Fall 2021

Topological biclustering ARTMAP

Raghu Yelugam

Department:

TOPOLOGICAL BICLUSTERING ARTMAP

by

RAGHU YELUGAM

A THESIS

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

in

COMPUTER ENGINEERING

2021

Approved by:

Dr. Donald C Wunsch II, Advisor
Dr. Cihan Dagli
Dr. Hamidreza Modares

PUBLICATION THESIS OPTION

This thesis consists of the following two papers, formatted in the style used by the Missouri University of Science and Technology.

Paper I: Pages 18-39, TopoBARTMAP: Biclustering ARTMAP with or without Topological Methods in a Blood Cancer Case Study, have been accepted by the IEEE International Joint Conference on Neural Networks 2020.

Paper II: Pages 40-78, TopoBARTMAP: Topological Biclustering ARTMAP, are submitted to IEEE Transactions on Neural Networks and Learning Systems Journal.

ABSTRACT

Detection of gene mutations is central for assessing genetic factors affecting disease predisposition, genetic causes of a particular disease, and gene-targeted treatment. DNA microarray methods are widely used to detect mutations by contrasting the expression levels of thousands of genes together under varying experimental conditions. The experimental conditions could be diseased cell states compared with the normal cell states. Biclustering, a robust exploratory data analysis tool, can be applied to microarray data to detect subsets of genes that co-express highly only for a subset of experimental conditions. Such detection is crucial for gaining insights into gene regulatory networks, differential gene expression, and gene-disease associations to identify candidate genes for further study. However, biclustering fails to identify functional associations between genes within a bicluster and group functionally related genes that might not co-express significantly.

This work presents a novel biclustering algorithm, TopoBARTMAP, which combines a biclustering ARTMAP (BARTMAP) with a topological ART (TopoART) to improve the quality of biclustering. Whilst producing a graphical representation of space, topological clustering can identify arbitrarily shaped clusters in space that are difficult to detect otherwise. These methods find application in analyzing disease-specific gene subgroups and disease progression. TopoBARTMAP inherits, from TopoART, the ability to detect arbitrarily shaped biclusters whilst remaining robust to noise. These capabilities of TopoBARTMAP are rigorously demonstrated in the study with 35 benchmark cancer datasets. Further, the benchmarking study underpins the statistically significant performance improvement observed in comparison to other compared methods. Using the breast cancer dataset containing expression levels of 39,326 genes observed over 38 samples, the graphical output of TopoBARTMAP is analyzed to detect intra-bicluster-gene associations within the dataset.

ACKNOWLEDGMENTS

I dedicate this thesis to my late grandparents.

I want to acknowledge the continuous academic, moral, and financial support provided by my advisor, *Prof Donald C Wunsch II*. Without his constant encouragement, meticulous guidance to fine-tune abstract ideas, and him agreeing to supervise my thesis after the third semester here at Missouri S&T, this work would have not been possible.

I profoundly thank *Dr Hamidreza Modares* for the encouraging me to assess the mathematical validity of conceived ideas, which etched an indelible impression on the mathematical rigour need for the practice undertaken from the beginning of my endeavours. I deeply thank *Prof Cihan Dagli* for this extensive instructions and sharing his invaluable time and experience. I am very thankful to both for agreeing to serve as my thesis committee members.

No words would suffice to express my gratitude to my fellow researcher, and mentor in ways, *Dr Leonardo E B Da Silva* for his instrumental support by giving me good advice that led to many substantial improvements in the presented work. I wish to especially thank *Austin Vandegriffe* for explicating fundamentals of topology and helping me with LaTex manuscript.

I want to specially thank *Dr Kurt Kosbar* for providing me an opportunity to work as a teaching assistant and ECE Staff and Graduate Support Office Staff for their continuous support.

I wish to thank for immense support I have received from many that I had the opportunity to meet and share experiences and become friends with at Missouri S&T. Finally, I want to thank my family for their unwavering support and constant encouragement in my pursuits, be it academic or professional.

TABLE OF CONTENTS

Page

PUBLICATION THESIS OPTION iii

ABSTRACT iv

ACKNOWLEDGMENTS v

LIST OF ILLUSTRATIONS ix

LIST OF TABLES x

SECTION

1. INTRODUCTION 1

    1.1. GENE EXPRESSION 4

    1.2. ADAPTIVE RESONANCE THEORY 6

    1.3. TOPOLOGICAL DATA ANALYSIS 8

        1.3.1. Metric Space, Open Sets, and Connectedness 8

        1.3.2. Simplicial Complex and Nerve of a Cover 13

PAPER

I. TOPOBARTMAP: BICLUSTERING ARTMAP WITH OR WITHOUT TOPOLOGICAL METHODS IN A BLOOD CANCER CASE STUDY 18

    ABSTRACT 18

1. INTRODUCTION 19

2. BACKGROUND 20

2.1. FUZZY ART 21

2.2. TOPOART 23

2.3. BARTMAP 25

3. TOPOBARTMAP 27

4. EXPERIMENTATION 28

5. RESULTS AND DISCUSSION 32

6. CONCLUSION 33

ACKNOWLEDGMENT 36

REFERENCES 36

II. TOPOBARTMAP: TOPOLOGICAL BICLUSTERING ARTMAP 40

ABSTRACT 40

1. INTRODUCTION 41

2. RELATED WORK 42

2.1. TOPOLOGICAL ART 43

2.2. BICLUSTERING ARTMAP 46

3. TOPOLOGICAL BICLUSTERING ARTMAP 49

4. EXPERIMENTATION 53

4.1. DATA 53

4.2. CLUSTERING PERFORMANCE 54

4.3. METHODS 56

4.4. PARAMETER TUNING 57

5. RESULTS AND DISCUSSION 59

5.1. ORIGINAL DATA ORDERING EXPERIMENT 59

5.2. RANDOMISED DATA ORDERING EXPERIMENT 62

5.3. NCBI GEO GSE89116 GENE EXPRESSION STUDY-CASE 63

6. CONCLUSION 68

ACKNOWLEDGMENT 69

REFERENCES 70

SECTION

2. CONCLUSIONS 79

REFERENCES 80

VITA 91

LIST OF ILLUSTRATIONS

Figure Page

SECTION

1.1. The DNA microarray data and biclustering result. 3

1.2. Gene Expression Process diagram. 5

1.3. ART1 model network structure. 7

1.4. Examples of simplices. 14

1.5. Change in identified simplicial complexes with change in open-ball radius. 15

1.6. The Nerve of the open cover for the data-set shown in Sub-figure (a). 17

PAPER I

1. Block diagram of TopoBARTMAP. 27

PAPER II

1. Schematic of TopoBARTMAP. 50

2. Statistical difference plot for the performance of each method on original data ordering experiments. 61

3. Statistical difference plot for the performance of each compared method on the randomized data-sets. 63

4. The sample classes identified by TopoARTa module of TopoBARTMAP for GSE89116 data-set. 66

5. The *Topological Gene Bi-cluster Association Network* at correlation level of 0.5848. 67

LIST OF TABLES

Table Page

PAPER I

1. Data sets used for experimental analysis. 30

2. Biclustering results on original data sets with performance measured using Adjusted Rand Index. 34

3. Biclustering results on randomized order of observations data sets run with parameters fixed to optimal parameters found for original data set. 34

4. Biclustering results on randomized order of features data sets run with parameters fixed to optimal parameters found for original data set. 35

5. Biclustering results on randomized order of observation data sets run with GA optimization for TopoBARTMAP, BARTMAP, and TopoART and variation of ρ for Fuzzy ART. 35

6. Biclustering results on randomized order of feature data sets run with GAs. 35

PAPER II

1. Thirty-five benchmark cancer data sets used for experimentation. 55

2. Biclustering results on thirty-five benchmark cancer data-sets. 60

3. Statistical comparison of different algorithms made using post-hoc test for observed biclustering results reported in Table 2. 61

4. Observed biclustering results on randomized data-sets. 64

5. Post-hoc test results for observed performances reported in Table 4. 65

6. Identified highly active differentially expressed genes with corresponding observation prototypes and tumor categories. 68

SECTION

1. INTRODUCTION

With today's advancements in molecular biology and chemical analysis, we can comprehend the activity of several thousand genes during cell division. At a molecular level, genes are subsequences of Deoxyribonucleic acid (DNA) that encode the information needed to synthesize functional proteins which control the development of an organism [1]. The mechanism by which the cells synthesize functional products using genes is called as *gene expression*. Understanding gene expression illuminates the conditions required to produce observable traits, called *phenotype*, in the offspring after cell division. By regulating gene activity, i.e., by controlling the timing, location, and amount of synthesized protein, it is possible to alter the manifestation of a phenotype in an organism. Although gene regulation is a cellular process, several extrinsic factors, such as a change in temperature, nutrient availability, and PH-level, can affect gene expression and alter the development of an organism. Further, any change to the DNA sequence of the gene can affect the expression product.

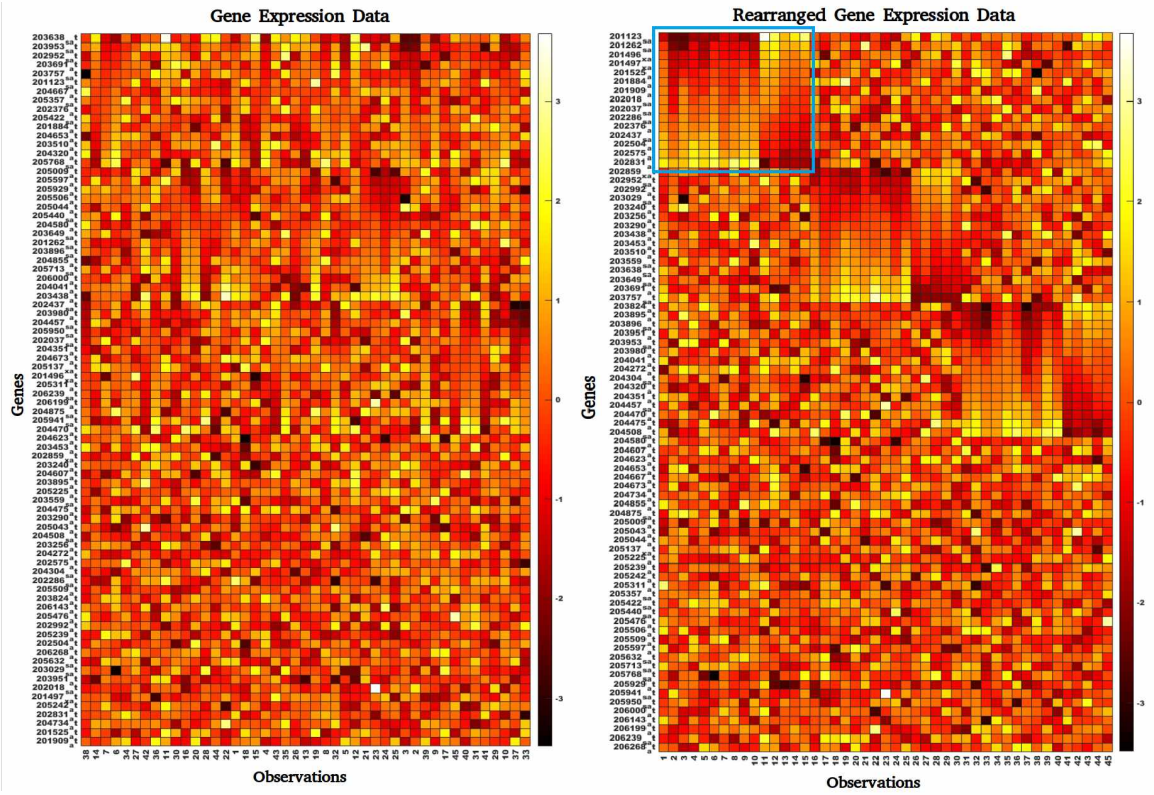
A permanent change to the DNA sequence that makes a gene is called a mutation. Since the functioning of a cell is intrinsically related to the functioning of proteins, any mutation to the DNA sequence of a gene can alter the coded protein, resulting in its malfunction. Mutations to the DNA can be hereditary or acquired, where the latter is acquired sometime during an organism's lifetime. Hereditary mutations occur in the germ cell and are inherited by all the cells of a new organism if they are not detected and repaired by the cell. Acquired mutations can be caused by environmental factors such as chemical exposures (such as cigarette smoke), pathogenic factors (such as viral infections), and drug usage [2, 3, 4, 5]. Although not all mutations are harmful, certain mutations are malignant

and can affect the predisposition of an organism. While Mendelian diseases are associated with hereditary mutations, diseases like cancer are related to acquired mutations; hence detecting mutations is pivotal for treating genetic conditions.

