simplex  
     number of columns from i-simplex  
     find intersection of (i-simplex and (i+1)-simplex)  
     find exclusion ((i-simplex  $\setminus$  (i+1)-simplex)  
 A1 = ZERO matrix of exclusion (rows and columns from above)  
 I1 = Identity square matrix of intersection (based of rows)  
 Now, juxtapose A1 and I1 based on priority of intersection.  
 i = +1  
 If i < dimension of complex  
 Then {Go to instruction M}  
 Else done and continue

### 3.4 Mathematical Foundation for Sheaf Cohomology and Data Analysis

This subsection provides a mathematical foundation for analyzing the behavior of a system based on its potential to exchange data, possible failure in data exchange, detection of noise in the system, and recognition of the redundant or complimentary sensors. There are two sides of this spectrum:

1. One can deploy a small number of sophisticated “global” sensors with high signal complexity and precise readings.
2. In contrast, one can deploy a large number of small, coarse, “local” devices that may have large uncertainties in their readings.

Dealing with the two sides of the spectrum requires challenging data management. The challenge is to specify which type of mathematics is useful in analyzing the above scenarios.

In this subsection, the distributed (spatial and temporal) information system under consideration is fixed, the simplicial complex associated with this system is denoted by  $\mathbf{K}$ ,

and the sheaf of vector spaces over  $\mathbf{K}$  is denoted by  $F$ . The notation  $(\mathbf{K}; F)$  is used for such a system representation and the notation  $(C^\bullet(\mathbf{K}; F), d)$  is for its corresponding Čech (cochain) complex.

**Definition 3.6** The family

$$(f(S_p))_{S_p \in \mathbf{K}} \in C^p(\mathbf{K}; F) = \bigoplus_{S_p \in \mathbf{K}} F(S_p) = \bigoplus_{S_p \in \mathbf{K}} \text{Stalk}(S_p) \quad (21)$$

is called a  $p$ -integrating family if  $(f(S_p))_{S_p \in \mathbf{K}} \in \text{Ker } d^p$ .

**Remark 3.1** The 0-integrating families are global sections.

*Proof.* This is a result from theorem 3.2 and the fact that  $H^0(\mathbf{K}; F) = \frac{\text{Ker } d^0}{\text{img } d^{-1}} = \text{Ker } d^0$ .

**Definition 3.7** The sum of two vector spaces  $V$  and  $W$  is defined to be the span of the union of their basis. It is denoted by  $V+W$  or  $\text{span } V \cup W$ .

**Definition 3.8** Suppose  $S = \{S_1, S_2, \dots, S_m\}$  is the family of sensors in a system. A 1-refinement of this family is the subset  $S - \{S_i\}$  where  $F(S_i) \subseteq \text{span } \bigcup_{j \neq i} F(S_j)$ . The subset  $S - \{S_i\}$  is called the 1-refined family. Inductively the 1-refinement of the  $(n-1)$ -refined family is called the  $n$ -refinement of the family.

**Definition 3.9** The family  $S = \{S_1, S_2, \dots, S_m\}$  is non-refinable if there is no  $S_i$  for which  $F(S_i)$  is contained in  $\text{span } \bigcup_{j \neq i} F(S_j)$ .

**Definition 3.10** The maximal non-refined subset of the family  $S = \{S_1, S_2, \dots, S_m\}$  of sensors is defined as the set of significant sensors in the system.



**Main Theorem 3.3** A family of sensors represented by the vertex set  $\{S_1, S_2, \dots, S_t\}$  in the simplicial complex representation  $\mathbf{K}$  for the information system is a family of significant sensors if the local sections  $f(S_i) \in F(S_i)$  form a minimal span of the 0-integrating families. More precisely the significant sensors  $S_1, S_2, \dots, S_t$  satisfy the following equation:

$$F(S_1) + F(S_2) + \dots + F(S_t) = \text{Ker } d^0 \quad (22)$$

*Proof.* Suppose the information system has the set  $V = \{S_1, S_2, \dots, S_m\}$  as its vertex set and the set  $E = \{e_1, e_2, \dots, e_k\}$  as the set of its edges. Then

$$C^0(\mathbf{K}; F) = \bigoplus_{i=1, \dots, m} F(S_i) ; \quad C^1(\mathbf{K}; F) = \bigoplus_{j=1, \dots, k} F(e_j) \quad (23)$$

Since  $C^0 \xrightarrow{d^0} C^1$ , then  $d^0$  is a  $k \times m$  block matrix. For  $f = (f(S_1), f(S_2), \dots, f(S_m)) \in C^0(\mathbf{K}; F)$ , the  $j$ -th row of the matrix  $d^0(f)$  is  $\sum_{i=1}^m F(S_i \rightarrow e_j) f(S_i)$ .

The equality  $d^0(f) = 0$  is equivalent to the following system of equations for  $f$ :

$$\sum_{i=1}^m F(S_i \rightarrow e_j) f(S_i) = 0 \quad ; \quad j = 1, \dots, k \quad (24)$$

The solution space of the above system of equations is, on one hand, the vector space  $\text{Ker } d^0$  and, on the other hand, has as a basis, the union of the basis for those  $F(S_i)$  for

which  $F(S_i)$  is not contained in the span of  $\bigcup_{j \neq i} F(S_j)$ . With re-index modification after refinement:

$$F(S_1) + F(S_2) + \dots + F(S_t) = \text{Ker } d^0 \text{ and the proof is complete Q.E.D.}$$

### 3.5 Summary

An analysis of data, encoding and translating heterogeneous data into common language are modeled by stalks. The fusion of data extracted from multiple sensors is modeled by a sheaf. The methodology studies the behavior of the system based on the detection of noise, possible failure in data exchange and recognition of the redundant or complimentary sensors.

To verify the validity of the above-referenced method and to bring the information of the system into a mathematical language, the required definitions are presented. The main classification theorem is presented to bring up a necessary and sufficient condition for a sensor to be significant.

## CHAPTER 4

### APPLICATIONS

This chapter is devoted to the application of the modeling from the previous chapters. First, the methodology is applied to study the wildfire threat monitoring [56]. In this example heterogeneous data are gathered from a variety of in-the-field stations, each with a potentially different set of sensors for temperature, wind, humidity, smoke, and hotspots in the infrared spectrum. Satellite images or aerial photography are also used. Second, the example of air traffic monitoring with multiple sensors of various types is applied [57]. Heterogeneous data are gathered from variety of sensor clusters: GPS satellites, radar stations, airport surface detectors, and smart IR (infrared) sensors.

In both examples, a duplication of the sensors of the same type is possible. The individual sensors may come online or gone offline at irregular intervals of time and space and may become permanently disabled. Therefore, the structure, availability, granularity, and quality of the data may vary by data source and type.

#### **4.1 Part 1: Example of Wildfire Threat Monitoring**

It has been reported that for the last decade, each year, more than 100,000 wildfires and forest fire threats have occurred in all countries. Which type of mathematics can be applied to analyze the collaboration of the sensors to monitor the possibility of such a natural disaster? The mathematical framework to collect local information and apply it into global environmental data utilizes the simplicial complex and sheaf models. The construction of a simplicial complex and sheaf data structure is applied to answer the question, “Do multiple

cells (sensors) work together? If so, how?” Multiple sensors of various types monitor regions for wildfires. To make the detection more precise, duplication of the sensors of the same type is considered. Sensors of the same type communicate and report a common information. The heterogeneous data are received by the sensors of various types in the region of detection at time  $t = t_0$ . The types of the sensors, their duplication numbers, and the heterogeneous data received by the sensors are shown in Tables 4.1 and 4.2.

Table 4.1. Sensors and Duplication Numbers ( $t = t_0$ ) for Wildfire Monitoring

Sensor type	Number of Sensors time $t=t_0$
Satellite Camera, C	n
CO2 Detector, O	m
IR Detector, R	p
Flame Detector, D	q

Table 4.2. The Heterogeneous Data ( $t = t_0$ ) for Wildfire Monitoring

Sensors vs. Data	Fire Size $F, \mathbb{R}^2$	Intensity $I, \mathbb{R}$	Temperature $T, \mathbb{R}$	Smoke size $S, \mathbb{R}^2$
Satellite Camera, C	✓			✓
CO2 Detector, O		✓	✓	✓
IR Detector, R			✓	
Flame Detector, D	✓		✓	

#### 4.1.1 The Construction of the Simplicial Complex

The integration of the received heterogeneous data are modeled by the simplicial complex structure as shown in Figure 4.1.

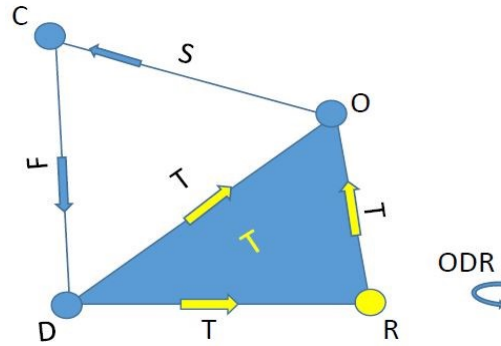


Figure 4.1. Simplicial complex model with oriented simplices for the wildfire threat monitoring at time  $t=t_0$ .

To obtain the desired measurements (homology groups) from the extracted data, orientation of the simplices in the simplicial complex model is required. The colored arrows represent the oriented simplices. The filled triangle ODR represents the shared data between the three sensors O, D and R. The hollow triangle OCD shows that there are no shared data among the three sensors O, C and D.

#### 4.1.2 Homology calculation at time $t = t_0$

The chain vector spaces are:

$C_0$  = The  $\mathbb{R}$ -vector space generated by the 0-simplices C, O, R, D as basis elements

$$= \{a_1 C + a_2 O + a_3 R + a_4 D : a_i \in \mathbb{R}\} = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$C_1$  = The  $\mathbb{R}$ -vector space generated by the 1-simplices CD, OC, RO, DO, DR as basis elements

$$= \{b_1 CD + b_2 OC + b_3 RO + b_4 DO + b_5 DR: b_i \in \mathbb{R}\} = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$C_2$  = The  $\mathbb{R}$ -vector space generated by the only 2-simplex ODR as the basis element.

$$= \{e ODR: e \in \mathbb{R}\} = \mathbb{R}$$

The linear boundary operators  $d_0, d_1, d_2$  are given by:

$$d_0 : C_0 \rightarrow 0$$

$$d_0 (a_1 C + a_2 O + a_3 R + a_4 D) = a_1 d_0 C + a_2 d_0 O + a_3 d_0 R + a_4 d_0 D = 0$$

(Since the boundary of a vertex is zero). Consequently  $\text{Ker } d_0 = C_0 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$ .

Now:

$$d_1 : C_1 \rightarrow C_0$$

$$\begin{aligned} d_1 (b_1 CD + b_2 OC + b_3 RO + b_4 DO + b_5 DR) &= b_1 d_1 CD + b_2 d_1 OC + b_3 d_1 RO + b_4 d_1 DO \\ &+ b_5 d_1 DR \end{aligned}$$

$$= b_1 (D-C) + b_2 (C-O) + b_3 (O-R) + b_4 (O-D) + b_5 (R-D)$$

$$= (b_1 - b_5 - b_4) D + (-b_1 + b_2) C + (-b_2 + b_3 + b_4) O + (-b_3 + b_5) R$$

To compute the  $\text{Img } d_1$ , consider the following equation:

$$(b_1 - b_5 - b_4) D + (-b_1 + b_2) C + (-b_2 + b_3 + b_4) O + (-b_3 + b_5) R = a_1 C + a_2 O + a_3 R + a_4 D.$$

Compare the coefficients to obtain:

$$b_1 - b_5 - b_4 = a_4$$

$$-b_1 + b_2 = a_1$$

$$-b_2 + b_3 + b_4 = a_2$$

$$-b_3 + b_5 = a_3$$

Sum up the above equations to get  $a_1 + a_2 + a_3 + a_4 = 0$ . The degree of freedom in this equation is 3 and consequently:

$$\text{Img } d_1 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \text{ and } H_0 = \frac{\text{Ker } d_0}{\text{Img } d_1} = \frac{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}}{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}} = \mathbb{R} \text{ and the dimension of } H_0 = 1.$$

To calculate the  $\text{ker } d_1$  consider the equality:

$$(b_1 - b_5 - b_4) D + (-b_1 + b_2) C + (-b_2 + b_3 + b_4) O + (-b_3 + b_5) R = 0.$$

Since D, C, O, R are basis elements for the vector space  $C_0$ , each coefficient must be zero:

$$b_1 - b_5 - b_4 = 0$$

$$-b_1 + b_2 = 0$$

$$-b_2 + b_3 + b_4 = 0$$

$$-b_3 + b_5 = 0$$

As a result:  $b_1 = b_2$ ,  $b_3 = b_5$ ,  $b_1 - b_5 - b_4 = 0$ ,  $-b_2 + b_3 + b_4 = 0$ .

The degree of freedom for this equation is 2 and  $\text{Ker } d_1 = \mathbb{R} \oplus \mathbb{R}$ .

$$d_2 : C_2 \rightarrow C_1$$

$$d_2(e \text{ ODR}) = e \cdot d_2(\text{ODR}) = e(\text{DR} - \text{OR} + \text{OD}) = e \text{ DR} - e \text{ OR} + e \text{ OD}$$

To compute the  $\text{Img } d_2$ , consider the following equation:

$$e \text{ DR} - e \text{ OR} + e \text{ OD} = b_1 \text{ CD} + b_2 \text{ OC} + b_3 \text{ RO} + b_4 \text{ DO} + b_5 \text{ DR}.$$

Compare the coefficients to obtain:

$$e = b_5, -e = -b_3, e = -b_4, b_1 = b_2 = 0$$

The degree of freedom is 1 and  $\text{Img } d_2 = \mathbb{R}$ . Consequently  $H_1 = \frac{\text{Ker } d_1}{\text{Img } d_2} = \frac{\mathbb{R} \oplus \mathbb{R}}{\mathbb{R}} = \mathbb{R}$ .

To compute  $H_2 = \frac{\text{Ker } d_2}{\text{Img } d_3}$ , consider the fact that there is no  $d_3$  and  $\text{Img } d_3 = 0$ .

Since  $d_2(e \text{ ODR}) = e \text{ DR} - e \text{ OR} + e \text{ OD} = 0$ , then  $e = 0$ , and  $\text{Ker } d_2 = 0$ . As a consequence

$$H_2 = \frac{\text{Ker } d_2}{\text{Img } d_3} = 0.$$

The remaining higher dimensional homology groups  $H_d$  ( $d > 2$ ) are all zero.