There are several methods for identifying mutations to the DNA sequence of the gene. These methods utilize the basic chemical properties of DNA, or the enzymes that act upon it during cell division [6]. Single base-pair mutations are detected by comparing each base pair in the gene sequence to that of the normal gene. More significant mutations are detected by studying chromosomes during cell division (also known as cytogenetics), and molecular diagnosis [7]. Cytogenetic methods like Fluorescence in situ hybridization (FISH) use fluorescent probes that bind to specific regions (DNA sequence) of chromosomes, and fluorescent microscopy is used to find the binding locations. Molecular methods [8, 9] use nucleic acid-based analytical methods to assess the genetic makeup. These methods are further classified as methods for detecting novel mutations and for detecting known mutations. To identify the former, methods based on the mobility of molecules under a uniform electric field (electrophoresis) are often used. To check for the latter, known mutations, methods that sequence DNA, exploit complementary binding properties of DNA strand, or produce a copy of DNA, like Polymerase chain reaction [10], are used.

DNA microarray or DNA chips [11, 12] use complementary bind properties of DNA strands to detect known mutations. These methods allow for measuring the expression levels of thousands of genes in one experiment, thus identify multiple known mutations in DNA. Furthermore, microarrays are often used to study gene expression profiles under different experimental conditions. The experimental conditions investigated could be but are not limited to stages of disease development, the time course of a cellular process, disease cases compared with controls, and various phenotypes [13]. Such studies with microarrays generate an immense amount of data as matrices. Frequently clustering methods are used to analyze, compare, and contrast different patterns in the microarray data [14, 15, 16]. Figure 1.1a shows an example of DNA microarray gene expression data.

Clustering methods are a family of analytical methods used for grouping objects, where each group is referred to as a cluster. The grouping is such that elements within a cluster are homogenous and differ significantly from the elements of other clusters [17]. Clustering of microarray is used to identify genes that have similar expression patterns across various experimental conditions. Such grouping is key for identifying genes associated with the regulatory process studied in the experiment (cite from TopoBARTMAP-II paper). Furthermore, two-way clustering or *Biclustering* is used to identify subsets of genes expressed only during a subset of experimental conditions. Figure 1.1b shows the result of biclustering the example gene expression data.



(a) (b)

Figure 1.1. The DNA microarray data and biclustering result. The Sub-figure (a) presents a DNA microarray gene expression matrix with rows representing individual genes and columns representing experimental conditions or observations. Each entry of a row indicates the expression level of the corresponding gene for the corresponding column experimental condition. The Sub-figure (b) presents the output of biclustering the expression data. The rows and columns are rearranged to show the biclusters in the data.

Biclustering methods allow for comparing expression profiles of genes while simultaneously comparing experimental conditions such as diseases. Such clustering identifies differential gene expression under varying conditions and can be applied to identify genes with altered expression, hence mutations. A subset of studied genes and the experimental conditions they are highly expressed define a bicluster. Each bicluster defines a differential gene module whose expression is either inhibited or promoted due to the changes in the associated genes.

This thesis presents a biclustering method based on Adaptive Resonance Theory, a neural learning theory. The presented method, Topological Biclustering ARTMAP, combines biclustering with Topological Data Analysis to identify arbitrarily shaped biclusters and within bicluster gene relations, ultimately to improve the quality of identified biclusters. The rest of the introduction is organized as follows. Section 1.1 briefs on gene expression and microarray data. In Section 1.2 necessary fundamentals of Adaptive Resonance Theory are presented, while Section 1.3 presents Topological data analysis.

1.1. GENE EXPRESSION

It is well known that cells archive the instructions needed for building proteins in DNA, the hereditary unit. Portions of this double helix molecule are transcribed to RNA, and this RNA is translated to proteins - which carry out specific cellular functions. Depending on the needs, cells can adjust the protein production by up-regulating or down-regulating the expression of a gene. Despite the identical genetic information shared by all the cells that make an organism, the cells that differ in function use different catalysts to express parts of DNA specific to their role. Gene expression, or the process by which proteins are synthesized from the genes, follows two steps.

DNA as such is made of four nitrogenous compounds, namely Adenine (A), Thymine (T), Guanine (G), and Cytosine (C), a sugar called deoxyribose, and phosphate base. A single nitrogenous compound with deoxyribose and phosphate base form nucleotides. DNA

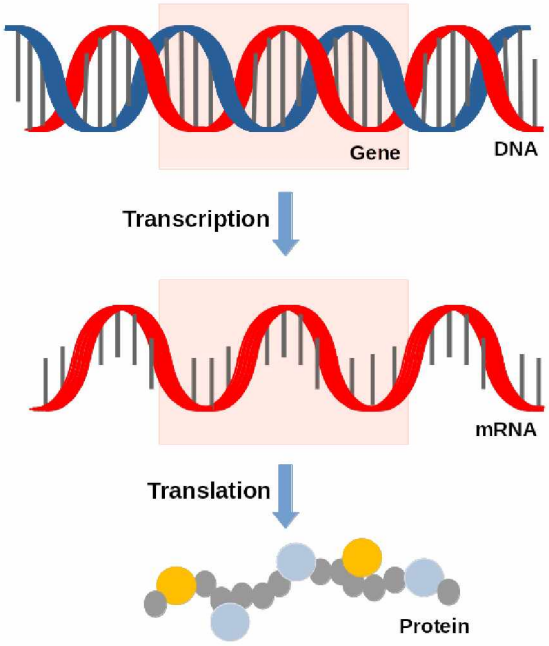


Figure 1.2. Gene Expression Process diagram.

strands are made of nucleotides held together by the covalent bond between the sugar of one nucleotide with the phosphate of the other. In a DNA helix, while nucleotides with Adenine base always pair with Thymine base, nucleotides with Guanine always pair with Cytosine. This specific pairing of nucleotides of one strand with another strand of DNA helix is called complementary base-pairing. The complementary pairs are held together by the hydrogen bonds. The first step of gene expression is building a Ribonucleic acid (RNA) molecule complementary to a portion of a strand of DNA molecule.

The production of RNA is directed by an enzyme called RNA polymerase. At first, the RNA polymerase detects a region very next to the gene, the promoter region. The enzyme binds to the DNA at the promoter region and breaks the hydrogen bonds between the nucleotides separating DNA strands. Next, polymerase builds an RNA with ribonucleic acid nucleotides (ribonucleotides), where the sugar is ribose instead of deoxyribose. Further, the Thymine base is replaced with Uracil (U). An RNA chain is initiated by bonding A or G with the nucleotides of the DNA strand used for RNA construction. Once the first nucleotide

is connected, the successive nucleotides are joined to elongate the chain. As the RNA is elongated with the addition of consecutive nucleotides, the polymerase moves along the DNA, adding incoming ribonucleotides. The two DNA strands bond at the disjoint region and intertwine as soon as the polymerase passes. Once polymerase reaches the end region of a gene called terminators, the RNA chain is terminated. This whole process is called *transcription*. Transcription products include Messanger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA).

In the later step, the mRNA is translated to make proteins; hence, it is called *translation*. The Figure 1.2 summarizes the transcription and translation steps of gene expression. The translation step happens in the cell's ribosome, where the rRNA directs the catalysis of protein synthesis. While mRNA dictates the order in which amino-acids should be connected, the tRNA links amino-acids as per the order. Each of the twenty essential amino-acids (footnote) is coded by three consecutive nucleotides of mRNA. Such a sequence of three nucleotides is called a codon. Each tRNA molecule has two binding sites. One of these sites has a complementary pairing sequence for a specific codon; this sequence forms an anticodon. The other site binds to amino-acid coded by the anticodon and codon. During Translation, the tRNA bond with the complementary codon while bringing an amino-acid to the ribosome. The amino-acids are joined as the ribosome moves along the mRNA, forming a polymer - protein. Once the ribosome reaches a stop codon, the Translation terminates.

1.2. ADAPTIVE RESONANCE THEORY

Adaptive Resonance Theory(ART) is a neural theory proposed to explain how the brain processes information. The intuition behind the theory is that stable code development happens due to the resonance between the bottom-up sensory stimulus and top-down expectation signals. Essentially the ART learning model consists of two interacting fields, field F1 that receives sensory input and field F2 that produces a top-down expectation signal based on the short-term memory(STM) activity of neurons at the field F1. At field

F1 the top-down expectation is compared with the bottom up sensory information. If the input signals match with the top-down expectation, then resonance ensues. During resonance, the long-term memory (LTM) traces between the two interacting fields are modified. Such a modification aids in the quick recognition of the learnt patterns. However, in case of a mismatch, the corresponding highly active top-down signal generating neurons are inhibited. This inhibition does not affect the other neurons and they continue to produce the top-down expectation signals. The process of inhibition and comparison continues till either a resonance occurs or all the neurons in field F2 are shut down and new neurons learn the bottom-up sensory signals. Thus this theory provides a solution to stability plasticity problem.

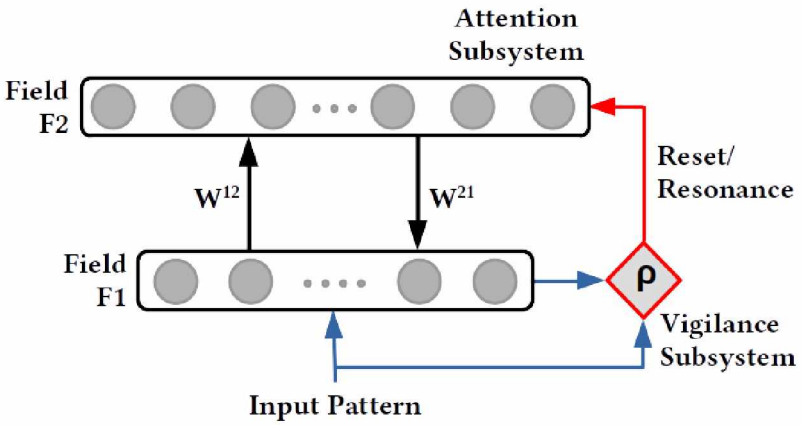


Figure 1.3. ART1 model network structure. In the figure, the long-term memory traces are indicated by the weights W^{12} and W^{21}, vigilance value is represented by ρ.

Several winner takes all neural network models based on ART were developed to perform supervised and unsupervised learning. In the ART1 and ART2 models, the top-down expectations are stored as prototypes that are compared with the actual features of the bottom-up sensory signals. The comparison gives rise to a measure of associability with the category represented by each prototype. As long as this similarity is not lower than a predefined threshold, vigilance value, the sensed signal is considered as a member of the expected category. Further, the filters connecting from field F1 to active category and the filters connecting from field F2 to F1 that generate expectation signal are tuned. In

case of failure to meet the vigilance criterion, the current winning neuron is shut off and is maintained in this state through the current input presentation. A new winner is selected form uninhibited neurons followed by a measure of similarity between the patterns. The process is repeated until vigilance criterion is satisfied or a new prototype is defined with the input signal that can not be associated to any of the learnt prototypes. Figure 1.3 shows the neural network structure for the ART1 learning model.

1.3. TOPOLOGICAL DATA ANALYSIS

Topological Data Analysis, TDA in short, is an emerging applied field in Algebraic Topology and Computational Geometry. Primarily motivated by the idea of utilising topological and geometric methods, TDA extracts insightful information on the shape and structure of data originating from a metric space. Often provided data to these methods arises as point clouds in high dimensional Euclidean space on which techniques like clustering analysis would fail to provide meaningful information (refer carlsson).

For a given finite point cloud data set, TDA methods usually follow the following pipeline [18]. TDA procedures start the analysis by assuming a metric defined on the space from which the data originates, thus a metric space is defined. A continuous structure is built on the data using this metric. Often such a continuous structure built is a simplicial complex or a nerve of the open cover, which will be defined eventually. Such construction involves the identification of connected spaces. Upon construction of the structure, one may either reconstruct the shape using triangulation methods, identify the associated topology, visualise the shape of the data, or study the geometric shapes using persistent homology.

In what follows next, the fundamentals of TDA are discussed, whilst providing relevant proofs for using TopoART for TDA.

1.3.1. Metric Space, Open Sets, and Connectedness. It is seldom the case that finite point clouds data reveals the inherent geometric and topological properties of space it originates. Of interest is the continuity of the space and variation in the *connectedness* of

the space. The quantification of the connectedness of the space is provided by the distance metric used. Therefore, it is essential to define the metric and a metric space, which are defined as following [19].

Definition 1.3.1 (Metric Space) *A* Metric Space $(S, \mathbf{d})$ *is a set* S *together with a function* $\mathbf{d}$ *which associates with each pair* $p, q \in S$ *a real number* $\mathbf{d}(p, q)$ *such that*

(a) $\mathbf{d}(p, q) \geq 0$ *for all* $p, q \in S$

(b) $\mathbf{d}(p, q) = 0$ *if and only if* $p = q$

(c) $\mathbf{d}(p, q) = \mathbf{d}(q, p)$ *for all* $p, q \in S$

(d) $\mathbf{d}(p, q) \leq \mathbf{d}(p, r) + \mathbf{d}(q, r)$ *for all* $p, q, r \in S$ *(due to triangle inequality).*

Thus a metric space can be seen an order pair $(S, \mathbf{d})$, where $\mathbf{d}$ is the metric, or distance measure defined as a function $\mathbf{d}\colon S \times S \longrightarrow \mathbb{R}$. For example $(\mathbb{R}^2, \|\bullet\|_2)$ is a metric space with Euclidean distance as measure. Since most data originates from $\mathbb{R}^d$, hence forth the set S is assumed to be a subset of $\mathbb{R}^d$ unless explicitly stated.