Results from calculation of the homology for the simplicial complex at time  $t = t_0$  are as follows:

$H_0 = \mathbb{R}$  (  $\dim H_0 = 1$  ), meaning that the simplicial complex is one connected.

$H_1 = \mathbb{R}$  (  $\dim H_1 = 1$  ), meaning that there is a one dimensional hole in this simplicial complex.

$H_n = 0$  for  $n > 1$ , meaning that in this simplicial complex there are no voids in dimension higher than 2D.



### 4.1.3 The Sheaf Construction

Each simplex carries some information. The information space is represented by a vector space assigned to each simplex. This assignment is the stalk over each simplex and carries all information about the data. It can be transferred to its neighboring nodes to analyze the system.

The stalk assignments are as follows:

$$\text{Stalk } C = F(C) = \{\text{Size of fire } \mathbb{R}^2, \text{ Size of Smoke } \mathbb{R}^2\} \simeq \mathbb{R}^2 \oplus \mathbb{R}^2$$

$$\text{Stalk } O = F(O) = \{\text{Intensity } \mathbb{R}, \text{ Temperature } \mathbb{R}, \text{ Size of smoke } \mathbb{R}^2\} \simeq \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2$$

$$\text{Stalk } R = F(R) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$$

$$\text{Stalk } D = F(D) = \{\text{Size of Fire } \mathbb{R}^2, \text{ Temperature } \mathbb{R}\} \simeq \mathbb{R}^2 \oplus \mathbb{R}$$

$$\text{Stalk } CO = F(OC) = \{\text{Size of smoke } \mathbb{R}^2\} \simeq \mathbb{R}^2$$

$$\text{Stalk } CD = F(CD) = \{\text{Size of fire } \mathbb{R}^2\} \simeq \mathbb{R}^2$$

$$\text{Stalk } OD = F(OD) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$$

$$\text{Stalk } OR = F(RO) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$$

$$\text{Stalk } DR = F(DR) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$$

$$\text{Stalk } ODR = F(ODR) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$$

The restriction maps are shown in Figure 4.2. The construction of these maps are encoded in the pseudocode for restriction maps that is discussed in Section 3.3.

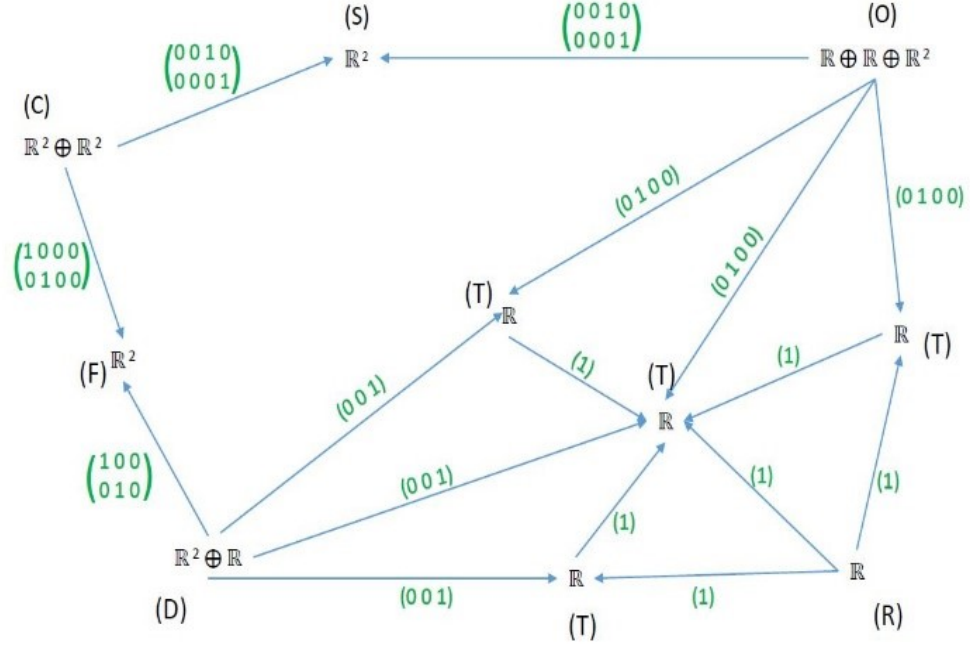


Figure 4.2. Sheaf, stalks and restriction maps associated with the simplicial complex for the wildfire threat monitoring example at time  $t = t_0$ .

#### 4.1.4 Sheaf Cohomology Calculation at Time $t = t_0$

The cochain vector spaces are:

$$\mathcal{C}^0 = F(C) \oplus F(O) \oplus F(R) \oplus F(D) = (\mathbb{R}^2 \oplus \mathbb{R}^2) \oplus (\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2) \oplus \mathbb{R} \oplus (\mathbb{R}^2 \oplus \mathbb{R})$$

$$\mathcal{C}^1 = F(OC) \oplus F(CD) \oplus F(DO) \oplus F(RO) \oplus F(DR) = \mathbb{R}^2 \oplus \mathbb{R}^2 \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$$

$$\mathcal{C}^2 = F(ODR) = \mathbb{R}$$

An element in  $\mathcal{C}^0$  is of the form  $(f(C), f(O), f(R), f(D)) \in \mathcal{C}^0$ , where  $f(C) \in F(C)$ ;  $f(O) \in F(O)$ ;  $f(R) \in F(R)$ ;  $f(D) \in F(D)$  are the local sections. In a similar way an element of  $\mathcal{C}^1$  is of the form  $(f(OC), f(CD), f(DO), f(RO), f(DR)) \in \mathcal{C}^1$  with the local sections

$$f(OC) \in F(OC); f(CD) \in F(CD); f(DO) \in F(DO); f(RO) \in F(RO); f(DR) \in F(DR).$$

These notations are applied in the computation of the coboundary maps.

The coboundary map  $d^0 : C^0 \rightarrow C^1$  is the  $\mathbb{R}$ -linear operator given by a  $5 \times 4$  dimensional block matrix  $d^0 = (a_{ij})$ ;  $i = 1, \dots, 5$ ;  $j = 1, \dots, 4$ . The detailed calculations are:

Components of  $f = (f(C), f(O), f(R), f(D)) \in C^0$  are given by:

$$f(C) = (f(C)_{fire}, f(C)_{smoke}) \in \mathbb{R}^2 \oplus \mathbb{R}^2$$

$$f(O) = (f(O)_{intensity}, f(O)_{temperature}, f(O)_{smoke}) \in \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}^2$$

$$f(R) = (f(R)_{temperature})$$

$$f(D) = (f(D)_{fire}, f(D)_{temperature}) \in \mathbb{R}^2 \oplus \mathbb{R}$$

With these notations the rows of the 4-dimensional block vector  $d^0 f = (a_{ij})f$  are given by:

$$\begin{aligned} a_{1j}f &= F(C \rightarrow OC)f(C) + F(O \rightarrow OC)f(O) + F(R \rightarrow OC)f(R) \\ &\quad + F(D \rightarrow OC)f(D) \end{aligned}$$

$$\begin{aligned} a_{2j}f &= F(C \rightarrow CD)f(C) + F(O \rightarrow CD)f(O) + F(R \rightarrow CD)f(R) \\ &\quad + F(D \rightarrow CD)f(D) \end{aligned}$$

$$\begin{aligned} a_{3j}f &= F(C \rightarrow DO)f(C) + F(O \rightarrow DO)f(O) + F(R \rightarrow DO)f(R) \\ &\quad + F(D \rightarrow DO)f(D) \end{aligned}$$

$$\begin{aligned} a_{4j}f &= F(C \rightarrow RO)f(C) + F(O \rightarrow RO)f(O) + F(R \rightarrow RO)f(R) \\ &\quad + F(D \rightarrow RO)f(D) \end{aligned}$$

$$\begin{aligned} a_{5j}f &= F(C \rightarrow DR)f(C) + F(O \rightarrow DR)f(O) + F(R \rightarrow DR)f(R) \\ &\quad + F(D \rightarrow DR)f(D) \end{aligned}$$

Calculating each row to obtain:

$$a_{11} = + \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; [\text{OC} : \text{C}] = +1$$

$$a_{12} = - \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; [\text{OC} : \text{O}] = -1$$

$$a_{13} = 0_{21}, a_{14} = 0_{23} \quad ; [\text{OC} : \text{R}] = [\text{OC} : \text{D}] = [\text{OC} : \text{O}] = 0$$

$$a_{21} = - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}; [\text{CD} : \text{C}] = -1$$

$$a_{22} = 0_{24}; [\text{CD} : \text{O}] = 0$$

$$a_{23} = 0_{21}; [\text{CD} : \text{R}] = 0$$

$$a_{24} = + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} ; [\text{CD} : \text{D}] = +1$$

$$a_{31} = 0_{14}; [\text{DO} : \text{C}] = 0$$

$$a_{32} = + [0 \ 0 \ 1] , [\text{DO} : \text{O}] = +1$$

$$a_{33} = 0_{11}; [\text{DO} : \text{R}] = 0$$

$$a_{34} = - [0 \ 0 \ 1] ; [\text{DO} : \text{D}] = -1$$

$$a_{41} = 0_{14} ; [\text{RO} : \text{C}] = 0$$

$$a_{42} = + [0 \ 1 \ 0 \ 0] ; [\text{RO} : \text{O}] = +1$$

$$a_{43} = - 1_{11} ; [\text{RO} : \text{R}] = -1$$

$$a_{44} = 0_{13} ; [\text{RO} : \text{D}] = 0$$

$$a_{51} = 0_{14} ; [\text{DR} : \text{C}] = 0$$

$$a_{52} = 0_{14} ; [\text{DR} : \text{O}] = 0$$

$$a_{53} = +1_{11} ; [\text{DR} : \text{R}] = +1$$

$$a_{54} = -[0 \ 0 \ 1] ; [\text{DR} : \text{D}] = -1$$

Here  $0_{ij}$  is the zero metrix with i-rows and j-columns.

The rows of the vector  $d^0 f = (a_{ij})f$  are:

$$d^0(f)(OC) = a_{1j}f = + \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} f(C) - \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} f(O) + 0_{21} + 0_{23}$$

$$d^0(f)(CD) = a_{2j}f = - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} f(C) + 0_{24} + 0_{21} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} f(D)$$

$$d^0(f)(DO) = a_{3j}f = 0_{14} + [0 \ 0 \ 1]f(O) + 0_{11} - [0 \ 0 \ 1]f(D)$$

$$d^0(f)(RO) = a_{4j}f = 0_{14} + [0 \ 1 \ 0 \ 0] f(O) - 1_{11}f(R) + 0_{13}$$

$$d^0(f)(DR) = a_{5j}f = 0_{14} + 0_{14} + 1_{11}f(R) - [0 \ 0 \ 1]f(D)$$

To compute the  $(\text{Ker } d^0)$ , notice that  $d^0 f = (a_{ij})f = 0$  if and only if:

$$f(C)_{\text{smoke}} = f(O)_{\text{smoke}} = \text{M}$$

$$f(C)_{\text{fire}} = f(D)_{\text{fire}} = \text{N}$$

$$f(O)_{\text{temperature}} = f(D)_{\text{temperature}} = f(R)_{\text{temperature}} = \text{P}$$

As a conclusion the element  $f = (f(C), f(O), f(R), f(D)) \in C^0$  belongs to  $\text{Ker } d^0$  if and

only if:

$$f(C) = (N, 0, 0, M)$$

$$f(O) = (0, \text{arbitrary}, P, M)$$

$$f(R) = (0, 0, P, 0)$$

$$f(D) = (N, 0, P, 0)$$

The zero components represent data that are not reported by the sensor. Consequently:

$$F(C) + F(O) = \text{Ker } d^0 \text{ or } F(D) + F(O) = \text{Ker } d^0 \quad (25)$$

As a result from calculations based on theorem 3.4.7, the significant sensors are either  $\{C, O\}$  or  $\{D, O\}$  and  $\text{Ker } d^0 \cong (\mathbb{R}^2 \oplus \mathbb{R}^2) \oplus (\mathbb{R} \oplus \mathbb{R})$  or  $\text{Ker } d^0 \cong (\mathbb{R}^2 \oplus \mathbb{R}) \oplus (\mathbb{R} \oplus \mathbb{R}^2)$ .

The zero cohomology is calculated from  $H^0 = \frac{\text{ker } d^0}{\text{img } d^{-1}} \cong \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$ .

Since there is no  $(d^{-1})$ , then  $\text{img } d^{-1} = 0$ , and the dimension of  $H^0$  is 6.

To calculate the first cohomology  $H^1 = \frac{\text{ker } d^1}{\text{img } d^0}$ , it is required to calculate  $\text{ker } d^1$  and  $\text{img } d^0$  separately. Since the matrix  $d^0$  has the number of 6 independent columns (calculated by MATLAB), the rank of the matrix is 6. For the calculation of  $\text{ker } d^1$ , consider the following:

$$d^1 : C^1 \rightarrow C^2 \quad ; \quad d^1 = (b_{1j}) \quad ; \quad j = 1, \dots, 5$$

$$b_{11} = 0_{12} \quad , \quad b_{12} = 0_{12} \quad , \quad b_{13} = -1_{11} \quad , \quad b_{14} = 1_{11} \quad , \quad b_{15} = 1_{11}$$

$$[\text{ODR} : \text{OC}] = [\text{ODR} : \text{CD}] = 0 \quad , \quad [\text{ODR} : \text{DO}] = -1 \quad , \quad [\text{ODR} : \text{RO}] = [\text{ODR} : \text{DR}] = 1$$

to get  $d^1 = (0_{12}, 0_{12}, -1_{11}, 1_{11}, 1_{11})$ .

For  $f = (f(OC), f(CD), f(DO), f(RO), f(DR)) \in C^1$ :

$$d^1(f) = (b_{1j})f = -f(DO) + f(RO) + f(DR) \quad (26)$$

From the above equation  $d^1(f) = 0$  if and only if  $f(RO) + f(DR) = f(DO)$  and

$$f(OC) = (0, 0, 0, f(OC)_{smoke}) = (0, 0, 0, W) ; W \in \mathbb{R}^2$$

$$f(CD) = (f(CD)_{fire}, 0, 0, 0) = (E, 0, 0, 0) ; E \in \mathbb{R}^2$$

$$f(RO) = (0, 0, f(RO)_{temperature}, 0) = (0, 0, U, 0)$$

$$f(DR) = (0, 0, f(DR)_{temperature}, 0) = (0, 0, V, 0)$$

$$f(DO) = (0, 0, f(DO)_{temperature}, 0) = (0, 0, U+V, 0)$$

Where  $W, E \in \mathbb{R}^2$  and  $U, V \in \mathbb{R}$ . Therefore

$$\ker d^1 = F(OC) \oplus f(CD) \oplus f(RO) \oplus f(DR).$$

Since the matrix  $d^0$  has the number of 6 independent columns (calculated by MATLAB),

$$\text{the rank of the matrix is 6. As a result } H^1 = \frac{\ker d^1}{\text{Im } d^0} = \frac{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}}{\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}} = 0.$$

$$d^2 : C^2 \rightarrow 0 \text{ and } \ker d^2 = C^2 = \mathbb{R}.$$

On the other hand the rank of the matrix  $d^1 = (0_{12}, 0_{12}, -1_{11}, 1_{11}, 1_{11})$  is 1,  $\text{Img } d^1 = \mathbb{R}$  and  $H^2 = \frac{\ker d^2}{\text{Img } d^1} = \frac{\mathbb{R}}{\mathbb{R}} = 0$ . The higher cohomology groups ( $H^d$  for  $d > 1$ ) will also be zero.