Lemma 1 *Let the metric* $\mathbf{d}$ *for all* $\mathbf{x}, \mathbf{y} \in [0, 1]^{2d}$ *be defined as the following,*

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) = 1 - \frac{|\ \mathbf{x} \wedge \mathbf{y}\ |}{d}$$

where $x_i \wedge y_i = \min(x_i, y_i)$ *and* $|\ \mathbf{x}\ | = \sum_{i=1}^{2d} (\mathbf{x}_i)$. *If the space* $\mathcal{S}$ *is defined as,*

$$\mathcal{S} = \left\{ [z_1, z_2, \cdots, z_d, z_1^c, z_2^c, \cdots, z_d^c]^T \ :\ \forall \mathbf{z} \in [0, 1]^d \right\}$$

where $z_i^c = 1 - z_i$, *then the space* $\mathcal{S}$ *equipped with the distance measure* $\mathbf{d}$ *forms a metric space.*

Proof $\forall \mathbf{x}, \mathbf{y} \in \mathcal{S}$, *we have:*

$$
\begin{aligned}
\mathbf{d}(\mathbf{x},\mathbf{y}) &= 1 - \frac{\mid \mathbf{x} \wedge \mathbf{y} \mid}{d} \\
&= 1 - \frac{\sum_{i=1}^{2d} \mid x_i \wedge y_i \mid}{d} \\
&= 1 - \frac{\sum_{i=1}^{2d} \min(x_i, y_i)}{d} \\
&= 1 - \sum_{i=1}^{d} \frac{\min(x_i, y_i) + \min(x_{d+i}, y_{d+i})}{d} && \because x_i \geq y_i \implies 1 - y_i \geq 1 - x_i \\
& && (\min(x_{i+d}, y_{i+d}) = 1 - \max(x_{i+d}, y_{i+d})) \\
&= 1 - \sum_{i=1}^{d} \frac{\min(x_i, y_i) + 1 - \max(x_i, y_i)}{d} \\
&= 1 - 1 + \sum_{i=1}^{d} \frac{\max(x_i, y_i) - \min(x_i, y_i)}{d} \\
&= \sum_{i=1}^{d} \frac{\mid x_i - y_i \mid}{d} \geq 0 && \because \max(x_i, y_i) - \min(x_i, y_i) = \mid x_i - y_i \mid \geq 0
\end{aligned}
$$

$\therefore \quad \mathbf{d}(\mathbf{x},\mathbf{y}) \geq 0 \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{S}$

Let $\mathbf{x}, \mathbf{y} \in \mathcal{S} \quad \ni \mathbf{x} = \mathbf{y}$. *Then,*

$$
\begin{aligned}
\mathbf{d}(\mathbf{x},\mathbf{x}) &= 1 - \frac{\mid \mathbf{x} \wedge \mathbf{x} \mid}{d} \\
&= 1 - \frac{\sum_{i=1}^{2d} x_i}{d} \\
&= 1 - \frac{\sum_{i=1}^{d} (x_i + 1 - x_i)}{d} && \because x_{i+d} = 1 - x_i \\
&= 1 - \frac{\sum_{i}^{d} 1}{d} \\
&= 0
\end{aligned}
$$

$\therefore \quad \mathbf{d}(\mathbf{x},\mathbf{y}) = 0 \quad \textit{iff} \quad \mathbf{x} = \mathbf{y}.$

$\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{S}$ *we have,*

$$\begin{aligned}
\mathbf{d}(\mathbf{x},\mathbf{z}) + \mathbf{d}(\mathbf{y},\mathbf{z}) &= 1 - \frac{|\ \mathbf{x} \wedge \mathbf{z}\ |}{d} + 1 - \frac{|\ \mathbf{y} \wedge \mathbf{z}\ |}{d} \\
&= \sum_{i=1}^{d} \frac{|\ x_i - z_i\ |}{d} + \sum_{i=1}^{d} \frac{|\ z_i - y_i\ |}{d} \quad \textit{from proof for} \quad \mathbf{d}(\mathbf{x},\mathbf{y}) \geq 0 \\
&= \sum_{i=1}^{d} \frac{|\ x_i - z_i + |\ z_i - y_i\ |}{d} \\
&\geq \sum_{i=1}^{d} \frac{|\ x_i - y_i\ |}{d} \qquad \because |\ x_i - z_i\ | + |\ z_i - y_i\ | \geq |\ x_i - y_i\ |
\end{aligned}$$

$\mathbf{d}(\mathbf{x},\mathbf{y}) = \mathbf{d}(\mathbf{y},\mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{S}$ *is a direct consequence of* $\min(x, y) = \min(y, x)$.

Therefore $\mathbf{d}$ *is metric in space* $\mathcal{S}$.

With notion of metric come the notions of open and closed balls and neighbourhood of a point. Open and closed balls can be defined as [19].

Definition 1.3.2 (Open Ball) *Given a metric space* $(S, \mathbf{d})$*, a point* $p \in S$*, and some real* $\epsilon > 0$*, the* ***open ball*** $B(p, \epsilon)$ *in* $(S, \mathbf{d})$ *centred at* p *of radius* ϵ *is the subset of* S *defined by*

$$B(p, \epsilon) = \{q \in S : \mathbf{d}(p, q) < \epsilon\}$$

the ***closed ball*** $B[p, \epsilon]$ *in* $(S, \mathbf{d})$ *centred at* p *of radius* $\epsilon > 0$ *is the subset of* S *defined by*

$$B[p, \epsilon] = \{q \in S : \mathbf{d}(p, q) \leq \epsilon\}.$$

A ***neighbourhood*** *of a point* $p \in S$ *is a subset* N *of the metric space* $(S, \mathbf{d})$ *such that there exists an open ball* $B(p, \epsilon) \subset U$.

For every point $p \in S$, if the non-empty collection τ of neighbourhoods $\{N : N$ is a neighbourhood of $p \in S\}$ obey the following four axioms then the collection τ is called a ***Topology*** [20] :

(a) p lies in each of its neighbourhoods.

(b) The intersection of neighbourhoods of p itself is a neighbourhood of p.

(c) Every set containing neighbourhood of p is a neighbourhood of p.

(d) If N is a neighbourhood of point p and if N° is a set such that $N^\circ = \{u \in N : N \text{ is a neighbourhood of } u\}$ then N° is a neighbourhood of p.

The structure formed by order pair (S, τ) is called ***Topological space***.

With open balls defined, it is possible to construct open sets with a collection of open balls. Thus an open set can be defined as following [19].

Definition 1.3.3 *A subset U of metric space $(S, \mathbf{d})$ is open, if for every $p \in U$ there is an open ball centred at p in U.*

Intuitively a subset U in metric space $(S, \mathbf{d})$ is open if there exist some real $\epsilon > 0$ such that for every $p, q \in U$ the distance $\mathbf{d}(p, q)$ is less than ϵ. Similarly a closed set can be defined as the following.

Definition 1.3.4 *A subset U of metric space $(S, \mathbf{d})$ is closed, if the set $S \setminus \{U\}$ is open.*

Given a subset U of a space, a collection of subsets $\{U_i\}_{iin\mathbb{I}}$ such that $I = \{1, 2, \cdots\}$ are said to *cover U* if $U \subseteq \bigcup_{i \in \mathbb{I}} U_i$. If the cover is formed by a collection of open sets, then such a cover is called as *open cover.*

Defining the open sets gives a way to define connectedness of the metric space under consideration. If a metic space is connected, every point belonging to the space has an open ball of some radius with the point under consideration that belongs to the metric space. Intuitively thus, a space is connected if it can not be represented as union of two disjoint open sets. Formally the connected space can be defined, based on the fact that the metric space $(S, \mathbf{d})$ and $\emptyset$ are closed open sets [19], as

Definition 1.3.5 *A metric space* $(S, \mathbf{d})$ *is connected if the only sets of the space* $(S, \mathbf{d})$ *that are closed and open are* $\emptyset$ *and the space itself.*

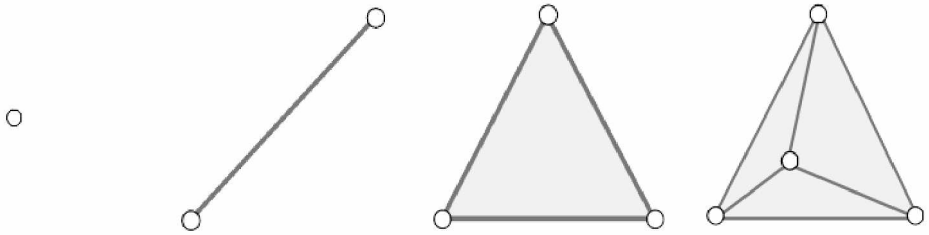
It ought to be noted that in a connected topological space (S, τ) , each point $u \in S$ has a collection of neighbourhoods or open sets centred at that point such that the aforesaid four axioms hold. However, it is not always the case that a topological space has to be connected. Connectedness plays a crucial role in identification of similar objects, *cluster analysis* or *clustering*. By varying the radius r of a ball B_r centred at point p, it is possible to identify other points of the space which exhibit similar properties. The clusters or the partitions of the data (a set of points originating from $\mathbb{R}^d$) are subjective to the radius of the ball under consideration.

1.3.2. Simplicial Complex and Nerve of a Cover. Let a collection of points $u_0, u_1, u_2, \cdots u_k$ be such that $u_i \in S \ \forall i \in \{0, 1, 2, \cdots, k\}$. The linear combination of these points $\sum_{i=0}^{k} \alpha_i u_i$ spans a hyperplane $\forall \alpha_i \in \mathbb{R}$ such that $\sum_{i=0}^{k} \alpha_i = 1$. If two linear combinations $x = \sum_{i=0}^{k} \alpha_i u_i$ and $y = \sum_{i=0}^{k} \beta_i u_i$ are equal if and only if $\forall i[\alpha_i = \beta_i]$, then the points are said to be affinely independent [21]. Given a such $k+1$ affinely indpendent points $u_0, u_1, u_2, \cdots u_k$, the smallest convex set containing them defined by the linear combination

$$x = \sum_{i=0}^{k} \alpha_i u_i$$

where $\alpha_i > 0$ and $\sum_{i=0}^{k} \alpha_i = 1$ as the ***simplex of dimension k (or k-simplex)***. For example, Figure 1.4 shows basic simplices as presented in [21].

It is possible to approximate a topological space (S, τ) (with $S \subset \mathbb{R}^d$) using a collection of simplices that can represent the geometric shape of the topological space under consideration [22]. The idea is based on covering surface of the geometric shape with a grid of triangles - that form simplices. Such an approximaiton method is called as *Triangulation* (please refer to chapter-6 of [20] for further details). The approximating



(a) o-simplex (b) 1-simplex (c) 2-simplex (d) 3-simplex

Figure 1.4. Examples of simplices. From left to right:A vertex, an edge, a triangle, and a tetrahedron. An edge has two vertices. A triangle has three vertices and three edges. A tetrahedron has four vertices and six edges forming four faces [21].

collection of simplices fit together in space by either having a vertex, an edge, or a face. Such a collection of simplices form a *simplicial complex*, which is formally defined as the following [20].

Definition 1.3.6 *A **Simplicial Complex** K is a finite collection of simplices such that whenever a simplex lies in the collection then so does all of its faces, and whenever two simplices in the collection intersect, the intersection is either empty or a common face of both.*

An immediate question would be how to construct a *simplicial complex* given a set of points $\mathcal{A}$ from some Euclidean space? The simplest way to construct a *simplicical complex* identify which set of points are with a certain distance from each other and establish a link, which act as edges, each pair considering them as vertices of the *simplicical complex*. This can be achieved by using open balls of some radius $r > 0$. Notice that this leads to identification of connected subspaces of the entire topological metric space under consideration and that this connectivity is subjective to the radius of the ball. By increasing the radius of the ball from some $r \to 0$, it can be noticed that simplicial complex changes from without any edges to one containing all possible links between vertices, forming a convex hull, for some radius $r = r_c$ and any further increase in radius ($r > r_c$) doesn't have any change in

the shape of the complex as shown in the Figure 1.5. For intermediate value of the radius $0 < r < r_c$ each simplicial complex represents a partition (or cluster) with in the data.

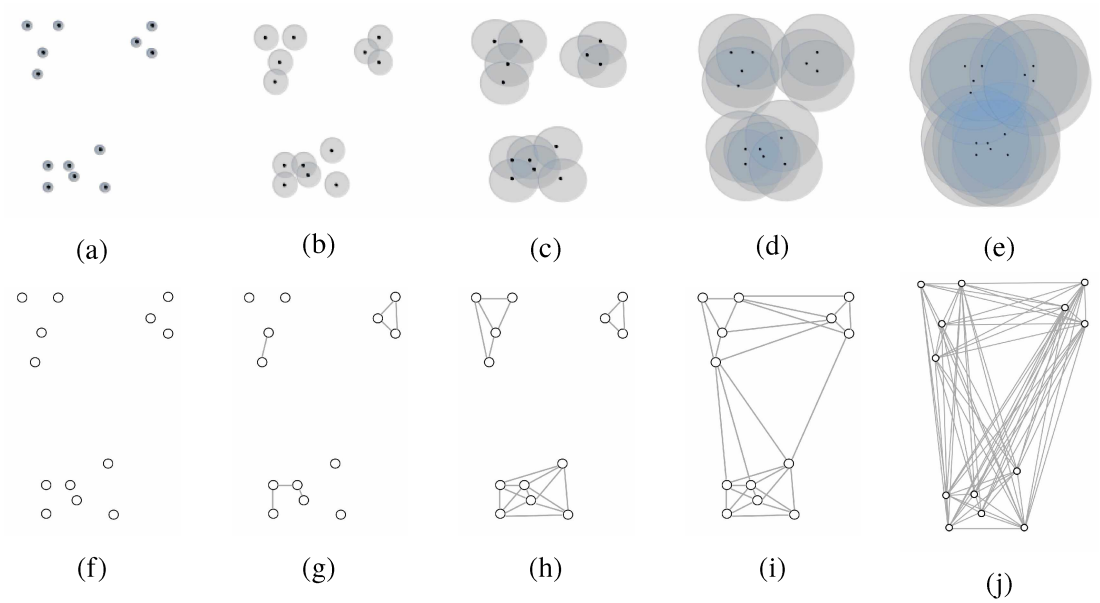


Figure 1.5. Change in identified simplicial complexes with change in open-ball radius. Top from left to right: (a) the point cloud data with a ball around each point indicating its neighbourhood. (b-c) As the radius of the ball increases the number of intersections between neighbourhoods increase and partitions of interest emerge. (d-e) Beyond certain radius of the ball, the point cloud can be considered as a single cluster due to intersecting neighbourhoods. Bottom left to right: corresponding simplicial complexes for Sub-figures (a), (b), (c), (d), and (e) respectively. Notice that for (j) the identified simiplicial complex has an edge between each pair of vertices.

A simplicial complex can also be constructed using collection of subsets C of space. With each subset represented as a vertex, the intersection can be represented as a link or an edge between them. Such representation lends a way to find the shape formed by the subsets and unshroud the connectivity between them. Further, the identified simplicial complex provides a graphical representation. This leads to another question: Would the simplicial complex be topological equivalent of the shape formed by the subsets under consideration? The answer to the question is given by *Nerve Theorem*. However, it ought to define *Nerve*.

Definition 1.3.7 *A **Nerve** is a finite collection of subsets of C such that all its non-empty subcollections have a common intersection which is not empty.*

$$Nrv(C) = \{X \subseteq C \mid \cap X \neq \varnothing\}$$

It should be noted that Nerve of a collection C can be seen to contain a simplicial complex which is closed under subset operation, i.e, every subset of a sub-collection is contained in the complex. The Figure 1.6 illustrates the Nerve of the open cover, shown in Sub-figure 1.6c, for the data-set shown in Sub-figure 1.6a. For a point cloud data, collection C is a cover of the point cloud data. The $Nrv(C)$ might not always be topological equivalent of the object from which the data originates. However, if certain conditions are met, then Nerve of cover C can be a weak equivalent of the object. The weak notion of equivalence is defined by *Homotopy* [1]. The idea behind *Homotopy* is that if two objects share topological similarity then it is possible find a continuous path in the space of functions that define the structure of two objects under consideration. This means one object can be deformed into another and therefore weak equivalence.

Definition 1.3.8 *Let $\mathcal{P}$ and $\mathcal{Q}$ be topological spaces and $f, g : \mathcal{P} \to \mathcal{Q}$ be continuous functions. The function f is **homotopic** to g if there exists a continuous function $\mathcal{H} : \mathcal{I} \times \mathcal{P} \to \mathcal{Q}$ such that $\mathcal{H}(0, x) = f(x)$ and $\mathcal{H}(1, x) = g(x)$ and the homotopy between f and g is represented as $f \simeq g$.*

All convex spaces and continuous spaces without any holes, a sphere for example, can be shrunk into a point, hence homotopic to a point. Further, spaces that can be continuously be shrunk to a point are called as *contractible*. The sufficient condition for Nerve of a cover to be homotopic to the subspace is given by *Nerve Theorem*[18, 20, 21], which will be stated without any proof.

[1] from Greek homos = same, topos = place

Theorem 1.3.1 (Nerve Theorem) *Let C be a finite open cover of topological space $\mathcal{S}$. If intersection of any collection of subsets in C is either contractible or empty, then the $Nrv(C) \simeq \mathcal{S}$.*

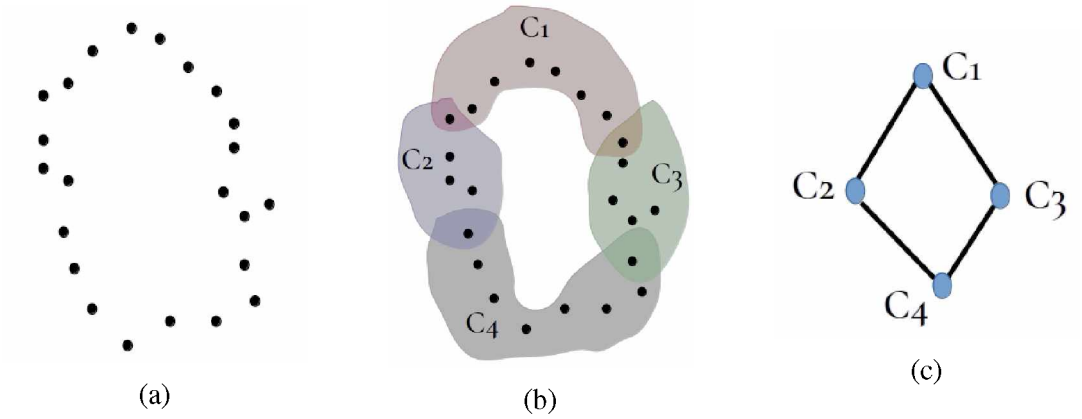


Figure 1.6. The Nerve of the open cover for the data-set shown in Sub-figure (a). Sub-figure (b) shows the open cover formed by sets C_1, C_2, C_3, and C_4. Sub-figure (c) shows the simplicial complexes or the nerve of the open cover with each set in sub-figure (b) represented by a node. Note that each edge of simplex indicates the intersections between the two sets of open cover.

PAPER

I. TOPOBARTMAP: BICLUSTERING ARTMAP WITH OR WITHOUT TOPOLOGICAL METHODS IN A BLOOD CANCER CASE STUDY

Raghu Yelugam[1], Leonardo Enzo Brito da Silva[1,2], Donald C. Wunsch II[1]

[1] Applied Computational Intelligence Laboratory,
Missouri University of Science and Technology, Rolla, MO, USA

[2] CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil

Email: ry222@umsystem.edu, {leonardoenzo, wunsch}@ieee.org

ABSTRACT

Biclustering is a special case of subspace clustering that has become viable in several domains. Particularly, in genomic data analysis, biclustering has been used to identify conditions under which a subset of genes are highly co-expressed, while topological data analysis has been used to analyze disease-specific subgroups, evolution, and disease progression. In this work, we combine biclustering with topological data analysis to achieve the best of both methods. We present TopoBARTMAP - produced by hybridizing BARTMAP, an adaptive resonance theory (ART)-based biclustering method, with TopoART, a topology learning ART network - in order to identify topological associations between biclusters. TopoBARTMAP outperformed both TopoART and BARTMAP in the experimental analysis on six benchmark blood cancer data sets. In some cases, BARTMAP may nevertheless be preferred due to implementation simplicity.

1. INTRODUCTION

Biclustering, at it's crux, is the problem of finding a subset of samples with high association across a subset of features. It has produced significant results in many different fields, starting with the application to gene expression data analysis [1]. Over the past decade, several algorithms were proposed under co-clustering [2], biclustering [3, 4], and subspace clustering [5]. Biclustering applied to gene expression data is used to identify subsets of conditions under which subsets of genes are highly co-expressed. Such identification is essential for gaining insights into regulatory networks [6] and gene-disease associations. However, biclustering can fail to completely uncover gene regulatory networks [7] due to the inability to identify functional associations between genes within a bicluster and to group functionally related genes that might not co-express significantly.

In an attempt to mitigate this problem, topological data analysis is becoming popular for gene expression data [8, 9]. In general, topological analysis extends clustering to identify local relationships in the data. Depending on the proximity of clusters, the identified clusters represented by cluster points are connected to generate a high-level graphical representation of the underlying shape of the space. Since topological methods discover the geometric structure within the data, these methods are sensitive to large and small scale patterns, invariant to noise, independent of the coordinate system, and can produce a compressed representation of data [10, 11, 8]. In practice, topological data analysis methods, such as Mapper [12], had proven to be effective in identifying subgroups of cancers [13], understanding genome dynamics [14] and cellular differentiation [15], and disease progression [16], which would otherwise be difficult to detect using traditional clustering methods applied to biomedical research [17].

In this work, we take a step towards combining topological data analysis with biclustering. To do so, we combine BARTMAP [18] with TopoART [19] to produce TopoBARTMAP, which can simultaneously identify the structure within the data and the biclusters. This algorithm, while identifying inter-bicluster relationships, is resilient to noise

due to the prototype pruning mechinism of TopoART. While the literature contains a few topological biclustering algorithms, such as BiTM [20], wBiTM [21], and SKB[7], to the best of our knowledge TopoBARTMAP is the first adaptive resonance theory (ART) [22] based topological biclustering algorithm. Amongst the other methods, the BiTM (Biclustering using Topological Maps) identifies biclusters by simultaneously clustering rows and columns by identifying a discrete topology. The biclustering method wBiTM, based on BiTM, is a feature group weighting method that uses topological maps for biclustering. Both BiTM and wBiTM produce topological maps that enable data visualisation. SKB, Skeleton Biclustering, is capable of identifying and mining missing genes in the biclusters while building inter-bicluster and intra-bicluster (relationships among genes within a bicluster) relationship skeletons. This method uses hierarchical biclustering and gene ontology annotations.

The remainder of this paper is organized as follows: we start with a brief review of related ART models [22] in Section 2, which is necessary to contextualize the main contribution of this paper introduced in Section 3, which is the TopoBARTMAP model. The details concerning the experimental design are presented in Section 4, while Section 5 reports and discusses the biclustering results on six benchmark blood cancer data sets. Although in our experimentation we compare TopoBARTMAP only with one biclustering algorithm, BARTMAP, in other papers [23, 18, 24] BARTMAP was already compared with other ART-based and non-ART-based methods and found to be superior. Finally, Section 6 summarizes the findings of this paper.

2. BACKGROUND

This section presents an overview of Fuzzy ART [25], TopoART [19] and BARTMAP [18]. We refer the reader to [26, 27, 28, 29, 30] for a comprehensive treatment of ART models and their applications.

2.1. FUZZY ART

This section exists to make the paper self-contained. The reader familiar with Fuzzy ART can skip to Section II.B. Readers new to ART will gain useful context from this section but will need the cited papers to understand ART well.

Several winner-take-all neural models have been proposed to perform supervised and unsupervised learning based on ART. Of them, Fuzzy ART, described in [26], extends the binary ART-1 by using fuzzy operations and is capable of recognising stable codes (prototypes) in response to real or binary-valued inputs. A typical ART-based learning model consists of two interacting fields: F1 and F2. Field F1 receives input as a vector X which is normalized and complement coded, i.e., if $x = [x_1, x_2, ..., x_d]$ denotes the normalized sensory input vector with d features, then $X = [x, 1 - x^c]$, where $x_i^c = 1 - x_i$. It is customary to consider each dimension of the input X as a node. Field F2 consists of several prototypes (neurons) - one per node - which are used to categorize the input pattern. The field F1 generates bottom-up sensory information which produces activity across each prototype in field F2 using bottom-up memory traces. The highly active prototype produces a top-down expectation signal based on the top-down memory traces. At the orientation subsystem, this expectation signal is compared with the sensory information produced by field F1 for similarity; this is known as vigilance criterion. If the similarity is greater than a certain threshold called the vigilance value, resonance ensues between the highly active prototype and field F1; the corresponding memory traces are updated. In any other case, the highly active prototype is shut down, and a search is done to find the next active prototype that produces a matching expectation. If no such prototype is found, a new node is created to register the current input, and this prototype is said to be committed.