Results from calculation of sheaf cohomology (data analysis) at time  $t=t_0$ :

$H^0 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$  ( $\dim H^0 = 6$ ), meaning that at time  $t_0$  the significant sensors are  $\{C, O\}$  or  $\{D, O\}$ . The global information (section globalization) is extracted from the sensors  $\{C, O\}$  or  $\{D, O\}$ .

$H^1 = 0$ . The first cohomology group  $H^1 = \frac{\ker d^1}{\text{Img } d^0}$  characterizes the families of sections on the edges that come from the families of sections on the vertices. More precisely it figures out the number of 1-integrating families that do not belong to  $\text{Img } d^0$ . For the case in which the first cohomology group becomes zero, it means that all sections of the form  $f = (f(OC), f(CD), f(DO), f(RO), f(DR)) \in C^1$  which are also 1-integrating families come from families of sections on the sensors.

$H^2 = 0$  since there are no  $n$ -simplices for  $n > 1$ .

#### 4.1.5 Time Changes from $t=t_0$ to $t=t_1$

At time  $t=t_1$  the Table 4.1 has been changed to the Table 4.3. The  $i$ -number of CO2 sensor detectors go out of mission.



Table 4.3. Sensors and Duplication Numbers, Wildfire Monitoring ( $t = t_1$ )

Sensor type	Number of sensors time $t=t_1$
Satellite Camera, C	n
CO2 Detector, O	$m - i$
IR Detector, R	p
Flame Detector, D	q

As a result temperature is no longer detected by the sensor O. Table 4.4 shows the change in Table 4.2.

Table 4.4. The Heterogeneous Data for Wildfire Monitoring ( $t = t_1$ )

Sensors vs. data	Fire Size $F, \mathbb{R}^2$	Intensity $I, \mathbb{R}$	Temperature $T, \mathbb{R}$	Smoke size $S,$ $\mathbb{R}^2$
Satellite Camera, C	✓			✓
CO2 Detector, O		✓	out of mission	✓
IR Detector, R			✓	
Flame Detector, D	✓		✓	

The new simplicial complex is shown in Figure 4.3.

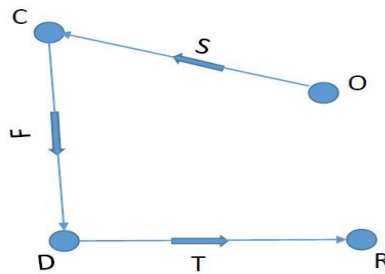


Figure 4.3. Simplicial complex model with oriented simplices for the wildfire threat monitoring at time  $t=t_1$ .

Similar to the calculations for the homology groups at time  $t = t_0$ , the calculations at time  $t = t_1$  are as follows:

$H_0 = \mathbb{R}$  ( $\dim H_0 = 1$ ), meaning that one connected simplicial complex exists.

$H_1 = 0$  ( $\dim H_1 = 0$ ), meaning that there is no 1-dimensional hole in the simplicial complex.

$H_n = 0$  for  $n > 1$ , meaning that in this simplicial complex there are no voids in dimension higher than 2D.

#### 4.1.6 The Sheaf Construction

The following new stalks are shown in Figure 4.4:

Stalk  $C = F(C) = \{\text{Size of fire } \mathbb{R}^2, \text{Size of Smoke } \mathbb{R}^2\} \simeq \mathbb{R}^2 \oplus \mathbb{R}^2$

Stalk  $O = F(O) = \{\text{Intensity } \mathbb{R}, \text{Size of smoke } \mathbb{R}^2\} \simeq \mathbb{R} \oplus \mathbb{R}^2$

Stalk  $R = F(R) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$

Stalk  $D = F(D) = \{\text{Size of Fire } \mathbb{R}^2, \text{Temperature } \mathbb{R}\} \simeq \mathbb{R}^2 \oplus \mathbb{R}$

Stalk  $CO = F(OC) = \{\text{Size of smoke } \mathbb{R}^2\} \simeq \mathbb{R}^2$

Stalk  $CD = F(CD) = \{\text{Size of fire } \mathbb{R}^2\} \simeq \mathbb{R}^2$

Stalk  $DR = F(DR) = \{\text{Temperature } \mathbb{R}\} \simeq \mathbb{R}$

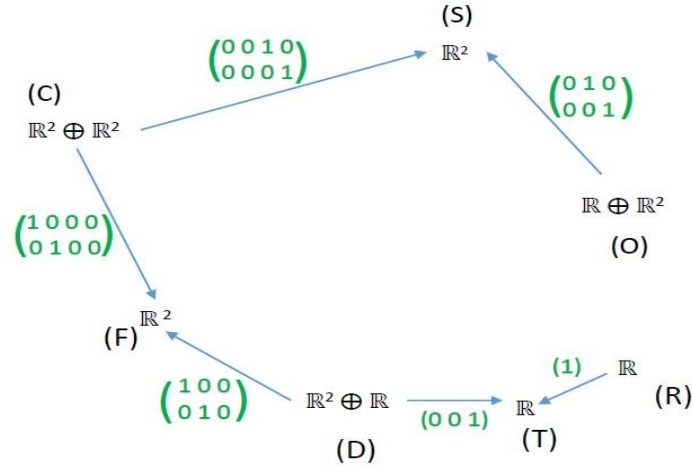


Figure 4.4. Sheaf, stalks and restriction maps associated with the simplicial complex for the wildfire threat monitoring for time  $t=t_1$ .

Similar to the sheaf cohomology calculations for time  $t = t_0$ , the cohomology calculations for time  $t=t_1$  are as follows:

$H^0 = \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$  ( $\dim H^0 = 6$ ), meaning that at time  $t_1$  the significant stalks are on O and D. The global information (section globalization) is extracted from O and D. In this situation the sensor C is no more significant although the detectors that no longer work, are from the CO2 detector O.

$H^1 = 0$ . The same interpretation applies here as in case  $t=t_0$ .

$H^2 = 0$ , since there are no 2-simplices.

#### 4.1.7 Discussion

From the calculations it is seen that the significant sensors are changed when there is a change in the number of sensors or if some sensors become inactive or out of mission.

The changes in the homology groups from time  $t=t_0$  to time  $t=t_1$  are:

$$H_0(t_0) = \mathbb{R} \rightarrow H_0(t_1) = \mathbb{R} \text{ (which is expected)}$$

$$H_1(t_0) = \mathbb{R} \rightarrow H_1(t_1) = 0 \text{ (the hole disappears)}$$

The changes in the sheaf cohomology groups from time  $t=t_0$  to time  $t=t_1$  are:

$$\text{Dimension } H^0(t_0) = 6 \rightarrow \text{Dimension } H^0(t_1) = 6$$

The dimension of the first cohomology group remains the same but the calculations show that the change in the stalks results in the change of the significant sensors from  $\{O, C\}$  to  $\{O, D\}$ .

## 4.2 Part 2: Example of Air Traffic Monitoring

Air traffic monitoring is one of the crucial complex systems to detect and estimate the location, velocity and flight direction of a large number of various airplanes approaching an airport. At an airport, multiple sensors of various types monitor the region. To make the detection more precise, consider duplication of the sensors of the same type. Consider cluster of GPS satellites, cluster of radar stations, cluster of airport surface detectors and cluster of smart IR (infrared) sensors for air traffic monitoring. Figures 4.5 and 4.6 show an air route and an air traffic and monitoring system. Numerous heterogeneous data acquisition must be integrated.



Figure 4.5. Air route and traffic control centers in the United States and its territories [58].

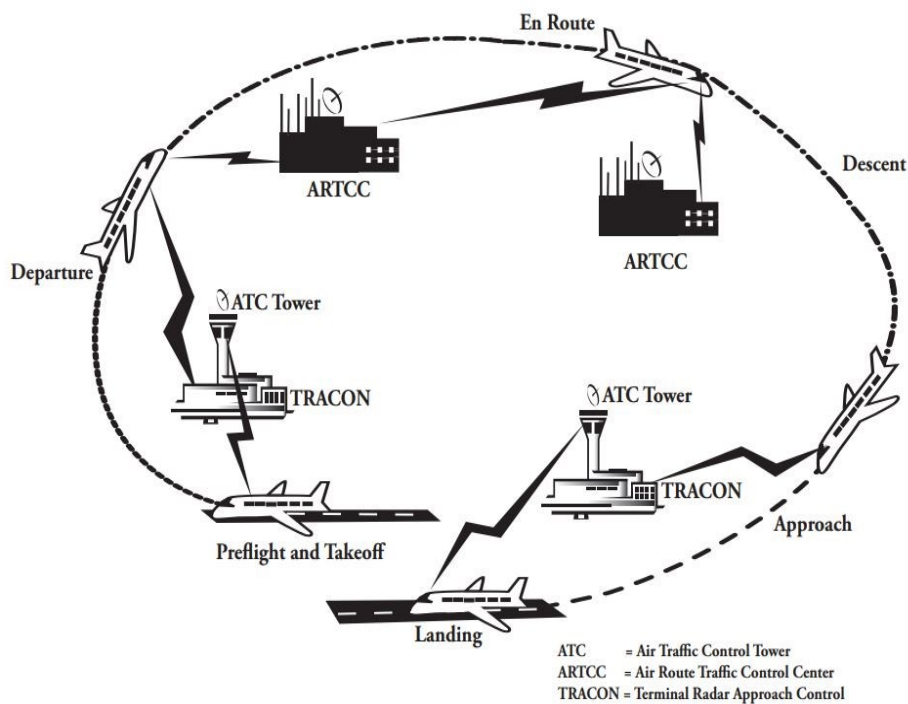


Figure 4.6. An example of air traffic monitoring system including air traffic control tower, air route traffic control center, and terminal radar approach control [58].

Sensors of the same type communicate and report common data, as shown in Table 4.5.

Table 4.5. Sensors and Duplication Numbers for Air Traffic Monitoring ( $t = t_0$ )

Sensor type	Number of sensors time $t=t_0$
Radars (R)	n
GPS (G)	m
Airport Surface Detectors(K)	p
IR Sensors (I)	q

1. Aircraft Status (E), Space of measurement =  $\mathbb{R}$
2. Aircraft Coordinates (C), Space of measurement =  $\mathbb{R}^3$
3. Direction (D), Space of measurement =  $\mathbb{R}^3$
4. Speed (S), Space of measurements =  $\mathbb{R}$

The heterogeneous data received at time  $t=t_0$  are given in the table 4.6. The measured subjects in the table are:

Table 4.6. The Heterogeneous Data for Air Traffic Monitoring ( $t = t_0$ )

Sensors vs. data	(E)	(C)	(D)	(S)
Radars (R)		✓	✓	✓
GPS (G)		✓		
Airport Surface Detectors (K)	✓		✓	
IR Sensors (I)	✓			✓

#### 4.2.1 The Construction of the Simplicial Complex

The oriented simplicial complex structure model is shown in Figure 4.7.

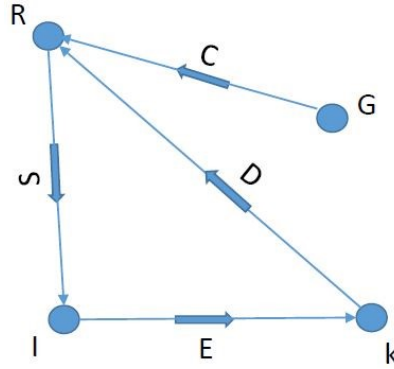


Figure 4.7. Simplicial complex model with oriented simplices for the air traffic monitoring at time  $t = t_0$ .

#### 4.2.2 The Sheaf Construction

Each simplex in the simplicial complex has a characteristic that is represented by assigning additional information to the simplex. To model this assignment, a stalk associated with the information is assigned to each simplex. It carries all of the information about the data and its neighboring nodes and enables the analysis of the system. The assigned spaces and the stalks are shown in Figure 4.8, as follows:

$$\text{Stalk } R = F(R) = \{\text{Aircraft Coordinates } \mathbb{R}^3, \text{Direction } \mathbb{R}^3, \text{Speed } \mathbb{R}\} \simeq \mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \mathbb{R}$$

$$\text{Stalk } G = F(G) = \{\text{Aircraft Coordinates } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } K = F(K) = \{\text{Aircraft Status } \mathbb{R}, \text{Direction } \mathbb{R}^3\} \simeq \mathbb{R} \oplus \mathbb{R}^3$$

$$\text{Stalk } I = F(I) = \{\text{Aircraft Status } \mathbb{R}, \text{Speed } \mathbb{R}\} \simeq \mathbb{R} \oplus \mathbb{R}$$

$$\text{Stalk } RG = F(RG) = \{\text{Aircraft Coordinates } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } RK = F(RK) = \{\text{Direction } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } RI = F(RI) = \{\text{Speed } \mathbb{R}\} \simeq \mathbb{R}$$

$$\text{Stalk } IK = F(IK) = \{\text{Aircraft Status } \mathbb{R}\} \simeq \mathbb{R}$$

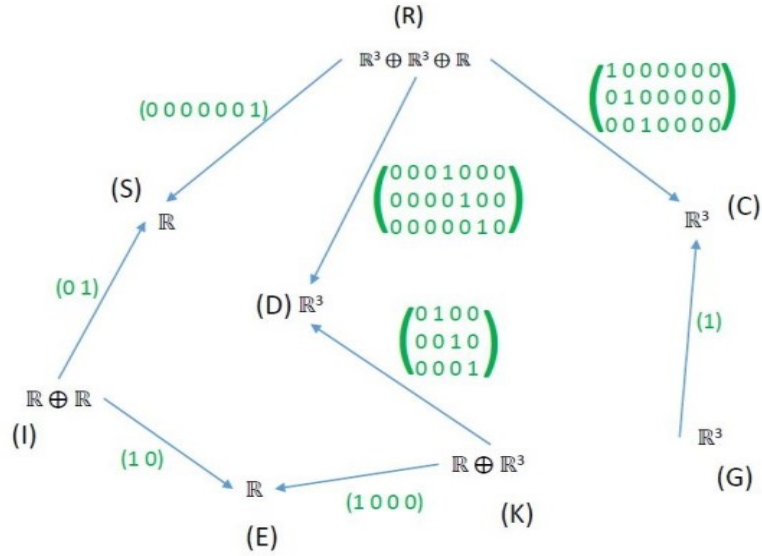


Figure 4.8. Sheaf, stalks and restriction maps associated with the simplicial complex for the air traffic monitoring example at time  $t = t_0$ .

#### 4.2.3 Homology and Sheaf Cohomology

Based on the algorithms for calculation of the homology groups in subsection 2.2.2, the following results at time  $t=t_0$  are obtained.

$H_0 = \mathbb{R}$  (dimension of  $H_0 = 1$ ), meaning one connected simplicial complex exists.

$H_1 = \mathbb{R}$  (dimension  $H_1 = 1$ ), meaning a 1-dimensional hole in this simplicial complex exists.

$H_n = 0$  for  $n > 2$ , meaning there are no voids in dimension bigger than 2D in this simplicial complex.

From the algorithm for calculation of the sheaf cohomology groups (data analysis) in subsection 3.3 it is seen that

Dimension  $H^0 = 8$ , meaning that at time  $t_0$  significant stalks are on R and I. The global information (section globalization) is extracted from the sensors R and I.



$H^1 = 0$ , meaning that all 1-integrating families come from sections on the sensors.

$H^n = 0$  ( $n > 1$ ), since there are no  $n$ -simplices for  $n > 1$ .

#### 4.2.4 Time Changes from $t=t_0$ to $t= t_1$

Suppose at time  $t=t_1$  the  $i$ -number of airport surface detectors are out of mission.

Table 4.5 has been changed to Table 4.7.

Table 4.7. Sensors and their Duplication Numbers for Air Traffic Monitoring ( $t = t_1$ )

Sensor type	Number of sensors time $t=t_0$
Radars (R)	n
GPS (G)	m
Airport Surface Detectors(K)	P - <b>i</b>
IR Sensors (I)	q

As a result the aircraft status (E) is no longer detected by the airport surface detectors (K). Table 4.8 shows the change that occurs in Table 4.6.

Table 4.8. The Heterogeneous Data for Air Traffic Monitoring ( $t = t_1$ )

Sensors vs. data	(E)	(C)	(D)	(S)
Radars (R)		✓	✓	✓
GPS (G)		✓		
Airport Surface Detectors (K)	<b>Out of mission</b>		✓	
IR Sensors (I)	✓			✓

The new simplicial complex is shown in Figure 4.9.

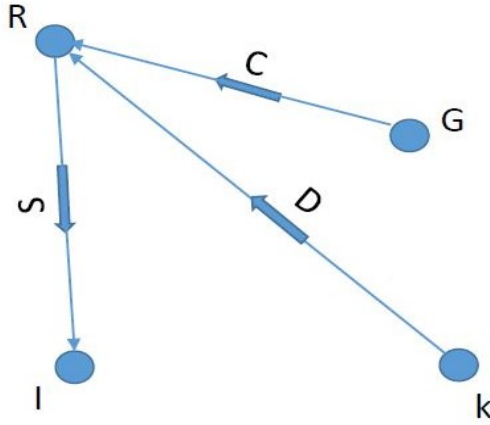


Figure 4.9. Simplicial complex model with oriented simplices for the air traffic monitoring at time  $t = t_1$ .

The new stalks are:

$$\text{Stalk } R = F(R) = \{\text{Aircraft Coordinates } \mathbb{R}^3, \text{ Direction } \mathbb{R}^3, \text{ Speed } \mathbb{R}\} \simeq \mathbb{R}^3 \oplus \mathbb{R}^3 \oplus \mathbb{R}$$

$$\text{Stalk } G = F(G) = \{\text{Aircraft Coordinates } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } K = F(K) = \{\text{Aircraft Status } \mathbb{R}, \text{ Direction } \mathbb{R}^3\} \simeq \mathbb{R} \oplus \mathbb{R}^3$$

$$\text{Stalk } I = F(I) = \{\text{Aircraft Status } \mathbb{R}, \text{ Speed } \mathbb{R}\} \simeq \mathbb{R} \oplus \mathbb{R}$$

$$\text{Stalk } RG = F(RG) = \{\text{Aircraft Coordinates } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } RK = F(RK) = \{\text{Direction } \mathbb{R}^3\} \simeq \mathbb{R}^3$$

$$\text{Stalk } RI = F(RI) = \{\text{Speed } \mathbb{R}\} \simeq \mathbb{R}$$

The sheaf, stalks and restriction maps for the simplicial complex at time  $t = t_1$  are shown in

Figure 4.10.

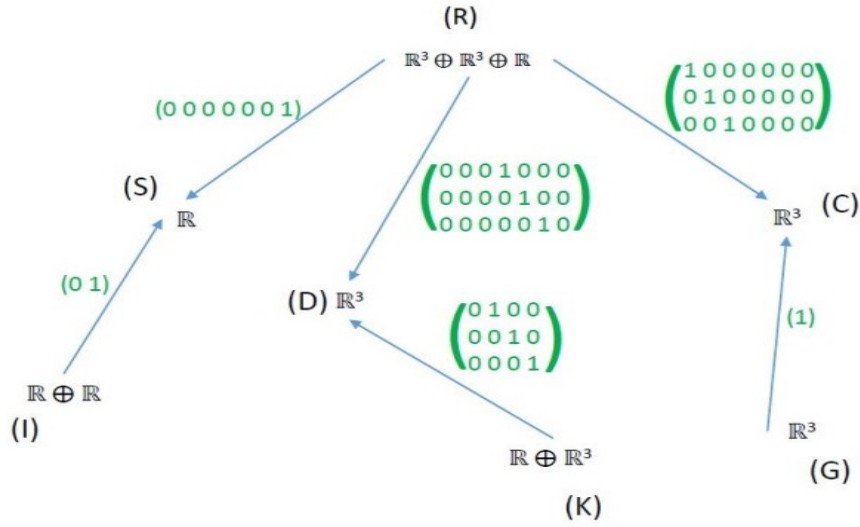


Figure 4.10. Sheaf, stalks and restriction maps associated with the simplicial complex for the air traffic monitoring example at time  $t=t_1$ .

Based on the algorithms for calculation of the homology groups in subsection 2.2.2, for the simplicial complex at time  $t=t_1$  the following results are obtained.

$H_0 = \mathbb{R}$  (dimension of  $H_0 = 1$ ), meaning that one connected simplicial complex exists.

$H_1 = 0$  (dimension  $H_1 = 0$ ), meaning that no 1-dimensional hole in this simplicial complex exists.

$H_n = 0$  for  $n > 2$ , meaning that no voids in dimension greater than 2D exist in this simplicial complex.

From the algorithms for calculation of the sheaf cohomology groups (data analysis) in subsection 3.3:

Dimension of  $H^0 = 9$ , meaning that at time  $t=t_1$  significant stalks are on K, G and I. The global information (section globalization) is extracted from the sensors K, G and I.

$H^1 = 0$ , meaning that all 1-integrating families come from sections on the sensors.

$H^n = 0$  ( $n > 1$ ), since there are no  $n$ -simplices for  $n > 1$ .

#### 4.2.5 Feedback from the Example

By the change in the number of sensors, when some sensors become inactive or broken, the changes in the simplicial complex homology and the sheaf cohomology, from time  $t=t_0$  to time  $t=t_1$ , occur. As a result the significant sensors are changed:

$$H_0(t_0) = \mathbb{R} \rightarrow H_0(t_1) = \mathbb{R} \text{ (as expected)}$$

$$H_1(t_0) = \mathbb{R} \rightarrow H_1(t_1) = 0 \text{ (the hole disappears)}$$

$$\text{Dimension } H^0(t_0) = 8 \rightarrow \text{Dimension } H^0(t_1) = 9$$

The change in the stalks results in the change of the significant sensors from  $\{R, I\}$  to  $\{K, G, I\}$ , and also the change in the dimension of the zero cohomology group.

### 4.3 Summary

Applications of the methodology in the previous chapters are described by the two case studies: one from the wildfire threat monitoring and the other from the air traffic monitoring.

Both cases are distributed information systems that deal with temporal and spatial fusion of heterogeneous data obtained from multiple sources, where the schema, availability, and quality vary.

Behavior of both systems is explained thoroughly in terms of the detection of the failure in the system. The redundant and complimentary sensors are recognized.

The mathematical foundations in Chapter 3 prove the validity of these processes.

## CHAPTER 5

### ALTERNATIVE SOLUTION

This chapter is devoted to the comparison between the sheaf theoretic method and the alternative method that does not apply the sheaf theory. Without utilizing the sheaf theory method, multiple tables are required to extract data from sensors. In the following, the two methods are compared in terms of time and space complexity. It is found that when the data are more heterogeneous the sheaf theory method makes the solution less complex with respect to time and space.

#### 5.1 Solving the Fire Monitoring with Alternative Tools

To address the wildfire monitoring, the construction of a grid of measured points for  $p$  types of sensors is required. This is of order  $O(n)$ . In this case  $p=4$ . Consider the following thresholds for the measurement of each sensor:

Satellite Camera = Sat\_Threshold

CO2 Detector = CO2\_Threshold

IR Detector = IR\_Threshold

Flame Detector = Flame\_Threshold

For the sake of simplicity in addressing the general issue, consider the region of interest to be rectangular. At time  $t = t_0$  consider the tables 5.1, 5.2, 5.3 and 5.4, each with  $H$  rows and  $G$  columns for reporting the sensor measurements.

Table 5.1. Signals from Satellite Cameras,  $\text{dim} = H \times G$ , complexity =  $O(n^2)$ 

S11	S12	S13	...	...	...	...						S1G
S21	S22	S23	...	...	...	...						S2G
...												
....												
...												
SH1	SH2	SH3	...	...	...	...						SHG

Each cell in Table 5.1 acquires a measurement from satellite cameras. These measurements will be compared with Sat\_Threshold.

Table 5.2. Signals from CO2 Detectors,  $\text{dim} = H \times G$ , complexity =  $O(n^2)$ 

C11	C12	C13	...	...	...	...						C1G
C21	C22	C23	...	...	...	...						C2G
...												
....												
...												
CH1	CH2	CH3	...	...	...	...						CHG

Each cell in Table 5.2 acquires a measurement from CO2 detectors. These measurements will be compared with CO2\_Threshold.

Table 5.3. Signals from IR Detectors,  $\text{dim} = H \times G$ , complexity =  $O(n^2)$ 

R11	R12	R13	...	...	...	...						R1G
R21	R22	R23	...	...	...	...						R2G
...												
....												
...												
RH1	RH2	RH3	...	...	...	...						RHG

Each cell of Table 5.3 acquires a measurement from IR detectors. These measurements will be compared with IR\_Threshold.

Table 5.4. Signals from Flame Detectors,  $\text{dim} = H \times G$ , complexity =  $O(n^2)$ 

F11	F12	F13	...	...	...	...						F1G
F21	F22	F23	...	...	...	...						F2G
...												
....												
...												
FH1	FH2	FH3	...	...	...	...						FHG

Each cell of Table 5.4 acquires a measurement from flame detectors. These measurements will be compared with Flame\_Threshold.

### 5.1.1 The Pseudocode to Confirm the Fire in Each Cell

```

For i = 1 to p ← p number of sensor types
  For j = 1 to H ← number of rows
    For k = 1 to G ← number of columns
      Compare (i,j,k) >= sensor thresholds

```

```

        If (yes) then fire
        Else no fire
    End
End
End
End

```

Results from the pseudocode for this alternative solution show that in terms of space complexity the order is  $O(n^3)$  and in terms of time complexity the order is  $O(n^3)$ . By comparing with sheaf and topology methods:

Space complexity of sheaf topology ( $O(n^2)$ ) < Space complexity of alternative method ( $O(n^3)$ ). Time complexity of sheaf topology ( $O(n^{2.5})$ ) < Time complexity of alternative method ( $O(n^3)$ ).

The only time the alternative method gives a better time and space complexity is when there is only one homogeneous sensor type. When  $p = 1$ :

Space complexity of sheaf topology ( $O(n^2)$ ) = Space complexity of alternative method ( $O(n^2)$ ). Time complexity of sheaf topology ( $O(n^{2.5})$ ) > Time complexity of alternative method ( $O(n^2)$ ). Based on the sensors measurements the results of existence of fire is in Table 5.5:

Table 5.5. Existence of Fire Based on Sensors Measurements Time  $t = t_0$

11	12	13	...	...	...	...						1G
21	22	23	...	...	...	...						2G
...												
....												
...												
H1	H2	H3	...	...	...	...						HG



Based on the reports from the fire department the results of existence of fire is in Table 5.6:

Table 5.6. Existence of Fire Based on Reporting from Fire Department Time  $t = t_0$

11	12	13	...	...	...	...						1G
21	22	23	...	...	...	...						2G
...												
....												
...												
H1	H2	H3	...	...	...	...						HG

It is obvious that some of the sensors are reporting wrong signals due to the defection or broken. A comparison of the two tables, cell by cell, yields information about the defective sensors and also shows which sensors cover the region and report correct information. The lower bound for this procedure is of order  $O(n^3)$ .

With the application of the sheaf theory approach in heterogeneous sensors, the time complexity is of order  $O(n^{2.5})$ . This is better than the time complexity of alternative method which is of order  $O(n^3)$ . The space complexity from sheaf theory method is  $O(n^2)$ , which is also better than space complexity of alternative method which is of order  $O(n^3)$ .

## 5.2 Summary

The comparison between the sheaf theory and the alternative methodologies is described to present further proof of the validity of the sheaf theory method.

It is shown that when the nature of the data is more heterogeneous, the sheaf theory method has less computational complexity in both space and time.

## CHAPTER 6

### NOISE AND INCONSISTENCY

#### 6.1 Consistency Radius

In sheaf theory when some assignments as local sections are inconsistent, the “Consistency Radius” emerged. The question is: “Are there any error detections and corrections to correct the discrepancy in the sheaf theory?” The answer is YES. There is a way to do some error detection and correction in a sheaf. This is how it works.