Bottom-up memory traces can be represented using top-down memory traces, which essentially is a key feature of Fuzzy ART. In this work, we represent the bottom-up memory traces as a vector for the j^{th} node in field F2 as weights $W_j = [W_1^j, W_2^j, ..., W_{2d}^j]$. Initially, field F2 consists of uncommitted prototypes. Instead of using two different memory traces,

Fuzzy ART has traces or weights (W) connecting F1 to F2 and F2 to F1, i.e., the prototypes are directly compared with the input signal. Because the input, X, to Fuzzy ART is normalized and complement coded, which is implicitly assumed in the models derived from Fuzzy ART, all the uncommitted node weights are initially set to a value of 1. During the input presentation, field F2 has an uncommitted neuron along with committed neurons. A measure of activity is computed using the category choice function defined by

$$T_j = \frac{|X \wedge W_j|}{\alpha + |W_j|}, \tag{1}$$

where $|\cdot|$ is the ℓ_1 norm, $\wedge$ is the fuzzy AND operation defined by $(X \wedge W_j)_i = min(x_i, W_i^j)$, $X \in \Re^{2d}$, $W \in \Re^{2d}$, and $\alpha > 0$ helps in breaking the tie when more than one prototype is a fuzzy subset of the input pattern. A winning neuron J is selected using winner-take-all rule. A category match is measured using a match function defined by

$$M_J = \frac{|X \wedge W_J|}{|X|}. \tag{2}$$

If M_J is greater than the vigilance parameter, ρ, it meets the vigilance criterion. In case of failure to satisfy, this winning neuron is shut off and the next winning neuron is tested against the vigilance criterion. The search process continues until the vigilance criterion is satisfied. When a neuron is found that satisfies the vigilance criterion, its weight vector or prototype is updated using the learning law defined as

$$W_J = \beta(W_J \wedge X) + (1 - \beta)W_J, \tag{3}$$

where $0 < \beta \leq 1$ is the learning rate. The output of the network y^{F2} is set for each i^{th} node as

$$y_i^{F2} = \begin{cases} 0, & \text{if } i \neq J \\ 1, & \text{if } i = J \end{cases}. \tag{4}$$

In the case of fast learning, when an uncommitted prototype is selected, β is set to a value of 1. Each prototype learned this way summarizes all the input patterns it is associated with and thus represents either a cluster or cluster point. Due to complement coding, it is possible to estimate the size of the cluster and region of space that each prototype summarizes [31].

2.2. TOPOART

Topological ART, or TopoART [19], is a Fuzzy ART-based hierarchical clustering model that learns topologies present within the input data. All rules used for prototype learning of Fuzzy ART remain the same in TopoART. However, TopoART uses a counter, n_i, for each i^{th} prototype that records the number of samples it summarized. After τ input presentations (time steps), the prototypes with a counter less than the threshold ϕ, called candidate prototypes, are pruned from field F2. Hence, the prototypes with $n_i \geq \phi$ are called permanent prototypes. This pruning mechanism renders TopoART robust to noise. Further, along with the best matching prototype J_1, a second highly active prototype J_2 satisfying the vigilance criterion is selected (namely, the second-best). This prototype's weights are updated using the same update rule defined by Eq. (3), but with a smaller learning rate, i.e.,

$$\beta_{J_2} < \beta_{J_1}. \tag{5}$$

This procedure of identifying and updating the second-best prototype's weights increases noise robustness by making growth of prototypes in pertinent areas of the input space more likely [19].

To learn topologies, an edge is established between prototypes J_1 and J_2 (if J_2 can be found), creating a topological structure with prototypes acting as nodes. While edges formed with candidate prototypes are removed when these prototypes are pruned, edges between permanent prototypes remain stable. Each one of the permanent prototypes is assigned

with a label depending on its topological association with others, i.e, all the connected permanent prototypes are assigned with the same label. These labels are recorded in a vector - $\overrightarrow{C} =< l_i : l_i \in \{1, 2, ..., k\} >$, where k is the number of topologies identified and l_i is the label for the i^{th} permanent prototype in field F2. With the aid of this vector, one can identify the topology to which an input datum belongs. Unlike Fuzzy ART, the clusters are represented by these topologies rather than the prototypes, which enables the identification of arbitrarily shaped clusters. The input patterns that correspond to the deleted prototypes are reclassified by associating them to the permanent prototypes that yield the strongest responses. This activity is computed using the following rule [19]:

$$T_j = 1 - \frac{|(X \wedge W_j) - W_j|}{|X|}. \tag{6}$$

Further, TopoART achieves hierarchical clustering by cascading multiple Fuzzy ART modules, with each having an increasing vigilance parameter value. The vigilance parameters across the modules in the hierarchy are computed using the following rule,

$$\rho_{next} = \frac{1}{2}(1 + \rho_{previous}), \tag{7}$$

where ρ_{next} and $\rho_{previous}$ are the vigilance values for the next and previous modules respectively. The module at the lowest hierarchical level receives all the input patterns, whereas subsequent modules receive an input pattern if and only if that pattern activates a permanent prototype in the previous module. When new data is presented, inference is made using Eq. (6) and the output y^{F2} for each module of TopoART is computed with Eq. (4) using only permanent prototypes.

2.3. BARTMAP

Biclustering ARTMAP, or BARTMAP [18], is an ART-based two-way clustering method that identifies biclusters present in the data adaptively without the explicit requirement of prior knowledge of the number of biclusters. We refer the reader to [32] and [33] for detailed expositions on ARTMAP and Fuzzy ARTMAP as BARTMAP is inspired by the theory of Fuzzy ARTMAP [33]. Here, we present a brief treatment of BARTMAP while retaining pertinent details. In doing so, we use the following notation. We represent the gene expression data matrix as $G = (F, O)$, where $F = \{f_1, ..., f_N\}$ represents the set of N features (or rows) and $O = \{o_1, ..., o_M\}$ represents the set of M observations (or columns). An element $g_{ij} \in G$ represents the intensity of feature i in observation j. In using this notation, which we did for ease of generalization, we considered each gene as a feature and each sample as an observation throughout the exposition.

In construction, BARTMAP consists of two Fuzzy ART modules, ARTa and ARTb, connected via an inter-ART module. While the ARTb module receives features as input, the ARTa module receives observations as input. In essence, BARTMAP learns a mapping between the clusters found by the ARTa and ARTb modules across observations and features, respectively, such that local relationships are preserved and captured by the produced biclusters. The learning in BARTMAP progresses in two steps. The first step is to identify clusters across one of the dimensions, let us say the row dimension, of the matrix. This is achieved by using the ARTb module as a standard Fuzzy ART, thus resulting in k_f feature prototypes or clusters $\{F_i | i = 1, ..., k_f\}$. In the next step, observations (or columns) are presented to the ARTa module. At this stage, when a new observation o_k is presented, the potential associable cluster is identified using winner-take-all. Eventhough the winner qualifies via the vigilance test, at this point, it might not represent the observation unless the winner corresponds to an uncommitted prototype. If the winner corresponds to an uncommitted prototype, its weights are updated. However, if the winner is a committed prototype, then the current sample o_k is associated to it if and only if the sample exhibits a

similar trend across at least one feature cluster F_i, as do the other member observations that are associated to the winning prototype. This similarity is verified using the average Pearson correlation coefficient that is computed as

$$\eta(o_k, O_J, F_i) = \frac{1}{M_J}\sum_{l=1}^{M_J} r(o_k, o_l, F_i), \tag{8}$$

where,

$$r(o_k, o_l, F_i) = \frac{\sum_{t=1}^{N_i} (g_{o_k f_{it}} - \overline{g}_{o_k F_i})(g_{o_l f_{it}} - \overline{g}_{o_l F_i})}{\sqrt{\sum_{t=1}^{N_i} (g_{o_k f_{it}} - \overline{g}_{o_k F_i})^2}\sqrt{\sum_{t=1}^{N_i} (g_{o_l f_{it}} - \overline{g}_{o_l F_i})^2}}, \tag{9}$$

$$\overline{g}_{o_k F_i} = \frac{1}{N_i}\sum_{t=1}^{N_i} g_{o_k f_{it}}, \tag{10}$$

$$\overline{g}_{o_l F_i} = \frac{1}{N_i}\sum_{t=1}^{N_i} g_{o_l f_{it}}, \tag{11}$$

and o_k is a new observation, o_l belongs to observation cluster $O_J = \{o_1, \ldots, o_{M_J}\}$ with M_J observations, and the set $F_i = \{f_{i1}, \ldots, f_{iN_i}\}$ represents the i^{th} feature cluster with N_i features. The observation o_k becomes a member of cluster O_J if and only if $\eta(o_k, O_J, F_i)$ is greater than a threshold η_{th} for at least one of the feature clusters. Cluster O_J is represented by ARTa's winning prototype J; therefore, if the previously mentioned constraint is satisfied, then its weights are updated using Eq. (3). If the test fails, match tracking ensues [32], i.e, the winning prototype is inhibited by increasing the vigilance value ρ_a of ARTa sufficiently above its base value. With the inhibition, a search for the next winning prototype ensues. Once a suitable cluster is identified or a new cluster is created, the vigilance value ρ_a of ARTa is reset to its base value.

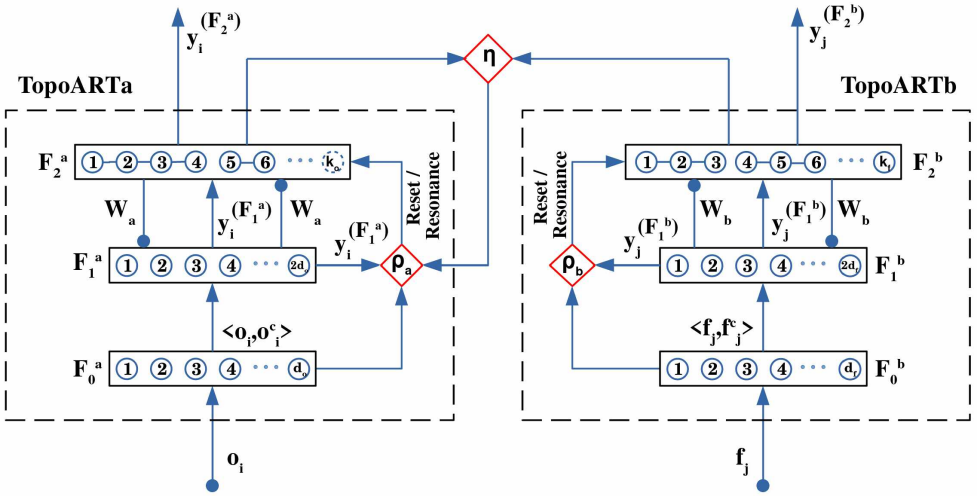


Figure 1. Block diagram of TopoBARTMAP. TopoBARTMAP consists of two TopoART modules whose clusters are related by correlation test module. While TopoARTb is shown to have learned k_f prototypes, TopoARTa is shown to have learned k_o prototypes (note that prototype labeled k_o is a temporary prototype).

3. TOPOBARTMAP

This section discusses the architectural novelty introduced in this paper. To understand the topological associations between the features (genes) and observations of a given gene expression matrix, $G = (F, O)$, BARTMAP is modified to produce topological biclusters. This is achieved by replacing both Fuzzy ART modules of BARTMAP with TopoART modules. The resulting biclustering method, TopoBARTMAP, is more robust to noise and can identify associations better than BARTMAP. While the construction and much of the training procedure remain the same, i.e., the training proceeds in two steps with a similarity/association test using average Pearson correlation, there are subtle differences rendered by the nuances in the training of TopoART.

Just as in the training of BARTMAP, one of the dimensions of matrix G, features F for example, are presented to TopoARTb during the first stage, resulting in identifying k_f prototypes $\{F_i | i = 1, ..., k_f\}$ and k topologies or clusters formed by these prototypes. During the second stage of training, the observations are presented to TopoARTa. Similar to

BARTMAP, a winning prototype is identified and verified for a match using the vigilance test (Eq. (2)) and associability with the similarity test (Eqs. (8) to (11)). If the presented observation exhibits the same trend as the other observations summarized by the winner J_1 across at least one of TopoARTb's k_f prototypes, it is associated with the winner. Otherwise, match tracking ensues and the search proceeds until either another representative prototype is found or an uncommitted prototype is assigned. Further, a second-best prototype J_2 is identified using winner-take-all and verified for the match. An edge is established if and only if the presented observation exhibits a trend similar to the other observations represented by the second winner across at least one of TopoARTb's k_f prototypes. If the similarity test fails, then the current second winner is inhibited and the next winner is searched for. This procedure terminates once a second winner passing the vigilance and similarity tests is found, or all the prototypes are inhibited. If a qualifying second-best prototype exists, then its weights are updated using Eq. (3). The pruning of candidate prototypes and reclassification of the observations represented by the pruned prototypes are conducted without any changes to TopoART. The learning results in the identification of topologies, which represent biclusters, enabling TopoBARTMAP to identify arbitrarily shaped biclusters. Similar to TopoART, the permanent prototypes are used for making an inference when new data are presented. The TopoBARTMAP architecture is depicted in Fig. 1, and its working procedure is summarized by Algorithm 1.