The consistency radius is the maximum distance between the value in a stalk and the values propagated along the restriction maps [59]. If an assignment consistency radius is not zero, it is definitely not a global section. Yet, if the sheaf model is trusted as being accurate, only the global sections should (in principle) be observed. Thus, what should be done is to find the global section that is nearest (in the appropriate assignment metric) to the assignment. That will typically replace all the values in the assignment with "better" ones. This approach often has been found to work quite well. Indeed, it seems to eliminate some standard algorithms for signal separation, which is an ongoing problem.

The downside is that the optimization problem to minimize the distance between the global section and the given assignment needs to be solved. Although it may not be easy to solve, in relatively simple cases, a straightforward “constrained least squares” might do the job. But this needs to be resorted to genetic algorithms that are commonly used to generate high-quality solutions to optimize and search problems. This is still an area that is open to research, since the problem is usually encoded as sheaves in several distinct ways. Different optimization problems are obtained, which may or may not vary

in how easy they are to solve. The occurrences of the discrepancy in sheaf model can be seen in Figures 6.1 through 6.4.

For error detection and error correction there are off the shelf approaches such as coding by Hamming, Huffman, Reed-Solomon, and Berlekamp-Massay [48], which give EBR (Error Bit Rate)  $1/10^9$ . Reaching to the lower EBR is another open research area.

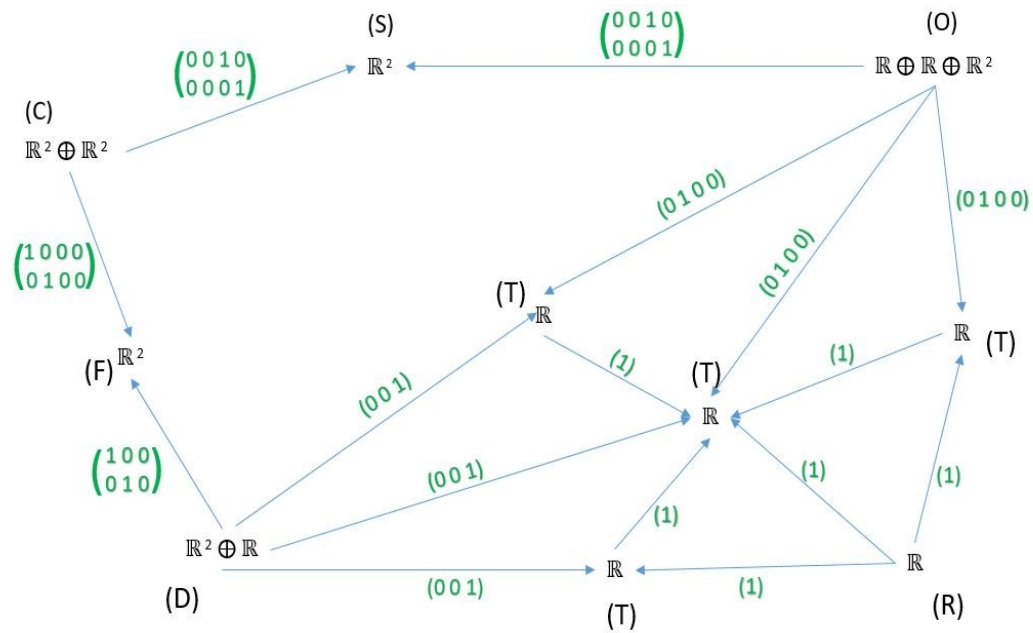


Figure 6.1. Sheaf of vector spaces on the partial order set associated with the example of wildfire threat monitoring system. The diagram commutes.

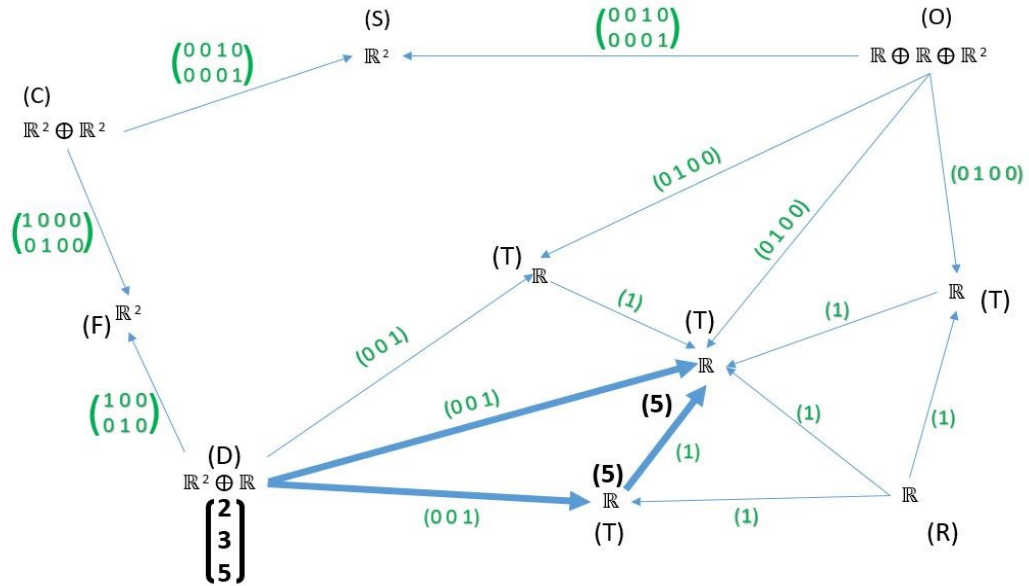


Figure 6.2. Relating to the example of wildfire threat monitoring a global section is an assignment that is consistent with restrictions.

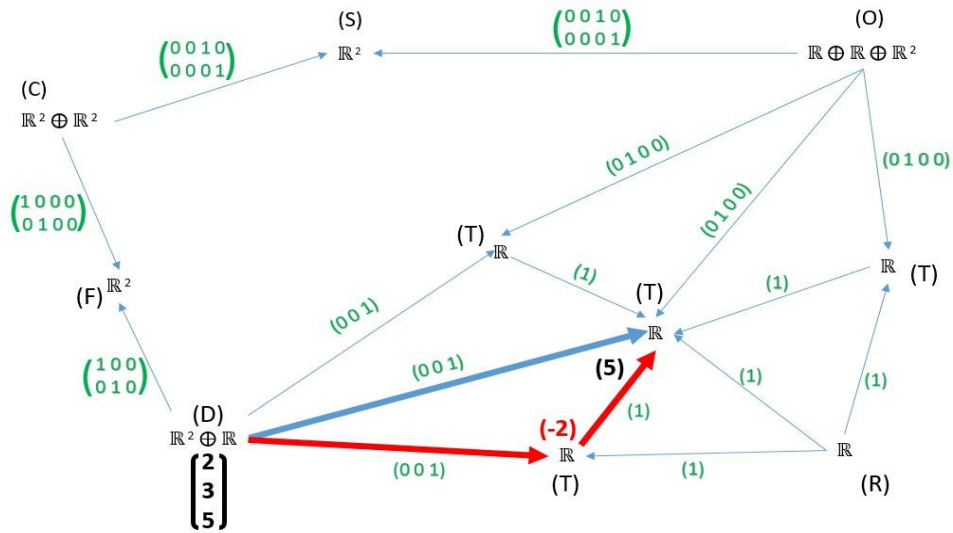


Figure 6.3. Relating to the example of wildfire threat monitoring due to noise some assignments are not consistent. They are partially consistent.

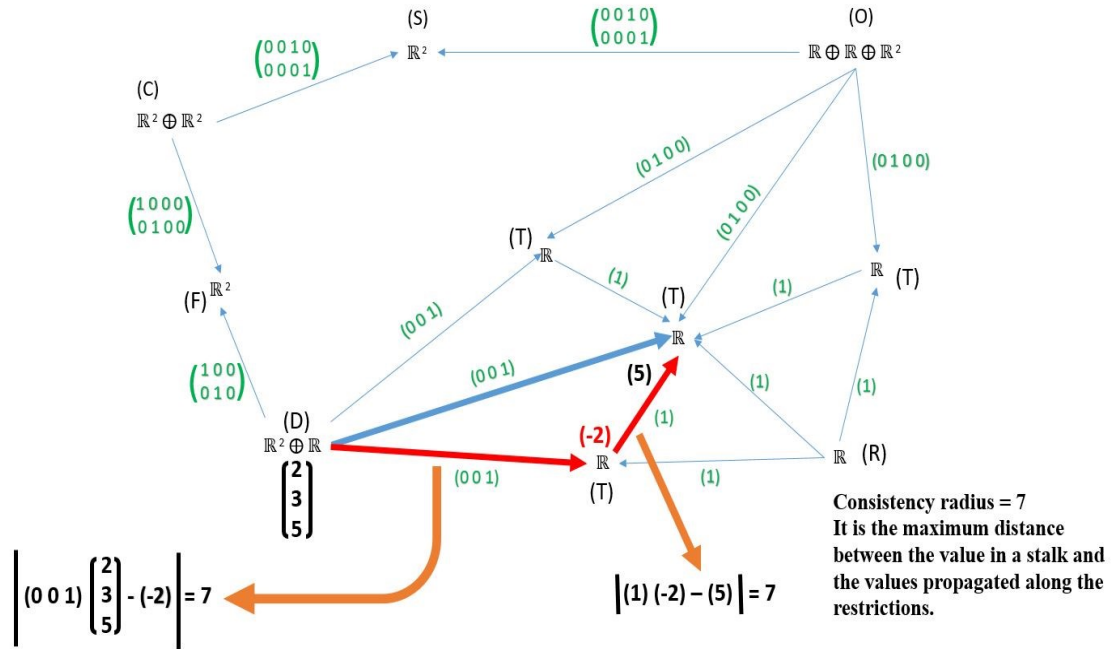


Figure 6.4. Consistency radius is the maximum distance between the value in a stalk and the values propagated along the restrictions.

## 6.2 How to Achieve the Desirable Consistency Radius

The method to achieve the desirable consistency radius is to deploy the supervised data input to the sensor integration, measuring the consistency radius and finding out the data quality estimation. The desirable consistency radius is obtained by calibrating the hardware, feeding these results as new input to the system and repeating the cycle until the desirable consistency radius threshold is obtained.

### **6.3 Summary**

This chapter is devoted to the case study in which the noise causes the sheafification of the system to be disturbed. Methods to detect error and make the corrections are stated. The noise is described from the consistency of the stalk assignments. The feedback process to achieve the desirable consistency radius is also discussed.

## CHAPTER 7

## CONCLUSION AND FUTURE WORK

## 7.1 The Feedback Process, Figure 7.1.

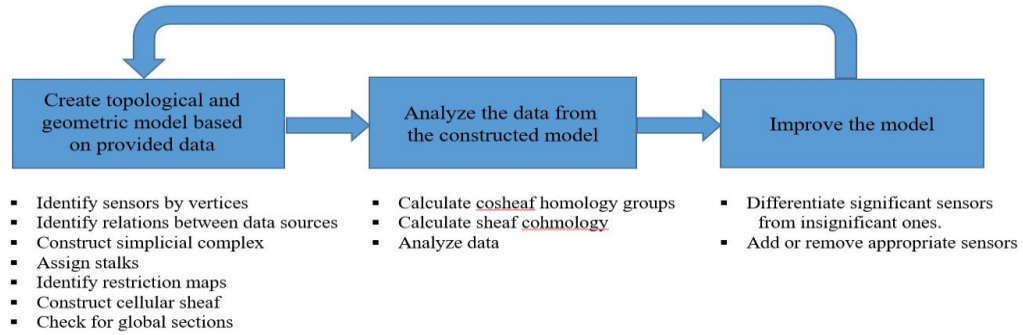


Figure 7.1. The feedback process.

## 7.2 Summary and Future Work

The focus of this dissertation research is to model temporal and spatial heterogeneous data fusion. The software utilized for computation of the matrix rank and the image and kernel is “MATLAB”. Some open problems are recommended for future work. Among them are:

- Addressing large or varied datasets (stalks)
- Statistical behavior of heterogeneous data fusion
- Dynamical persistence of sheaves
- Using machine learning technique to automate suggestions for the addition, removal, or changing of sensors
- Concept of cosheaf and sheaf, cosheaf duality [60] and [61]
- Simplicial complex  $\rightarrow$  Sheaf is done, what about sheaf  $\rightarrow$  Simplicial complex?



## REFERENCES

- [1] J. Thomas and L. Sael, "Overview of integrative analysis methods for heterogeneous data," *2015 International Conference on Big Data and Smart Computing (BIGCOMP)*, IEEE Computer Society, pp. 266-270, February 2015.
- [2] P. S. Medina and R. W. Doerge, "Statistical methods in topological data analysis for complex, high-dimensional data," in *27th Annual Conference on Applied Statistics in Agriculture*, Kansas State University Libraries, New Prairie Press, 2015.
- [3] M. Robinson, C. Joslyn, E. Hogan and C. Caoraro, "Conglomeration of heterogeneous content using local topology (CHCLT)," American University, Technical Report No. 2015-1, 2015.
- [4] V. Snasel, J. Nowakova, F. Xhafa and L. Barolli, "Geometrical and topological approaches to big data," *Future Generation Computer Systems*, vol. 67, pp. 286-296, 2017.
- [5] E. Munch, "A user's guide to topological data analysis," *Journal of Learning Analytics*, vol. 4, no. 2, pp. 47-61, 2017.
- [6] M. Lesnick, "Studying the Shape of Data Using Topology," *Institute for Advanced Studies (IAS) letter*, Princeton University Press, pp. 10-11, Summer 2013.
- [7] G. Carlsson, "Why topological data analysis works," AYASDI Technical Publication, January 2015.
- [8] S. Y. Moore, *The practice of statistics*, 4th ed., New York: W. H. Freeman and Company, 2012.
- [9] J. Ouellette, "The mathematical shape of things to come," *QuantamaGazine*, 4 October 2013.
- [10] P. Awasthi, A. S. Bandeira, M. Charikar, R. Krishnaswamy, S. Villar and R. Ward, "Relax, no need to round: Integrality of clustering formulations," in *ITCS '15, Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, 2015.
- [11] Y. Xu, "A general comparison between tokamak and stellarator plasmas," *Matter and Radiation at Extremes*, vol. 1, no. 4, pp. 192-200, 2016.

- [12] L. Li, W. Y. Cheng, B. S. Glicksberg, O. Gottesman, R. Tamler, R. Chen, E. P. Bottinger and J. T. Dudley, "Identification of type 2 diabetes subgroups through topological analysis of patient similarity," *Science Translational Medicine*, vol. 7, no. 311, p. 311ra174, October 2015.
- [13] S. Reddy, The Vietoris–Rips complexes of finite subsets of an ellipse of small eccentricity, Honors Thesis, Duke University, 2017.
- [14] K. Almgren, M. Kim and J. Lee, "Extracting knowledge from the geometric shape of social network data using topological data analysis," *Entropy*, vol. 19 (7), no. 360, 2017.
- [15] P. Y. Lum, G. Singh, A. Lehman, T. Ishkhanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson and G. Carlsson, "Extracting insights from the shape of complex data using topology," *Nature, Scientific Report*, vol. 3, no. 1236, 2013.
- [16] J. Feng, M. Gou and H. Hu, "Graph-based multi-sensor fusion for event detection," *International Journal of Control and Automation*, vol. 8, no. 10, pp. 145-154, 2015.
- [17] M. Andjelkovic, B. Tadic, S. Maletic and M. Rajkovic, "Hierarchical sequencing of online social graphs," *Physica A: Statistical Mechanics and its Applications, Elsevier*, vol. 436, pp. 582-595, October 2015.
- [18] S. Maletic and M. Rajkovic, "Consensus formation on a simplicial complex of opinions," *Physica A: Statistical Mechanics and its Applications, Elsevier*, vol. 397, pp. 111-120, March 2014.
- [19] H. Bacelar-Nicolau, F. Nicolau and A. Sousa, "Measuring similarity of complex and heterogeneous data in Clustering of Large datasets," *Biocybernetics and Biomedical Engineering, Nałecz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences*, vol. 29, no. 2, pp. 9-18, 2009.
- [20] C. Joslyn, E. Hogan and M. Robinson, "Towards a topological framework for integrating semantic information sources," in *semantic technology for intelligence, defense and security (STIDS)*, 2014.
- [21] F. Casanedo, "A review of data fusion techniques," *The Scientific World Journal, Hindawi Publishing Corporation*, vol. 2013, 2013.
- [22] E. Bareinboim and J. Pearl, "Casual inference and the data-fusion problem," *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, vol. 113, pp. 7345-7352, July 2016.

- [23] R. Zuech, T. M. Khoshgoftaar and R. Wald, "Intrusion detection and Big Heterogeneous Data: a Survey," *Journal of Big Data, Springer open Journal*, vol. 2:3, 2015.
- [24] G. L. Cattani and G. Winskel, "Presheaf models for concurrency," *Lecture Notes in Computer Science; Computer Science Logic, Springer, Berlin, Heidelberg*, vol. 1258, pp. 58-75, 1996.
- [25] Y. V. Srinivas, "A sheaf-theoretic approach to pattern matching and related problems," *Theoretical Computer Science, Elsevier*, vol. 112, no. 1, pp. 53-97, 1993.
- [26] H. D. Ehrich, J. A. Goguen and A. Sernadas, "A categorial theory of objects as observed processes," *Lecture Notes in Computer Science (LNCS), Springer, Berlin, Heidelberg*, vol. 489, pp. 203-228, 2005.
- [27] J. A. Goguen, "Sheaf semantics for concurrent interacting objects," *Mathematical Structures in Computer Science, Cambridge University Press*, vol. 2, no. 2, pp. 159-191, June 1992.
- [28] G. Malcolm, "Sheaves, objects and distributed systems," *Electronic Notes in Theoretical Computer Science, Elsevier*, vol. 225, pp. 3-19, 2009.
- [29] M. Robinson, "Understanding networks and their behaviors using sheaf theory," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Austin TX, 2013.
- [30] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.
- [31] F. H. Croom, *Basic Concepts of Algebraic Topology*, Springer-Verlag New York, 1978.
- [32] J. Binchi, E. Merelli, M. Ruco, G. Petri and F. Vaccarino, "jHoles: A tool for understanding biological complex networks via clique weight rank persistent homology," *Electronic Notes in Theoretical Computer Science, Elsevier*, vol. 306, pp. 5-18, 2014.
- [33] P. S. Alexandroff, "Uber den allgemeinen dimensionsbegriff und seine beziehungen zur elementaren geometrischen anschauung," *Mathematische Annalen*, vol. 98, pp. 617-635, 1928.
- [34] M. Doktorova, *Constructing simplicial complexes over topological spaces*, Hanover, New Hampshire: Dartmouth Computer Science Technical Report TR2012-721, Dartmouth College, 2012.

- [35] R. Ghrist and A. Muhammad, "Coverage and hole-detection in sensor networks via homology," in *4th International symposium, Information processing in sensor networks, IEEE Xplore*, 2005.
- [36] J. Kanno, J. G. Buchart, R. R. Selmic and V. Phoha, "Detecting coverage holes in wireless sensor networks," in *IEEE 17th Mediterranean Conference on Control and Automation, Makedonia Palace, Thessaloniki, Greece*, 2009.
- [37] V. D. Silva and R. Ghrist, "Homological sensor networks," *Notices of the American Mathematical Society*, pp. 10-17, 2007.
- [38] M. Mrozek and B. Batko, "Coreduction homology algorithm," *Discrete Comput. Geom.*, vol. 41, pp. 96-118, 2009.
- [39] P. Dlotko, R. Ghrist, M. Juda and M. Mrozek, "Distributed computation of coverage in sensor networks by homological methods," *Applicable Algebra in Engineering Communication and Computing, Springer*, vol. 23, no. 1-2, pp. 29-58, 2012.
- [40] "github," April 2013. [Online]. Available: <https://github.com/j2kun/computing-homology>.
- [41] C. Giusti, R. Ghrist and D. Bassett, "Two's company, three (or more) is a simplex: algebraic-topological tools for understanding higher-order structure in neural data," *Journal of Computational Neuroscience*, vol. 41, no. 1, pp. 1-14, 2016.
- [42] R. Ghrist, "Three examples of applied and computational homology," *Nieuw Archief voor Wiskunde*, vol. 5/9, no. 2, 2008.
- [43] B. R. Tennison, Sheaf theory, London Mathematical Society Lecture Note Series No. 20, Cambridge University Press, 1975.
- [44] L. Kastner, K. Shaw and A. L. Winz, "Cellular Sheaf Cohomology in Polymake," in *Combinatorial Algebraic Geometry*, vol. 80, G. G. Smith and B. Sturmfels, Eds., Fields Institute Communications book series, Springer, 2017, pp. 369-385.
- [45] M. Vybornov, "Constructible Sheaves on Simplicial Complexes and Kosul Duality," *Mathematical Research Letters*, vol. 5, pp. 675-683, 1998.
- [46] J. M. Curry, "Topological data analysis and cosheaves," *Japan Journal of Industrial and Applied Mathematics, Springer*, vol. 32, no. 2, pp. 333-337, 2015.
- [47] J. H. Hubbard, Teichmüller Theory and Applications to Geometry, Topology, and Dynamics, vol. 1, Matrix Editions, 2006.

- [48] M. Robinson, Topological signal processing, Mathematical engineering, Springer-Verlag Berlin Heidelberg, 2014.
- [49] J. Gallier and J. Quaintance, A gentle introduction to homology, cohomology and sheaf Cohomology, University of Pennsylvania Press, 2016.
- [50] K. Ueno, Algebraic geometry 2: sheaves and cohomology, vol. 197, Transactions of Mathematical Monographs, American Mathematical Society, 2001.
- [51] G. Harder, Lectures on Algebraic Geometry I: sheaves, cohomology of sheaves and applications to Riemann surfaces, 2nd ed., Vieweg Teubner Verlag, Springer Fachmedien Wiesbaden GmbH, 2011.
- [52] M. Robinson, "Asynchronous logic circuits and sheaf obstructions,," *Electronic Notes in Theoretical Computer Science, Elsevier*, vol. 283, pp. 159-177, 2012.
- [53] P. Dlotko, "A fast algorithm to compute cohomology group generators of orientable 2-Manifolds," *Pattern Recognition Letters, Elsevier*, vol. 33, no. 11, pp. 1468-1476, 2012.
- [54] P. Dlotko and R. Specogna, "A novel technique for cohomology computations in engineering practice," *Comput. Methods Appl. Mech. Engrg., Elsevier*, vol. 253, pp. 530-542, 2013.
- [55] J. Curry, R. Ghrist and V. Nanda, "Discrete morse theory for computing cellular sheaf cohomology," *Foundations of Computational Mathematics, Springer*, vol. 16, no. 4, pp. 875-897, 2016.
- [56] S. Mansourbeigi, "Sheaf theory as a mathematical foundation for distributed applications involving heterogeneous datasets," in *The 32nd IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA-2018)*, Cracow, Poland, 2018.
- [57] S. Mansourbeigi, "Sheaf theory approach to distributed applications: analyzing heterogeneous data in air traffic monitoring," *International Journal of Data Science and Analysis, Science Publishing Group*, vol. 3, no. 5, pp. 34-39, 2017.
- [58] L. Wilson, Eyes in the sky, D. Dahlseide, Ed., Minnesota Department of Transportation, Office of Aeronautics, Aviation Education Section, 2006.
- [59] M. Robinson, "Sheaves are the canonical data structure for sensor integration," *Information Fusion, Elsevier*, vol. 36, pp. 208-224, July 2017.

- [60] J. M. Curry, "Dualities between cellular sheaves and cosheaves," *Journal of pure and applied algebra, Elsevier*, vol. 222, pp. 966-993, 2018.
- [61] J. M. Curry, Sheaves, cosheaves and applications, Doctoral Thesis , University of Pennsylvania, 2014.

## APPENDICES

## Sheaf Theory Approach to Distributed Applications: Analysing Heterogeneous Data in Air Traffic Monitoring

Seyed Mansourbeigi

Department of Computer Science, College of Engineering, Utah State University, Logan, USA

Email address:

smansourbeigi@aggiemail.usu.edu, phy\_math\_ee@yahoo.com

To cite this article:

Seyed Mansourbeigi. Sheaf Theory Approach to Distributed Applications: Analysing Heterogeneous Data in Air Traffic Monitoring. *International Journal of Data Science and Analysis*. Vol. 3, No. 5, 2017, pp. 34-39. doi: 10.11648/ijds.20170305.11

Received: September 6, 2017; Accepted: September 23, 2017; Published: October 23, 2017

**Abstract:** The goal of the present article is to demonstrate a mathematical modeling for distributed applications. The present paper applies tools from topology and sheaf theory as an appropriate mathematical modeling to reflect interactions among elements of distributed applications resources. Sensors are characterized from their topological representations in distributed network system. This modeling is applied for the study of the air traffic monitoring system and discuss the model in detail.

**Keywords:** Cellular Sheaf, Stalks, Cosheaf Homology, Sheaf Cohomology

### 1. Introduction

The biggest engineering problems are fun math problems. So if you have hard engineering problem and you cannot crack, it almost always has a neat mathematics problem buried under there.

Distributed applications are applications or software that runs on multiple computers within a network at the same time and can be stored on servers. Data management is a key aspect of any distributed system. One of the main challenges in today's computer science research is the extraction of information from heterogeneous datasets. There have been numerous research that have shown the impact of mathematics in network modeling. The challenge is to understand how data is organized by turning data into information; information into knowledge; and eventually knowledge into wisdom. Geometry and topology are the natural modern approaches to handle complex heterogeneous data. The computations presented in this paper yields towards a bridge between modern geometry, topology and distributed systems and aims to introduce methods based on geometry and topology to detect and manage particular structures of the complex system. In recent years there has been researches on the application of sheaf theory to provide a semantic foundation for distributed applications [1] [2]. A sheaf can be thought of as a system of observations on a topological space, with the key property that consistent local observations can be uniquely pasted together to provide a global observation.

Application of sheaf theory in computer science has a long historical track. An early use of sheaf theory was a paper by Monteiro and Pereira [3]. They applied sheaf theory to study connections between event systems. As a foundation for the behavior of concurrent processes Ehrlich, Goguen and Semadas [4] and Goguen [5] and Cattani, G. L. and G. Winskel [6] applied sheaf theory. Cirstea [7] provided the semantics for a concurrent object-oriented programming language using sheaf theory. The motivation in this paper is inspired by the very recent applications of sheaf theory in computer science and software engineering. These applications can be found in [1] [2] [8] [9].

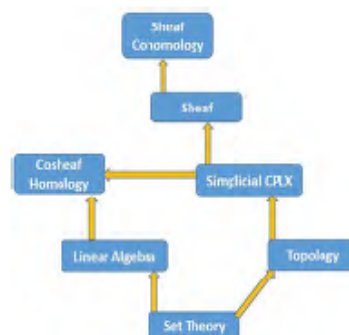


Figure 1. Road Map towards the Creation of the Modeling.



# Sheaf Theory as a Mathematical Foundation for Distributed Applications Involving Heterogeneous Data Sets

Seyed M-H Mansourbeigi  
College of Engineering  
Department of Computer Science  
Utah State University  
Logan, Utah, USA

smansourbeigi@aggiemail.usu.edu; phy\_math\_ee@yahoo.com

**Abstract**— The goal of the present article is to demonstrate a mathematical modeling for distributed applications. The present paper applies tools from topology and sheaf theory as an appropriate mathematical modeling to reflect interactions among elements of distributed applications resources. Simplicial complexes are topological models for the network structure. Behavior of the objects in distributed network system are represented by sheaves. This modelling is applied for the study of the wild fire threat.

**Keywords**— *Distributed Applications; Sheaf; Simplicial Complex, Stalks, Cosheaf Homology, Sheaf Cohomology*

## I. INTRODUCTION

Distributed applications are end-user systems consisting of software components running on multiple host machines that share resources and coordinate their actions to complete a task (or tasks) through message passing. They exist in nearly every industry and corner of the society, from social media to weather forecasting to grocery shopping. They differ from standalone applications in that they must handle partial failures, cope with unpredictable message transfer times, coordinate tasks without a global clock, and address a wide range of temporal challenges [1]. In addition, distributed applications often integrate data from multiple sources that differ in terms of their schemas, granularity, and quality. An example of such a system is a wildfire detection application that gathers data from a variety of in-the-field stations, each with a potentially different set of sensors for temperature, wind, humidity, smoke, and hotspots in the infrared spectrum. The individual stations may come online or go offline at irregular intervals and stations may become permanently disabled. Such a system might also make use of satellite images or aerial photography. Therefore, the structure, availability, granularity, and quality of the data would vary by data source and type.

All these and other issues tend to make distributed application complex. Furthermore, as their features become more sophisticated, the underlying complexities increase well beyond what current modeling languages and tools can manage. Therefore, software and data engineers need richer modeling tools that can help them express the characteristics of distributed applications, reason about those characteristics, and develop solutions that can optimize performance, resource utilization, and data quality. These tools should also support system-wide

analytics for uncovering emerging or changing characteristics, particularly in the presence of heterogeneous data coming from multiple sources with variable availability and quality. Section II provides some additional background on distributed applications and their characteristics, as well as a summary of tools commonly used for their modeling. Section II also discusses the problem integrating heterogeneous temporal data in more detail.