4. EXPERIMENTATION

To verify and compare the performance of TopoBARTMAP, we used six benchmark blood cancer data sets brought together by [34]. The corresponding information about the data sets, number of features, number of observations, and number of classes per data set is available in Table 1. Of these data sets, Alizadeh-V1 and Alizadeh-V2 [35] have a randomized order of observations, while the rest of the data sets have observations ordered according to the classes. We used MATLAB (Statistics and Machine Learning Toolbox)

Algorithm 1: TopoBARTMAP

Input : G, {TopoART a and b parameters}, η_{th}, δ
Output : $y^{(O)}$ observation classes
:
/* Notation */
O_j : j^{th} cluster of observations.
F_i : i^{th} cluster of features.
$\eta(o_k, O_j, F_i)$: correlation between o_k and O_j across F_i.
/* Training */
1 **for** $f \in F$ **do** // TopoARTb
2 | Train TopoARTb using $< \rho_b, \phi_b, \tau_b, \alpha_b, \beta_{J_1}, \beta_{J_2} >$
3 **end**
4 Initialize time step counter t_a to 0
5 **for** $o_k \in O$ **do** // TopoARTa
6 | Increment t_a
7 | Compute $T_j \; \forall j$ using Eq. (1)
8 | Find first winner: $J_1 = \arg\max_j (T_j)$
9 | Compute M_{J_1} using Eq. (2)
10 | **if** $M_{J_1} \geq \rho_a$ **then**
11 | | Compute $\eta(o_k, O_{J_1}, F_i) \; \forall \; F_i$ using Eq. (8)
12 | | **if** $\exists F_i | \eta(o_k, O_{J_1}, F_i) \geq \eta_{th}$ **then**
13 | | | Update W_{J_1} using Eq. (3), $n_{J_1} = n_{J_1} + 1$
14 | | | reset ρ_a
15 | | | Find second winner: $J_2 = \arg\max_{j, j \neq J_1} (T_j)$
16 | | | **if** $(M_{J_2} \geq \rho_a) \wedge \exists F_i [\eta(o_k, O_{J_2}, F_i) \geq \eta_{th}]$ **then**
17 | | | | Update W_{J_2} using Eq. (3)
18 | | | | Establish a link between J_1 and J_2
19 | | | **else if** $\forall F_i \forall J_2 (\eta(o_k, O_{J_2}, F_i) < \eta_{th})$ **then**
20 | | | | Stop the search
21 | | | **else**
22 | | | | Shut down J_2
23 | | | | Go to line 15
24 | | | **end**
25 | | **else if** $(\nexists (M_{J_1} \geq \rho_a)) \parallel (\rho_a \geq 1)$ **then**
26 | | | $W_{new} = o_k, n_{new} = 1$
27 | | | reset ρ_a
28 | | **else**
29 | | | Shut down J_1, $\rho_a = \rho_a + \delta$
30 | | | Go to line 8
31 | | **end**
32 | | **if** $t_a \bmod \tau_a = 0$ **then**
33 | | | Remove clusters with counter less than ϕ_a
34 | | **end**
35 **end**
36 Identify and label the topologies using the connected prototypes
/* Inference */
37 **for** $o_l \in O$ **do**
38 | Compute $T_j \; \forall j$ using Eq. (6)
39 | $y_l = \arg\max_j (T_j)$ and Cluster $= l_{y_l}$
40 **end**

Table 1. Data sets used for experimental analysis [4]. Here, Sl.no, N_o, N_f, and N_c refer to serial number, number of observations, features, and Classes within observations, respectively.

Sl.no	Data set	N_o	N_f	N_c
1	Alizadeh-V1 [35]	42	1095	2
2	Alizadeh-V2 [35]	62	2093	3
3	Alizadeh-V3 [35]	62	2093	4
4	Armstrong-V1 [37]	72	1081	2
5	Armstrong-V2 [37]	72	2194	3
6	Shipp-V1 [38]	77	798	2

The data sets are available at `https://github.com/padilha/biclustlib`.

and the Cluster Validity Analysis Platform toolbox [36] to conduct all the experiments. To enhance the reproducibility of this research, the MATLAB source code is available at the Applied Computational Intelligence Laboratory GitHub repository[1].

To investigate the performance of TopoBARTMAP on each of these data sets, we ran tests with a Genetic Algorithm (GA) [39] to optimize parameters $< \rho_a,\ \beta^a_{J_2},\ \phi_a,\ \tau^{\%}_a,\ \rho_b,\ \beta^b_{J_2},\ \phi_b,\ \tau^{\%}_b,\ \eta >$ with the Adjusted Rand Index [40] as the performance measure, where $\tau^{\%}_{-}$ represents the percentage of the total number of samples presented to each module. To measure the biclustering performance, we considered topologies found by TopoARTa as biclusters. Although topologies can be retrieved from the TopoARTb module, for the correlation test, individual prototypes rather than topologies were considered. Further, the vectors $< 0.0, 0.0, 0.0, 0.1, 0.0, 0.0, 0.0, 0.1, 0.0 >$ and $< 0.95, 1.0, 5.0, 0.3, 0.95, 1.0, 5.0, 0.3, 0.99 >$ were used as lower and upper bounds for parameters with the aforementioned order during optimization, along with the following constraints:

$$\phi_a - \tau^{\%}_a N_o \geq 0, \tag{12}$$

and

$$\phi_b - \tau^{\%}_b N_f \geq 0, \tag{13}$$

[1]https://github.com/ACIL-Group/TopoBARTMAP

where $N_o = \#\{observations\}$ and $N_f = \#\{features\}$. Note that $\tau_a = \lfloor \tau_a^{\%} N_o \rceil$ and $\tau_b = \lfloor \tau_b^{\%} N_f \rceil$. The GA was run 10 times for 25 generations with a population size of 200 agents. Throughout the experiments, β_a and β_b were set to 1, i.e., fast learning is used with the value of choice parameter α set to 0.001, match tracking step size set to 0.01, and data presented for one epoch.

BARTMAP, TopoART, and Fuzzy ART were run on the same data sets in order to compare their performance with TopoBARTMAP's. For BARTMAP, $< \rho_a, \rho_b, \eta >$ were optimized using GA, with the lower and upper bounds as given by vectors $< 0.0, 0.0, 0.0 >$ and $< 0.95, 0.95, 0.99 >$, respectively. For TopoART, $< \rho, \beta_{J_2}, \phi, \tau^{\%} >$ were optimized using GA, with the lower and upper bounds for these parameters given by vectors $< 0.0, 0.0, 0.0, 0.1 >$ and $< 0.95, 1.0, 5.0, 0.3 >$, respectively, as well as the inequality constraint

$$\phi - \tau^{\%} N_o \geq 0, \tag{14}$$

where $\tau^{\%}$ and τ are defined as in TopoBARTMAP. Note that unlike the standard TopoART network, which uses two modules for hierarchical purposes, here only one module was trained, and performance was measured using topologies identified by TopoART. While these other methods were run with GAs similar to TopoBARTMAP, Fuzzy ART was run with a parameter sweep. In particular, for running Fuzzy ART on each of the six data sets, ρ was varied amongst 4960 values uniformly sampled from $[0, 1]$, which corresponds to roughly the same amount of fitness evaluations of the GAs used to optimize the other ART networks. For all of these methods, fast learning was used, with the value of choice parameter α set to 0.001, and data were presented for one epoch. For BARTMAP, the match tracking step size was set to 0.01, similar to TopoBARTMAP.

All agglomerative clustering algorithms (which includes online methods), such as most unsupervised ART networks, are prone to ordering effects (see [41, 42, 43] and the references therein for further discussion and methods to mitigate order dependency). Therefore, once the optimal parameters that produce the best Adjusted Rand Index on each

of the six data sets were determined, tests were conducted to verify the hyper-parameter sensitivity of each method. To do this, each data set is randomized with respect to columns and rows separately to produce ten different data sets. Performance on each of these data sets was measured using the Adjusted Rand Index with model parameters set to the optimal values for the corresponding original data set. Later using GA with corresponding aforementioned settings, TopoBARTMAP and BARTMAP had their parameters optimized for the data sets produced by randomization to identify the sensitivity to order of presentation. As TopoART and Fuzzy were not sensitive to changes in feature order, using the respective settings mentioned previously, parameters were optimized for TopoART using GA, and Fuzzy ART's ρ was varied to measure the performance on data sets with randomization in order of observations.

5. RESULTS AND DISCUSSION

We compare the performance of TopoBARTMAP, BARTMAP, TopoART, and Fuzzy ART on the aforementioned experiments in this section.

Table 2 shows the clustering performance of the contenders measured using the Adjusted Rand Index. For the original data sets, TopoBARTMAP outperforms the other three algorithms on all six data sets. While not being a biclustering algorithm, TopoART outperforms BARTMAP on four out of six data sets, and its performance is comparable with TopoBARTMAP on the Armstrong-V1 [37] and Shipp-V1 [38] data-sets. On the other hand, BARTMAP outperforms TopoART but loses to TopoBARTMAP on Alizadeh-V1 and Alizadeh-V3 [35]. We attribute TopoART's better performance over BARTMAP to the topological learning and order of observation presentation. Of all the algorithms compared, Fuzzy ART fails to identify better clusters and consequently performs poorly.

The results of experimentation on data sets with a randomized order of observations and features with parameters fixed to optimal parameters are summarized in the Tables 3 and 4, respectively. Unsurprisingly, none of the contenders, TopoBARTMAP, BARTMAP,

TopoART, or Fuzzy ART, perform better with the order of observation presentation changed. Since TopoART and Fuzzy ART are impervious to the randomization of the features, the results summarized in Table 2 are valid for this case. Due to this, the performance of TopoBARTMAP and BARTMAP are juxtaposed in Table 4, where BARTMAP wins four out of six times in experimentation with randomized features. These tests reveal that TopoBARTMAP is more sensitive to parameterization than are BARTMAP or TopoART.

The results of order sensitivity analyses run with GA optimization (and parametric sweep for Fuzzy ART) are summarized in Tables 5 and 6. While the performance of each of the algorithms differs noticeably, TopoBARTMAP still outperforms the others. In contrast to the original-order experiments, in the randomized cases, BARTMAP outperforms TopoART on all randomized observation order data sets. To estimate the statistical significance of the performance differences on these data sets, a t-test was conducted among TopoBARTMAP, BARTMAP, and TopoART. The results indicate that the performances of TopoBARTMAP and TopoART, and of BARTMAP and TopoART are significantly different at 0.01 significance on all the data sets. However, the performance difference between TopoBARTMAP and BARTMAP varied with the data set under consideration. For data sets generated from Alizadeh-V1, Armstrong-V1, and Armstrong-V2, the performance differed significantly at 0.05 significance. For data sets generated from Alizadeh2000v1, Alizadeh2000v3, and Shipp2002v1 the performances are significantly different at 0.075, 0.10, and 0.15 significance levels, respectively. For the randomized feature order experiments, when run with GA, the results remained close to the original order experiments. TopoBARTMAP out-performs BARTMAP in all these experiments.

6. CONCLUSION

Our goal was to integrate BARTMAP with topological data analysis to improve the effectiveness of biclustering and robustness to noise, and to uncover complex relationships present in the gene expression data sets. We experimentally demonstrated that combining

TopoART with BARTMAP, which we call TopoBARTMAP, results in an algorithm that outperforms BARTMAP. TopoBARTMAP is more robust to order of presentation than either TopoART or BARTMAP. It is also more robust to noise and finds arbitrarily shaped biclusters.

The experiments on parameter sensitivity demonstrate that optimal TopoBARTMAP hyperparameters for one presentation order might not work for different presentation orders. Further, this behavior raises a question about the possibility of TopoBARTMAP being subject to overfitting, which requires further investigation. Due to the rigidness concerning parameters and its higher computational cost than BARTMAP, we recommend using Topo-BARTMAP when data are known to have complex structures that are otherwise not easily identifiable. In any other case, we recommend using BARTMAP.

Table 2. Biclustering results on original data sets with performance measured using Adjusted Rand Index. Best performances are reported in bold.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	Fuzzy ART
1	Alizadeh-V1	**0.4407**	0.4292	0.1899	0.0922
2	Alizadeh-V2	**1.0000**	0.8952	0.9186	0.8147
3	Alizadeh-V3	**0.5521**	0.5079	0.4852	0.4055
4	Armstrong-V1	**1.0000**	0.5565	**1.0000**	0.3379
5	Armstrong-V2	**1.0000**	0.8270	0.9583	0.6281
6	Shipp-V1	**1.0000**	0.9443	**1.0000**	0.1583

Table 3. Biclustering results on randomized order of observations data sets run with parameters fixed to optimal parameters found for original data set. Mean Adjusted Rand Index is used as metric for comparison. Best performances are reported in bold.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	FuzzyART
1	Alizadeh-V1	0.0091 ± 0.0662	0.0119 ± 0.0099	**0.0132 ± 0.0125**	-0.0056 ± 0.0003
2	Alizadeh-V2	**0.4327 ± 0.3773**	0.0419 ± 0.0304	0.0183 ± 0.0084	-0.0025 ± 0.0005
3	Alizadeh-V3	**0.2576 ± 0.1501**	0.1237 ± 0.0871	0.0220 ± 0.0146	-0.0079 ± 0.0000
4	Armstrong-V1	0.0284 ± 0.0536	**0.0455 ± 0.0579**	0.0075 ± 0.0073	0.0153 ± 0.0006
5	Armstrong-V2	0.0015 ± 0.0034	**0.1117 ± 0.0485**	0.0093 ± 0.01401	0.0135 ± 0.0005
6	Shipp-V1	0.0116 ± 0.0260	0.0259 ± 0.0195	**0.0277 ± 0.0150**	0.0066 ± 0.0002

Table 4. Biclustering results on randomized order of features data sets run with parameters fixed to optimal parameters found for original data set. Mean Adjusted Rand Index is used as metric for comparison. Best performances are reported in bold.