This paper proposes a conceptual modeling technique that addresses the problem of modeling and reasoning about distributed applications that work with heterogeneous data from multiple sources, where the schema, availability, and quality vary. This proposed technique uses Sheaf Theory; a mathematical tool for tracking data attached to open sets in a topological space; and prescribes a method for expressing a distributed application in terms of simplicial complexes as the topological data space equipped with stalks as space of information and the restriction maps to construct the associated cellular sheaf. Section III provides a summary of Sheaf Theory, prior to giving the explanation of the proposed modeling technique in Section IV. This section illustrates how this modeling technique can be used to capture essential characteristics of the wildfire application and answer questions, like a) are the stations providing sufficient information to track a real fire even when some of the stations may go offline, b) which types of data are redundant or complimentary, and c) are there “holes” in the spatial or temporal dimensions.

A modeling technique that can help engineers answer such questions could become a powerful tool for creating more advanced and adaptive distributed applications. Specifically, it should help engineers deal with complex issues arising from heterogeneous spatial and temporal data. Section V discusses these and other potential benefits in more detail, as well as current limitations and inefficiencies. Future research efforts aim to minimize the limitations and inefficiency. These are outlined in Section VI, along with some a summary of this paper’s contributions.

## II. BACKGROUND

Distributed applications are software systems with components running on two or more independent host machines and that communicate with each other via message passing over

## Geometry of physical systems on quantized spaces

Vida Milani<sup>\*,‡</sup>, Seyed M. H. Mansourbeigi<sup>†,§</sup> and  
 Stephen W. Clyde<sup>†,¶</sup>

<sup>\*</sup>*Department of Mathematics and Statistics*  
*Utah State University, Logan Utah, UT 84322, USA*

<sup>†</sup>*Department of Computer Science*  
*Utah State University, Logan Utah, UT 84322, USA*

<sup>‡</sup>*vida\_milani@yahoo.com*

<sup>§</sup>*smansourbeigi@aggiemail.usu.edu*

<sup>¶</sup>*swc@usu.edu*

Received 9 November 2016

Accepted 17 November 2016

Published 3 January 2017

We present a mathematical model for physical systems. A large class of functions is built through the functional quantization method and applied to the geometric study of the model. Quantized equations of motion along the Hamiltonian vector field are built up. It is seen that the procedure in higher dimension carries more physical information. The metric tensor appears to induce an electromagnetic field into the system and the dynamical nature of the electromagnetic field in curved space arises naturally. In the end, an explicit formula for the curvature tensor in the quantized space is given.

**Keywords:** Quantized space; quantized equations of motion; functional quantization; curved space; Q-meromorphic functions; electromagnetic field; metric tensor.

Mathematics Subject Classification 2010: 81T75, 53D55, 81R60

### 1. Introduction

Developments in quantum mechanics resulted in the discovery of non-commutative framework of mathematical models for physical systems [6]. The non-commutative version of the standard study of smooth manifolds lies in the representation of spaces by non-commutative function algebras [1, 3, 4]. A mathematical approach in transition to non-commutative formulation is through quantization of commutative algebras; assuming an appropriate set of non-commutative variables spanning a representation space [2, 9]. The quantized algebras provide appropriate models for physical systems; the physical concepts on these algebras can be well treated and the calculations can be simplified. The choice of the algebra varies from theory to theory. Different types of quantization provide different models for quantum theories [6, 13].



## COFIBRATIONS IN THE CATEGORY OF NONCOMMUTATIVE CW COMPLEXES

V. MILANI, S. M. H. MANSOURBEIGI AND A.-A. REZAEI

**ABSTRACT.** Cofibration in the category of noncommutative CW complexes is defined. The  $C^*$ -algebraic counterparts of topological mapping Cylinder and mapping cone are presented as examples of noncommutative CW complex cofibres. As a generalization, the concepts of noncommutative mapping cylindrical and conical telescope are introduced to provide more examples of noncommutative CW complex cofibres. Their properties and K-theoretic behavior are also studied in detail. It is seen that they carry the properties similar to the topological properties of their CW complex counterparts.

### 1. INTRODUCTION

The category of  $C^*$ -algebras and  $*$ -homomorphisms can be interpreted as the noncommutative counterpart of the category of topological spaces and continuous maps [1, 2, 8]. Its origin goes back to the Gelfand duality. The results of the paper [7] known as the Gelfand-Naimark theorem provide a duality between the topology of locally compact spaces and the algebraic structure of commutative  $C^*$ -algebras. The duality creates a dictionary between the two categories. Topological constructions such as cofibrations, mapping cylinder and mapping cone are translated into their  $C^*$ -algebraic counterparts [12, 13]. In the absence of commutativity, the dictionary may still contain noncommutative CW complexes (NCCW complexes) as the  $C^*$ -algebraic version of the topological CW complexes defined in [6]. The noncommutativity comes from the fact that noncommutative CW complexes are algebras of matrix-valued continuous functions. In [11], we studied some of the geometric properties of noncommutative CW complexes. In this paper, we are motivated by noncommutative constructions through NCCW complex examples and study their topological properties. In this regard the paper is organized as follows.

Section 2 is a review of basic tools: extensions, pullbacks, NCCW complexes and their primary properties. Section 3 is devoted to the study of cofibrations and cofibres in the category of NCCW complexes. In this section we explain the  $C^*$ -algebraic counterparts of the topological mapping cylinder and mapping cone.

---

Received July 8, 2014.

2010 *Mathematics Subject Classification.* Primary 46Lxx, 57Txx, 57Q05, 57Q12.

*Key words and phrases.*  $C^*$ -algebra; Cofibration; Cofibre; CW complex; K-group; mapping cone(cylinder); mapping conical(cylindrical) telescope; noncommutative CW complexes.



Appl. Gen. Topol. 18, no. 1 (2017), 1-11

doi:10.4995/agt.2017.2250

© AGT, UPV, 2017

## Algebraic and topological structures on rational tangles

VIDA MILANI<sup>a</sup>, SEYED M. H. MANSOURBEIGI<sup>b</sup> AND HOSSEIN FINIZADEH<sup>c</sup>

<sup>a</sup> Department of Mathematics and Statistics, Utah State University, Logan Utah, USA  
([vida.milani@yahoo.com](mailto:vida.milani@yahoo.com))

<sup>b</sup> Department of Computer Science, Utah State University, Logan Utah, USA  
([smansourbeigi@aggiemail.usu.edu](mailto:smansourbeigi@aggiemail.usu.edu))

<sup>c</sup> Department of Mathematics, Shahid Beheshti University, Tehran, Iran ([hfinizadeh@gmail.com](mailto:hfinizadeh@gmail.com))

Communicated by M. Sanchez

### ABSTRACT

*In this paper we present the construction of a group Hopf algebra on the class of rational tangles. A locally finite partial order on this class is introduced and a topology is generated. An interval coalgebra structure associated with the locally finite partial order is specified. Irrational and real tangles are introduced and their relation with rational tangles are studied. The existence of the maximal real tangle is described in detail.*

2010 MSC: 16T05; 11Y65; 18B35; 57M27; 57T05.

KEYWORDS: group Hopf algebra; locally finite partial order; tangle; pseudo-module; bi-pseudo-module; pseudo-tensor product; incidence algebra; interval coalgebra; continued fraction; tangle convergent.

### 1. INTRODUCTION

Rational tangles are not only beautiful mathematical objects but also have many applications in other fields such as biology and DNA synthesis [5]. The theory of tangles was invented in 1986 by Conway in his work [2]. He introduced the notion of rational tangles and with each rational tangle he associated a rational number by the continued fraction method. The associated rational

Received 30 March 2014 – Accepted 22 November 2016

## Discrete dynamics on noncommutative CW complexes

VIDA MILANI<sup>a</sup> AND SEYED M. H. MANSOURBEIGI<sup>b</sup>

<sup>a</sup> Dept. of Math., Faculty of Math. Sci., Shahid Beheshti University, Tehran 1983963113, Iran.  
School of Mathematics, Georgia Institute of Technology, Atlanta GA 30332, USA (v-milani@sbu.ac.ir;  
vmilani3@math.gatech.edu)

<sup>b</sup> Dept. of Electrical Engineering, Polytechnic University, NY 11747, USA  
Department of Electrical Engineering, SCCC, Brentwood, NY 11717, USA (mansous@sunysuffolk.edu)

### ABSTRACT

---

*The concept of discrete multivalued dynamical systems for noncommutative CW complexes is developed. Stable and unstable manifolds are introduced and their role in geometric and topological configurations of noncommutative CW complexes is studied. Our technique is illustrated by an example on the noncommutative CW complex decomposition of the algebra of continuous functions on two dimensional torus.*

---

2010 MSC: 46L85, 55U10, 54H20, 34D35.

KEYWORDS: closed hemi-continuous,  $C^*$ -algebra, CW complexes, discrete dynamical system, modified Morse function, noncommutative CW complex, open hemi-continuous, stable manifold, unstable manifold.

### 1. INTRODUCTION

The theory of CW complexes was invented by Whitehead in 1949 [14]. The concept of CW complex structures on topological manifolds has been a great development in the category of topological spaces [8]. It is a well known fact that the topology of a manifold can be reconstructed from the commutative  $C^*$ -algebra of continuous functions on it [7, 10]. In other words commutative  $C^*$ -algebras play as the dual concept for topological manifolds. Away from



## Morse theory for $C^*$ -algebras: a geometric interpretation of some noncommutative manifolds

VIDA MILANI\*, ALI ASGHAR REZAEI AND SEYED M. H. MANSOURBEIGI

### ABSTRACT

---

*The approach we present is a modification of the Morse theory for unital  $C^*$ -algebras. We provide tools for the geometric interpretation of noncommutative CW complexes. Some examples are given to illustrate these geometric information. The main object of this work is a classification of unital  $C^*$ -algebras by noncommutative CW complexes and the modified Morse functions on them.*

---

2010 MSC: 06B30, 46L35, 46L85, 55P15, 55U10.

KEYWORDS:  $C^*$ -algebra, critical points, CW complexes, homotopy equivalence, homotopy type, Morse function, Noncommutative CW complex, poset, pseudo-homotopy type,  $*$ -representation, simplicial complex.

### 1. INTRODUCTION

Morse theory is an approach in the study of smooth manifolds by the tools from calculus. The classical Morse theory provides a connection between the topological structure of a manifold  $M$  and the homotopy type of critical points of a function  $f : M \rightarrow \mathbb{R}$  (the Morse functions).

On a smooth manifold  $M$ , a point  $a \in M$  is a *critical point* for a smooth function  $f : M \rightarrow \mathbb{R}$ , if the induced map  $f_* : T_a(M) \rightarrow \mathbb{R}$  is zero. The real number  $f(a)$  is then called a *critical value*. The function  $f$  is a *Morse function* if i) all the critical values are distinct and ii) its critical points are non degenerate, i.e. the Hessian matrix of second derivatives at the critical points has a

---

\*Corresponding author.

## Quantum MIMO n-Systems and Conditions for Stability

Seyed M.H. Mansourbeigi

*Dept.of Electrical Engineering , Polytechnic Univ. Long Island Graduate  
Center, New York, USA*

e-mail: s.mansourbeigi@ieee.org

<sup>1</sup>Vida Milani

*Dept.of Math., Faculty of math.Sciences, Shahid Beheshti Univ., Tehran,  
1983963113, IRAN*

e-mail: v-milani@cc.sbu.ac.ir

### Abstract

In this paper we generalize the notion of a dynamical system to that of a quantum dynamical system and try to find some conditions for the stability of an n-D Quantum (MIMO) system  $P(X)$ . It contains two parts. The first part is to introduce the n-D Quantum MIMO systems where the coefficients vary in the algebra of Q-meromorphic functions. Then we introduce some conditions for the stability of the solutions of these systems. The second part is to show that this Quantum system has the n-D system as its quantum limit and the results for the SISO,SIMO,MISO,MIMO are obtained again as special cases.

**keywords:** Dynamical system; Q-Meromorphic functions; Functional quantization; Quantized space; Derivations

*AMS Math Subject Classification:* 46L55, 46L65, 28D, 37K, 54H

*Computing Classification System codes:* C4, F10

# Quantum Computing v.s. Conventional Computing: Near-Term Solution is Smart Distributed Systems

Seyed M. Mansourbeigi, Stephen W. Clyde, Department of Computer Science, USU

## Introduction

Are we getting to the end of Moore's law???? Moore's law which is held for 50 years tells computer power doubling every 18 months may begin to wane in the next 10 or so years. Imagine you buy a computer having the same power as the last year and the year before last year. Would you buy a new computer? Or upgrade? The present computer model is CPU with I/O and combination of logical gates, and software which controls these I/O's based on finite memory and Turing machines. Basically the above hardware software model surpasses the 6 and 1 bit. That is what we have in computer theory and real physics. Quantum computers are not based on only 0's and 1's, Turing machines and finite automata. Quantum computer computes on individual atoms and instead of 0's and 1's which are called bits, there are Qubits in quantum computer which is sometimes 1 and sometimes 0 and sometimes any number between 1 and 0. No longer computing would be just on 0's and 1's, so how would be the conventional programming possible? It would be very Challenging!!!! The operations of quantum computer is incredibly sensitive to vibrations, temperature, electromagnetic radiation and environmental effects and the quantum mechanics is for smaller elements in the universe such as atoms, electrons, photons, and subatomic particles. Nevertheless Quantum Computing is still in state of infancy, and have not started to make useful things, however experiment is getting better and better. So Quantum Computing will revolutionize the way we do computing and will not be the same architecture as regular silicon architecture. A near-term solution might be **Smart Distributed computer communication systems**.

## Materials and methods

From theoretical physics and engineering:  
Transistor operation, Quantum Mechanics, Quantum tunneling, Quantum superposition, Quantum entanglement, Schrodinger equation

From mathematics:  
Geometric topology, Algebra, Computational geometry, Discrete algebraic topology, Discrete Morse theory

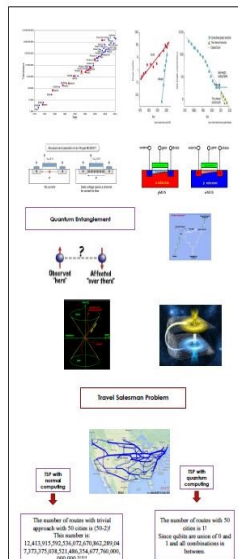


## Results

Every day we see more evidence which Mathematics will get more involve and more dominant in all areas of Engineering, Physics, Computer Science and so many other Sciences, and probably this trend will get more stronger for another quarter of century. And field of Mathematics rapidly expanding applications in areas of Geometry and Topology and Algebra. Recently, techniques and models borrowed from classical algebraic topology have yielded a variety of new applicable models for computer science problems. Techniques and models, borrowed from classical algebraic topology, represent a promising new approach to the theory of distributed computing. I was particularly interested in making the mathematical concepts applicable to the computer science community. M.P. Heuley and N. Shurey [10] were the first to use simplicial complexes to model decision tasks, and to formulate properties of decision tasks in terms of simplicial homology. It is intriguing to speculate whether some kind of topological classification of protocol complexes might yield a useful computational classification of objects. The problem of coordinating concurrent processes remains one of the central problems of distributed computing. Coordination problems arise at all scales in distributed and concurrent systems. Coordination is difficult because in modern multiprocessor systems, processes may be diverse, without waiting for a variety of reasons. Because the processes are asynchronous, a process cannot distinguish a failed process from a slow process; a program that solves a decision task is a protocol. Decision tasks are modeled by model reactive systems such as distributed, file systems, or digital control systems. An input value represents information entering the system from the outside world, such as a character typed at a keyboard, a message from another computer, or a signal from a sensor. An output value models effects on the outside world, such as a decision to commit a transaction, to dispense cash in banking system, or to launch a missile. The mathematical modeling of a simple task would be as follows:

Suppose we are given a set of  $n = 1$  sequential processes. This set of processes could be a set of  $n = 1$  sequential threads of control communicating by applying operations to objects in shared memory. Each process starts out with a private input value (A simplicial complex of initial configurations). The processes communicate for a while, then each process chooses a private output value (A simplicial complex of final configurations), and then halts. Examples of shared objects include message queues, machine variables, read-write variables, or objects of arbitrary abstract type. A set of  $n = 1$  mutually compatible initial or final process states is modeled as a  $d$ -dimensional simplex. A decision task for this set of processes is given by a tuple  $(O, \Delta, \lambda)$  where:

- $\Delta$  is a simplicial complex (A simplicial complex of initial configurations)
- $O$  is an output complex (A simplicial complex of final configurations)
- $\lambda$  is a simplicial map from  $\Delta$  to  $O$ : a map carrying each input  $n$ -simplex of  $\Delta$  to a set of  $n$ -simplices of  $O$ .



## Conclusions

The final question is:  
1- What kind of Hardware we will have for quantum Computing and when?  
2- What kind of programming tools and software and environment we need for this new revolutionized world of Quantum Computing and when?

But Near term and double solution is:  
Smart Distributed Systems

## Literature cited

- [1]- "Morse Theory for  $C^*$ -Algebras": A Geometric interpretation of some NC-Manifolds, Vito Micale, Seyed M. H. Mansourbeigi and Ali Akbari Farnesi, Applied General Topology, vol. 12, no. 2 (2011)
- [2]- "Discrete Dynamics on NCW complexes", Vito Micale, Seyed M. H. Mansourbeigi, Applied General Topology, vol. 14, no. 2 (2013)
- [3]- "Cellularity in the Category of NCW complexes", Vito Micale, Seyed M. H. Mansourbeigi and Ali Akbari Farnesi, Acta Math. Univ. Comenianae, vol. 85, no. 1 (2014)
- [4]- "Quantum KMO Systems and Conditions for Stability", Seyed M. H. Mansourbeigi and V. Micale, Proc. JCM, 2 (2008)
- [5]- "Two steps closer to Quantum Internet", IEEE Spectrum, Jan 2016
- [6]- "MOORE Theory and Design", B. L. Oberg, R.M. Warner, Oxford Univ. Press, 1999
- [7]- "Introduction to Quantum Mechanics", David J. Griffiths, Pearson Education Limited, 2014
- [8]- "Quantum Computing", Doreen Williams, Educational Value
- [9]- "Quantum Computing", UNIV Educational Value
- [10]- "The asynchronous computability theorem for  $n$ -resilient tasks", M.P. Heuley and N. Shurey, Proceedings 5th Annual ACM Symposium on Theory of Computing, 1993.

## Acknowledgments

The authors would like to thank the Department of Computer Science, USU



## VITAE

**SEYED M-H MANSOURBEIGI****Professional Profile**

- Electrical and system engineer, with 10+ years of experience at Utah State University and State University of New York, AT&T Company and Brookhaven National Laboratory (BNL) in Analog/Digital signal processing communication systems in electrical engineering systems and computer communications projects.
- Statistical analyst with advanced multi-task skills in designing and managing multiple electrical system projects for commercial, industrial, market, and scientific use.
- Creative, goal-driven problem solver who thrives on identifying and solving problems.
- Adjunct professor at Suffolk County Community College with demonstrated verbal and written communication skills to express complex technical terms in clear language to facilitate students' learning.
- Established reputation for maintaining high standards of personal and professional conduct. Demonstrated experience and strong work ethic in budgeting, schedules, and giving technical direction in managing dynamical systems in mechanical and electrical designs.

**Area of Interest**

- Statistical analog/digital communication software and hardware engineering, theoretical computer sciences, embedded systems.
- Power electronic systems, control (linear/nonlinear) systems engineering, quantum mechanics, signal processing and communications, electromagnetics and RF/wireless systems, chaos fractals dynamical system, dynamical of physical system.
- Energy resources & technology, signal processing, stochastic processing. Kalman and Wiener Filtering, neural network and machine learning, deep learning Tensorflow, phase lock loop, Markov process, and hidden Markov model.

## Skills

- Analog/digital communication electrical engineer
- Project management
- Innovative problem solver
- Regulatory compliance
- AIX UNIX; Microsoft office suit; SUN Solaris, LINUX, UBUNTU.
- SAS; C/C++; Windows; DB2, JAVA, C#, Python, Tensorflow.
- MATHLAB; ORACLE; LTSPICE, PostgreSQL, UML, FPGA, VLSI.
- System engineering, software engineering, embedded systems.

## Professional Experience

Utah State University College of Engineering Computer Science Department  
08/2015 - Present

*(Teaching assistant, research assistant, statistical consultant)*

- Teaching Assistant: CS3100, CS5700, CS5200:  
Participate in lectures in lieu of business trips of main lectures, grade students' projects and exams, maintain office hours for trouble shooting, and guide students.
- Research Assistant PostgreSQL and database programming for new born projects for State of Utah Health Division.
- Statistical consulting for two Ph.D. candidates, ANOVA, T-Test, Chi-Square Test.

SCCC - Suffolk County Community College

21/Sep/2009 - Present

*(Adjunct Professor)*

- Teach Industrial Control including: Electronics and Electrical Systems, Motor Control, PLC (Programming Logic Controller), Electric Drive Systems and Drivetrain Components in Electric Vehicles; Motor Drives and Vehicle Power Electronics; Safety Procedures for Working with High Voltages and Power Levels Typical of Electric Vehicles; Hydraulic; Pneumatic; Business Management.
- Involved in analog and digital communications engineering systems. Received Teacher Performance Award.
- Acquired appreciation letter from the dean of the faculty for facilitating education by applying new strategies and technologies to the course materials.

CTG (Computer Task Group), IBM

23/Jan/2006 - 01/Sep/2009

*System Engineering*

- Major contributor for high availability multi-processing for AIX UNIX and analog/digital communications.
- Planned the development of potential engineering projects and products by managing and setting up context manager software and hardware reposting the customer correspondences.
- Chairman of Signal Processing in IEEE (Institute Electrical Electronics Engineer).
- Involved with the complexity of the systems by thinking logically, and analyzing top down and bottom up approach, resulted in 15% time and 10% budget savings.

State University of New York (SUNY)

12/Jul/1999 - 20/Jan/2006

*Electrical Engineering*

- In migration IBM Z Operating Systems to AIX UNIX Operating Systems SUNY saved 14% in budget by thoroughly reviewing every detail in a project.
- Researched new methodologies and developed new procedures in green energy resources and technology (solar, wind, tidal, hydro, geothermal, coal) to apply principles of electrical theories to the projects.
- Involved in the development of potential engineering projects by employing dynamical systems, fractals and chaos to circuits and mechanical systems.
- Implemented battery technologies and battery charging techniques, to take care of computer systems uptime 24/7.

AT&T (*Electrical Engineering*)

05/Sep/1994 - 12/Jul/1999

- Managed and installed electrical equipment, components in system 5 TCP/IP network protocol for commercial, industrial, and scientific use.
- Job time constraint in critical business systems 24/7 by fixing a problem in a time-frame by quickly developing a solution, resulting in 12% time and 8% budget savings.
- Planned troubleshooting process, safety measures and testing for analog and digital communication in Code, Frequency and Time Division Multiple Access to ensure compliance.
- Participated in company training courses kept track of new developments in technology and worked with wireless energy transfer systems.

SUNY at Stony Brook/ Brookhaven National Lab

01/Jun/1987 - 01/Sep/1994

*System Engineering*

- Managed hardware, software, database and computer communication networks as system engineer in the breast cancer research in department of Preventive

Medicine and Brookhaven National Lab.

- Developed innovative and effective ways to improve computer systems saving 6% in time and 4% in budget for the Brookhaven National Laboratory.
- Taught discrete mathematics and probability theory for department of applied math and statistics.
- Participated in the development of ad hoc engineering projects in the department of preventive medicine with C/C++ Programming, SAS and SPSS languages.

## HONORS

- Member of the honor society "ETA KAPPA NU"
- Member of the honor "TAU BETTA PI".
- Member of "EPSILON PI EPSILON".

## MEMBERSHIPS and EDITORIAL BOARD

- Member of the editorial board of Universal Journal of Electrical and Electronic Engineering (HRPUB).
- Member of the Institute of Electrical and Electronics Engineering (IEEE).
- Reviewer of the American Mathematical Society (AMS).
- Member of the Association for Computer Machinery (ACM).

## EDUCATION

- Ph.D. College of Engineering Computer Science, Utah State University, 2018 (GPA 4.00/4.00).
- Master of Science Degree (M.Sc.), NYU Tendon School of Engineering, New York *Electrical Engineering*, 1998 (GPA 4.00/4.00).
- Master of Science (M.Sc.), NYU Tendon School of Engineering, New York *Computer Science*, 1993 (GPA 4.00/4.00).
- Master of Science (M.Sc.), State university of New York *Applied Mathematics and Statistics*, 1991 (GPA 3.64/4.00).
- Bachelor of Science (B.Sc.), State university of New York *Computer Science/Engineering*, 1987 (GPA 3.64/4.00).
- Bachelor of Science (B.Sc.), Tehran Polytechnic, Tehran, Iran *Civil/Electrical Engineering*, 1979 (GPA 3.86/4.00).

## CERTIFICATIONS

- PLC (Programming Logic Control 2009)
- Motor Control (2008)
- AIX Unix System Support and Administration (2001)
- Sun Solaris System Administration (2001)

## CONTINUING EDUCATION

Taking Online Courses in MIT, Yale, Stanford, Khan Academy and IIT India.

## JOURNAL PUBLICATIONS

- "Incidence of Stage III+ Retinopathy of Prematurity (ROP) in Tertiary Care Hospital": R.J. Smith, L.Hyman, B. Golub, R. Sosulski, S. Mansourbeigi, A. O. Varma, JAMA (July 8, 1991).
- "Quantum MIMO Systems and Conditions for Stability", Seyed M. H. Mansourbeigi and V. Milani. *Proc. ICM*, **2**, 713-719 (2008).
- "Morse Theory for  $C^*$ -Algebras: A Geometric interpretation of some Noncommutative Manifolds": Vida Milani, Seyed M. H. Mansourbeigi and Ali Asghar Rezaei, *Applied General Topology*, vol 12, No. 2, (2011).
- "Discrete Dynamics on noncommutative CW complexes", Vida Milani, Seyed M. H. Mansourbeigi, *Applied General Topology*, vol 14, No. 2 (2013).
- "Cofibrations in the Category of Noncommutative CW Complexes", Vida Milani, Seyed M.H. Mansourbeigi and Ali-Asghar Rezaei, *Acta Mathematica Universitatis Comenianae*, vol. 85 No.1 (2016).
- "Geometry of Physical Systems", Vida Milani, Seyed M.H. Mansourbeigi and Stephen W. Clyde, *International Journal of Geometric Methods in Modern Physics*, vol.14, no.3, (2017).
- "Algebraic and Topological Structures on Rational Tangles", Vida Milani, Seyed M.H. Mansourbeigi and Hossein Finizadeh, *Applied General Topology*, vol.18, no.1, (2017).
- "Sheaf Theory Approach to Distributed Applications: Analyzing Heterogeneous Data in Air Traffic Monitoring", Seyed M-H Mansourbeigi, *International Journal of Data Science & Analysis*, 3(5), (2017).

- "Sheaf Theory as a Mathematical Foundation for Distributed Applications Involving Heterogeneous DataSets", Seyed M-H Mansourbeigi, IEEE AINA 32<sup>nd</sup> international Conference on Advanced Information Networking and Applications Workshops, pp: 28-33, DOI 10.1109/WAINA.2018.00059, (2018).
- "Quantum Computing vs. Conventional Computing", Poster, Seyed M. Mansourbeigi, Stephen W. Clyde, Department of Computer Science, USU, 2016.

### **CONFERENCES AND PRESENTATION TALKS AND POSTER DAY**

- Mansourbeigi Seyed M.H. and Milani Vida; "Dirac structures on (Quantum) manifolds: The topological conditions for their integrability and the moduli space of integrable Dirac structures," International Conference "DIFFERENTIAL EQUATIONS and TOPOLOGY: Dedicated to the Centennial Anniversary of Lev Semenovich Pontryagin (1908-1988)," State University of Moscow, Moscow, Russia, June 17-22, 2008.
- Mansourbeigi Seyed M.H. and Milani Vida; "Morse Theory: A tool for geometric classification of the noncommutative CW Complexes," AMS Meeting 1059: University of New Mexico, Albuquerque, New Mexico, April 17-18- 2010 USA.
- Mansourbeigi Seyed M.H. ; "Sheaf Theory as a Mathematical Foundation for Distributed Applications Involving Heterogeneous Datasets," The 32nd IEEE International Conference on Advanced Information Networking and Applications (IEEE AINA-2018) May 16-18, 2018 Pedagogical University of Cracow, Poland.
- Mansourbeigi Seyed M.H. : Poster day April 2016, Quantum Computing vs. Conventional Computing: Near-Term Solution in Smart Distributed Systems, Utah State University, College of Engineering Computer Science Department.