Sl.no	DataSet	TopoBARTMAP	BARTMAP
1	Alizadeh-V1	0.0179 ± 0.0234	**0.0666 ± 0.0641**
2	Alizadeh-V2	0.3407 ± 0.4660	**0.8952 ± 0.0000**
3	Alizadeh-V3	0.2747 ± 0.1987	**0.2927 ± 0.0440**
4	Armstrong-V1	**0.3989 ± 0.3215**	0.2881 ± 0.1632
5	Armstrong-V2	**1.0000 ± 0.0000**	0.8152 ± 0.0000
6	Shipp-V1	0.6312 ± 0.5079	**0.9443 ± 0.0000**

Table 5. Biclustering results on randomized order of observation data sets run with GA optimization for TopoBARTMAP, BARTMAP, and TopoART and variation of ρ for Fuzzy ART. Adjusted Rand Index is used as metric for comparison. Best performances are reported in bold.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART	Fuzzy ART
1	Alizadeh-V1	**0.5893 ± 0.1120**	0.4485 ± 0.0853	0.1208 ± 0.0114	0.0658 ± 0.0012
2	Alizadeh-V2	**0.9826 ± 0.0237**	0.8836 ± 0.0401	0.3529 ± 0.0689	0.1166 ± 0.0013
3	Alizadeh-V3	**0.6006 ± 0.0754**	0.5708 ± 0.0896	0.1727 ± 0.0250	0.0954 ± 0.0005
4	Armstrong-V1	**0.9068 ± 0.0330**	0.7537 ± 0.0478	0.1567 ± 0.0204	0.1147 ± 0.0022
5	Armstrong-V2	**0.9149 ± 0.0482**	0.8142 ± 0.0776	0.1992 ± 0.0191	0.0990 ± 0.0021
6	Shipp-V1	**0.5297 ± 0.0902**	0.4116 ± 0.0660	0.2875 ± 0.0538	0.0328 ± 0.0000

Table 6. Biclustering results on randomized order of feature data sets run with GAs. Performance measured using Adjusted Rand Index. Best performances are reported in bold.

Sl.no	DataSet	TopoBARTMAP	BARTMAP
1	Alizadeh-V1	**0.5580 ± 0.1023**	0.5210 ± 0.0646
2	Alizadeh-V2	**0.9799 ± 0.0000**	0.8952 ± 0.0000
3	Alizadeh-V3	**0.5890 ± 0.0501**	0.5630 ± 0.0700
4	Armstrong-V1	**1.0000 ± 0.0000**	0.5640 ± 0.0364
5	Armstrong-V2	**0.9840 ± 0.0219**	0.8380 ± 0.0061
6	Shipp-V1	**1.0000 ± 0.0000**	0.9550 ± 0.0249

ACKNOWLEDGMENT

This research was sponsored by the Missouri University of Science and Technology Mary K. Finley Endowment and Intelligent Systems Center; the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance code BEX 13494/13-9; the Army Research Laboratory (ARL) and the Lifelong Learning Machines program from the DARPA / Microsystems Technology Office, and it was accomplished under Cooperative Agreement Number W911NF-18-2-0260. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES

[1] Cheng, Y. and Church, G. M., 'Biclustering of expression data.' in 'Ismb,' volume 8, 2000 pp. 93–103.

[2] Govaert, G. and Nadif, M., *Co-clustering: models, algorithms and applications*, John Wiley & Sons, 2013.

[3] Madeira, S. C. and Oliveira, A. L., 'Biclustering algorithms for biological data analysis: a survey,' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2004, **1**(1), pp. 24–45.

[4] Padilha, V. A. and Campello, R. J., 'A systematic comparative evaluation of biclustering techniques,' BMC bioinformatics, 2017, **18**(1), p. 55.

[5] Parsons, L., Haque, E., and Liu, H., 'Subspace clustering for high dimensional data: a review,' Acm Sigkdd Explorations Newsletter, 2004, **6**(1), pp. 90–105.

[6] Wang, H., Wang, W., Yang, J., and Yu, P. S., 'Clustering by pattern similarity in large data sets,' in 'Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data,' SIGMOD '02, Association for Computing Machinery, New York, NY, USA, ISBN 1581134975, 2002 p. 394–405.

[7] Chen, J., Ji, L., Hsu, W., Tan, K.-L., and Rhee, S. Y., 'Exploiting domain knowledge to improve biological significance of biclusters with key missing genes,' in '2009 IEEE 25th International Conference on Data Engineering,' IEEE, 2009 pp. 1219–1222.

[8] Patania, A., Vaccarino, F., and Petri, G., 'Topological analysis of data,' EPJ Data Science, 2017, **6**(1), p. 7.

[9] Carlsson, G., 'Topology and data,' Bulletin of the American Mathematical Society, 2009, **46**(2), pp. 255–308.

[10] Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G., 'Extracting insights from the shape of complex data using topology,' Scientific reports, 2013, **3**, p. 1236.

[11] Lockwood, S. and Krishnamoorthy, B., 'Topological features in cancer gene expression data,' in 'Pacific Symposium on Biocomputing Co-Chairs,' World Scientific, 2014 pp. 108–119.

[12] Singh, G., Mémoli, F., and Carlsson, G. E., 'Topological methods for the analysis of high dimensional data sets and 3d object recognition.' in 'SPBG,' 2007 pp. 91–100.

[13] Nicolau, M., Levine, A. J., and Carlsson, G., 'Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,' Proceedings of the National Academy of Sciences, 2011, **108**(17), pp. 7265–7270.

[14] Perkins, M. and Daniels, K., 'Visualizing dynamic gene interactions to reverse engineer gene regulatory networks using topological data analysis,' in '2017 21st International Conference Information Visualisation (IV),' ISSN 2375-0138, July 2017 pp. 384–389.

[15] Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., and Rabadan, R., 'Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development,' Nature biotechnology, 2017, **35**(6), p. 551.

[16] Ma, X., Gao, L., and Tan, K., 'Modeling disease progression using dynamics of pathway connectivity,' Bioinformatics, 04 2014, **30**(16), pp. 2343–2350, ISSN 1367-4803.

[17] Xu, R. and Wunsch, D. C., 'Clustering algorithms in biomedical research: A review,' IEEE Reviews in Biomedical Engineering, 2010, **3**, pp. 120–154, ISSN 1941-1189.

[18] Xu, R. and Wunsch II, D. C., 'BARTMAP: A viable structure for biclustering,' Neural Networks, 2011, **24**(7), pp. 709–716.

[19] Tscherepanow, M., 'TopoART: A topology learning hierarchical ART network,' in 'International Conference on Artificial Neural Networks,' Springer, 2010 pp. 157–167.

[20] Chaibi, A., Lebbah, M., and Azzag, H., 'A new bi-clustering approach using topological maps,' in 'The 2013 International Joint Conference on Neural Networks (IJCNN),' IEEE, 2013 pp. 1–7.

[21] Sarazin, T., Lebbah, M., Azzag, H., and Chaibi, A., 'Feature group weighting and topological biclustering,' in 'International Conference on Neural Information Processing,' Springer, 2014 pp. 369–376.

[22] Carpenter, G. A. and Grossberg, S., 'A massively parallel architecture for a self-organizing neural pattern recognition machine,' Computer vision, graphics, and image processing, 1987, **37**(1), pp. 54–115.

[23] Kim, S., *Novel approaches to clustering , biclustering algorithms based on adaptive resonance theory and intelligent control*, Ph.D. thesis, Missouri University of Science and Technology, 2016.

[24] Elnabarawy, I., Wunsch, D. C., and Abdelbar, A. M., 'Biclustering ARTMAP collaborative filtering recommender system,' in '2016 International Joint Conference on Neural Networks (IJCNN),' ISSN 2161-4407, July 2016 pp. 2986–2991.

[25] Carpenter, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system,' Neural Networks, 1991, **4**(6), pp. 759 – 771, ISSN 0893-6080.

[26] Carpetner, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns,' Technical report, Boston University Center for Adaptive Systems and Department of Cognitive . . . , 1991.

[27] Carpenter, G. A. and Grossberg, S., *Adaptive resonance theory*, Springer, 2010.

[28] Grossberg, S., 'Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world,' Neural Networks, 2013, **37**, pp. 1–47.

[29] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'A survey of adaptive resonance theory neural network models for engineering applications,' Neural Networks, 2019, **120**, pp. 167 – 203.

[30] Wunsch II, D. C., 'Admiring the Great Mountain: A Celebration Special Issue in Honor of Stephen Grossberg's 80th Birthday,' Neural Networks, 2019, **120**, pp. 1 – 4, special Issue in Honor of the 80th Birthday of Stephen Grossberg.

[31] Xu, R. and Wunsch, D., *Clustering*, volume 10, John Wiley & Sons, 2008.

[32] Carpenter, G. A., Grossberg, S., and Reynolds, J. H., 'Artmap: supervised real-time learning and classification of nonstationary data by a self-organizing neural network,' in '[1991 Proceedings] IEEE Conference on Neural Networks for Ocean Engineering,' 1991 pp. 341–342.

[33] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B., 'Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps,' IEEE Transactions on Neural Networks, Sep. 1992, **3**(5), pp. 698–713, ISSN 1941-0093.

[34] de Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermir, T. B., and Schliep, A., 'Clustering cancer gene expression data: a comparative study,' BMC bioinformatics, 2008, **9**(1), p. 497.

[35] Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M., 'Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling,' Nature, Feb 2000, **403**(6769), pp. 503–511, ISSN 1476-4687, doi:10.1038/35000501.

[36] Wang, K., Wang, B., and Peng, L., 'CVAP: validation for cluster analyses,' Data Science Journal, 2009, pp. 0904220071–0904220071.

[37] Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J., 'Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia,' Nature Genetics, Jan 2002, **30**(1), pp. 41–47, ISSN 1546-1718, doi:10.1038/ng765.

[38] Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. W., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C., and Golub, T. R., 'Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning,' Nature Medicine, Jan 2002, **8**(1), pp. 68–74, ISSN 1546-170X, doi:10.1038/nm0102-68.

[39] Eiben, A. E. and Smith, J. E., *Introduction to Evolutionary Computing*, Springer Publishing Company, Incorporated, 2nd edition, 2015.

[40] Hubert, L. and Arabie, P., 'Comparing partitions,' J. Classification, 1985, **2**(1), pp. 193–218.

[41] Brito da Silva, L. E. and Wunsch II, D. C., 'A study on exploiting VAT to mitigate ordering effects in Fuzzy ART,' in 'Proc. IEEE International Joint Conference on Neural Networks (IJCNN),' 2018 pp. 2351–2358.

[42] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'Dual vigilance fuzzy adaptive resonance theory,' Neural Networks, 2019, **109**, pp. 1 – 5.

[43] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence,' Neural Networks, 2020, **121**, pp. 208 – 228.

II. TOPOBARTMAP: TOPOLOGICAL BICLUSTERING ARTMAP

Raghu Yelugam[1], Leonardo Enzo Brito da Silva[1,2], Donald C. Wunsch II[1]

[1] Applied Computational Intelligence Laboratory,
Missouri University of Science and Technology, Rolla, MO, USA

[2] CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil

Email: ry222@umsystem.edu, {leonardoenzo, wunsch}@ieee.org

ABSTRACT

Biclustering is a powerful tool for exploratory data analysis such as social networking, data reduction, and differential gene expression studies. To improve the quality of biclustering and module extraction, a biclustering method, BARTMAP, was combined with a topological clustering algorithm, TopoART, to produce TopoBARTMAP. Topological clustering identifies connected regions that are difficult to find using other popular methods, and it produces a graphical representation. TopoBARTMAP inherits the ability to detect topological associations while performing data reduction. The capabilities of TopoBARTMAP are benchmarked to 35 cancer datasets and contrasted with other clustering methods. Assessments are made using the Adjusted Rand Index on ordered and shuffled data. TopoBARTMAP provided a statistically significant improvement over the other assessed methods. The produced graphical representation is refined to represent gene bicluster associations and is assessed on the NCBI GSE89116 dataset containing expression levels of 39,326 genes sampled over 38 observations.

Keywords: Adative Resonance Theory (ART), Biclustering, Topological Data Analysis, Gene Expression, and Co-expression

1. INTRODUCTION

With high throughput gene expression profiling thousands of genes simultaneously, the study of functional interactions among genes had led to gaining insights into the cell cycle and diseases with perturbations in environmental state [1]. Such studies necessitated the development of computational methods that could identify and associate genes to specific diseases, regulatory processes and uncovering regulatory networks, and disease progression and prognosis. Biclustering[2, 3, 4], one such method, has been widely used to identify genes corresponding to cancers.

Given the gene expression data, the bi-clustering algorithms identify the genes that co-express highly across a subset of all conditions and often uses non-time-series data. Co-expression identified by these methods provides insights into local associations within gene regulatory networks and are capable of identifying active functional modules [5, 6, 7, 8]. Nonetheless, such local information is insufficient to completely uncover the regulatory network without further analysis and biclustering methods are shown to perform relatively poor at module extraction [8]. Moreover, biclustering fails to explain the nature of interactions between genes within a bicluster, although identified genes can be functionally associated with other known genes through "Guilt By Association" principle. Furthermore, biclustering methods are prone to miss grouping genes related by a regulatory process if they do not co-express significantly and if the bi-clusters are of complex and nontrivial shape. Because topological methods are able to uncover complex-shaped clusters and underlying geometry, a combination of topological methods with biclustering might enable recovery of complex shaped biclusters.

Topological methods applied to high dimensional data, Topological Data Analysis (TDA)[9, 10, 11, 12, 13, 14], builds on clustering by identifying connected spaces and produces a graphical representation of the geometric structure, often as simplicial complexes[15], which renders topological methods with noise, shift, and scale invariance. These properties of interest and graphical representation had led to the application of topological methods

like Mapper [16] towards uncovering the disease progression [17, 18], identifying subgroups of cancer [19], cellular differentiation [20], and genome dynamics and reverse engineering gene regulatory networks [21]. In order to enhance biclustering with these properties and take a step towards using TDA with biclustering, we developed Topological Biclustering ARTMAP, TopoBARTMAP [22], which is a hybrid of BARTMAP [23] and TopoART [24], the latter is a topology learning neural clustering method.

In this article, we present a variation of TopoBARTMAP that uses the Maximal Information Coefficient (MIC) [25] as a measure for correlation between samples across feature clusters. We extend the results of our recent work with TopoBARTMAP [22] by testing the algorithm on a much broader collection of benchmark data-sets while demonstrating, via statistical analysis, the method's significant performance when compared to other methods. We further detail the performance of the method under randomization and introduce the *topological gene bicluster association network*. The latter summarizes the gene and condition cluster associations within a bicluster, produced by clustering using TopoBARTMAP on a breast cancer data-set [26].

The remainder of this article is organized as following. The pertinent ART models [27] principal for understanding the kernel of the contribution, i.e., TopoBARTMAP, are presented in Section 2. Section 3 presents the TopoBARTMAP model. Details germane for experimental study: data, methods, clustering performance measures, and parameter tuning procedure adopted are explicated in Section 4. Section 5 presents experimental observations and discusses their significance. In closing, Section 6 concludes with the findings of this paper.

2. RELATED WORK

This section presents a compendium on TopoART [24] and BARTMAP [23] thus furnishing principal details for explicating TopoBARTMAP. The reader is referred to [28, 29, 30, 31, 32] for a rigorous exploration of "Winner-Take-All" (WTA) learning models based on Adaptive Resonance Theory (ART) and their applications.

2.1. TOPOLOGICAL ART

Fundamentally, topology learning clustering algorithm Topological ART or TopoART [24] is built on Fuzzy ART. Readers are recommended [33, 31] for brief explanations, and [34, 28] for the original treatise on Fuzzy ART, respectively. TopoART, similar to typical ART-based learning models in construction, contains three layers of neurons, called fields, F0, F1, and F2. The initial layer F0 receives the input signal and transmits scaled and complement coded input to layer F1; i.e., if $\mathbf{x} = [x_1, x_2, ..., x_d]$ represents the scaled input vector, then $\mathbf{X} = [x_1, x_2, ..., x_d, x_1^c, x_2^c, ..., x_d^c]$ with $x_i^c = 1 - x_i$ represents the input to F1. Since the interactions between field F1 and F2 govern learning, the input to TopoART is assumed henceforth in the exposition to be scaled and complement coded, reducing the network to contain interacting fields F1 and F2.

The field F2 consists of several prototypes, with j^{th} prototype's weights represented by the vector $\mathbf{W_j} = [W_1^j, W_2^j, \cdots, W_{2d}^j]$ which are used for categorizing the input pattern. When an input is presented, due to bottom up activity at the layer F1, each prototype in F2 gets activated. The activation of each prototype can be computed with the choice function:

$$T_j = \frac{|\mathbf{X} \wedge \mathbf{W_j}|}{\alpha + |\mathbf{W_j}|}, \tag{1}$$

where $|\cdot|$ is the ℓ_1 norm, $\wedge$ is the fuzzy AND operation defined by $(X \wedge W_j)_i = min(x_i, W_i^j)$, $\mathbf{X} \in \Re^{2d}, \mathbf{W} \in \Re^{2d}$ with choice parameter $\alpha > 0$. With the activations computed, the best matching prototype or the highly active prototype in the field F2 is selected as the winner. This winner is checked to satisfy the vigilance criterion using the category match function:

$$m_j = \frac{|\mathbf{X} \wedge \mathbf{W_j}|}{|\mathbf{X}|}. \tag{2}$$

If the match m_j is above the vigilance threshold, ρ, the vigilance criterion is satisfied, and this causes resonance between the winner prototype and the field F1. Resonance leads to learning; the weight vector of the winner is moved in the direction to encompass the presented input:

$$\mathbf{W_j} = \beta(\mathbf{W_j} \wedge \mathbf{X}) + (1 - \beta)\mathbf{W_j} \tag{3}$$

where β is the learning rate. The output of the layer F2 is set as following:

$$y_i = \begin{cases} 1, & \text{if i = j} \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

If the vigilance criterion is not met, the winner neuron's activity is suppressed, and the next winning neuron is searched. If no such winner neuron is found, a new prototype is assigned in layer F2 to represent the input, and its weights are updated with β set to 1 in Eq. (3) to enable fast learning. Each neuron learned this way represents a hyper-cube that covers the summarized data points in the space from which they originate. Further, the size of the hyper-cube represented by i^{th} prototype, S_i, at any instance of learning is given by:

$$S_i = \sum_{k=1}^{d} \mid (1 - \mathbf{W^i_{d+k}}) - \mathbf{W^i_k} \mid . \tag{5}$$

Along with the winning prototype, a second highly active prototype satisfying the vigilance criterion, called second best, is selected. The weights of this second-best prototype are updated with a lower learning rate $\beta_{second-best} : \beta_{second-best} < \beta_{best}$ using the same update rule (i.e., Eq. (3)). An edge is established between these two prototypes indicating an intersection between the corresponding hyper-cubes. While an update to the weights of the second-best prototype moves it closer to the presented input pattern promoting growth in relevant regions of input space, the edges represent the connectivity of the space and help in uncovering geometric structures of interest within the data.

Further, each neuron in field F2 has a counter registering the number of samples summarized by the prototype. Once in every τ time steps, prototypes that do not summarize more than a predefined number of samples, ϕ, are removed from F2 layer to make the method less susceptible to noise. While the prototypes that are retained after this pruning mechanism are called permanent prototypes, those created between the two consecutive pruning cycles are called candidate prototypes. Further, only edges formed between the permanent prototypes are retained.

The retained edges are labeled depending on the connectivity between the prototypes; i.e., edges between prototypes belonging to the same cluster are assigned with the same label. These labels along with the vertices corresponding to labeled edges are recorded in a vector - $\overrightarrow{C} = \{< c_j >: j = 1, ..., l\}$, where l is the number of labels and c_j is the list consisting of all the prototypes that are connected by corresponding edges with label l. Each member list c_j of the vector $\overrightarrow{C}$ represents a complex-shaped cluster of the data. Further, each of the member list c_j can be used to construct the nerve of the cover formed by the prototypes. Since the intersection between any two prototypes is either empty or contractible (a hyper-cube), the nerve is homotopy equivalent[2] of the topological subspace from which the data originates.

After complete data presentation, when no further learning is desired, data associated with the pruned neurons can be reclassified by assigning them to permanent prototypes. The assignment is decided using WTA applied to the activation defined as the following [24]:

$$T_j = 1 - \frac{|(\mathbf{X} \wedge \mathbf{W_j}) - \mathbf{W_j}|}{|\mathbf{X}|}. \tag{6}$$

TopoART, further, refines the identified clusters by cascading modules of Fuzzy ART that receive input patterns that activate a permanent prototype in the preceding module. Each such module has a vigilance value higher than that of its preceding module. In addition to further improving noise robustness, the cascading produces a hierarchical effect. It is customary to calculate the vigilance value of subsequent modules using that of the lowest

[2]Due to Nerve Theorem [15].

level module, with the rule:

$$\rho_{next} = \frac{1}{2}(1 + \rho_{previous}), \tag{7}$$

where ρ_{next} and $\rho_{previous}$ represent vigilance values of next and previous modules respectively.

2.2. BICLUSTERING ARTMAP

Biclustering ARTMAP, or BARTMAP [23], is a Fuzzy ARTMAP-based [35, 31, 36] two-way clustering method that can adaptively identify biclusters present in the data without any prerequisite knowledge on the number of biclusters. BARTMAP was engineered by modifying the map-field of Fuzzy ARTMAP. This fundamental design change converts the latter from a supervised to an unsupervised learning method. In this context, the family of ARTMAP-based models that also accomplish such conversion includes the self-consistent modular ART (SMART) [37] and iCVI-ARTMAP [38], which perform hierarchical and cluster validity index-based clustering, respectively.

The notation adopted for the exposition on BARTMAP in this section is as follows: the presented data, related to genomics for instance, is denoted as a matrix $\mathbf{G} = (F, O)$. With $F = \{f_1, \cdots, f_N\}$ denoting the feature set of cardinality N, $O = \{o_1, \cdots, o_M\}$ denotes the set of M experimental conditions or observation during which the features F are observed. For a given experimental observation o_j, the perceived intensity of feature f_i is represented by the element $g_{f_i o_j} \in \mathbf{G}$ (or more succinctly, g_{ij}). Henceforward, for generality, the genes are considered as features and samples are considered as experimental observations in the article.

BARTMAP retains Fuzzy ARTMAP's modular construction, thereby consisting of two Fuzzy ART modules, ARTa and ARTb, interacting via the Inter-ART module, where the latter is reduced to measure the similarity. Unlike Fuzzy ARTMAP, however, the training procedure for BARTMAP progresses in two steps. The first step of training involves grouping features that exhibit similar intensity levels for all considered experimental observations of

the data matrix $\mathbf{G} = \left[g_{ij}\right]_{M\times N}$. For genomic data, such grouping fundamentally identifies genes that are likely involved in the same regulatory process of a biological pathway [39]. The ARTb module, which is a standard Fuzzy ART, is responsible for partitioning F into k_f clusters $\{F_i \mid i = 1, \cdots, k_f\}$.

During the second step, the experimental observations/conditions set O is partitioned with the ARTa module. For a presented observation, the activity of each prototype in ARTa module is computed with Eq. (1). A representative cluster is identified using WTA and is tested to meet the vigilance criterion (Eq. (2)). If a significant match is noticed, the observation is associated with the recognized winner if and only if it demonstrates a similar intensity pattern across at least one of the feature clusters as the other observations summarized by the winning prototype. While there are several similarity measures available in the literature [40], BARTMAP uses linear Pearson Correlation Coefficient [41] as the litmus test for the similarity; which was found to perform better than several other proximity measures when bicluster classes are known [42]. This homogeneous behavior is substantiated by checking if the average correlation measured between the presented observation and other represented observations η_{kj} is higher than a presupposed correlation threshold, η_{th}. The former is defined as:

$$\eta_{kj} = \frac{1}{M_j}\sum_{l=1}^{M_j} r_{kjl}, \tag{8}$$

where

$$r_{kjl} = \frac{\sum_{t=1}^{N_i}(g_{o_k f_{it}} - \overline{g}_{o_k F_i})(g_{o_{jl} f_{it}} - \overline{g}_{o_{jl} F_i})}{\sqrt{\sum_{t=1}^{N_i}(g_{o_k f_{it}} - \overline{g}_{o_k F_i})^2}\sqrt{\sum_{t=1}^{N_i}(g_{o_{jl} f_{it}} - \overline{g}_{o_{jl} F_i})^2}}, \tag{9}$$

$$\overline{g}_{o_k F_i} = \frac{1}{N_i}\sum_{t=1}^{N_i} g_{o_k f_{it}}, \tag{10}$$

$$\overline{g}_{o_{jl} F_i} = \frac{1}{N_i}\sum_{t=1}^{N_i} g_{o_{jl} f_{it}}, \tag{11}$$

and o_k is the presented observation, o_{jl} are members of the observation cluster $O_j = \{o_{j1}, \cdots, o_{jM_j}\}$ with cardinality M_j associated with the winning prototype, and $F_i = \{f_{i1}, \cdots, f_{iN_i}\}$ denotes a feature cluster with N_i members.

The winning prototype O_j qualifies to represent observation o_k only if η_{kj} is above η_{th}. If this were the case, then weights of the prototype O_j are updated using the learning rule expressed by Eq. (3). Nonetheless, if it were not the case, the current winning prototype is suppressed from activating by match tracking [35], which adequately increases the vigilance value ρ_a of ARTa above its base value. The increment in vigilance value forces for finding next qualifying winning prototype. The match tracking terminates upon identifying a winner prototype that qualifies for representing the current observation or the vigilance value is increased to 1. If match tracking increases vigilance value to its maximum, a new prototype is assigned to represent the current observation. The weights of this prototype are updated using fast learning by setting learning rate β in Eq. (3) to 1. After the termination of match tracking and before next observation presentation, the vigilance value ρ_a is readjusted to its base value.

Despite scaling, features that demonstrate co-variation for a subset of conditions can be grouped into different clusters by the ARTb module due to varying intensity levels. Due to the condition imposed for the association, nonetheless, an observation might highly correlate with other observations associated with the representative prototype across more than one feature cluster. Thus during the training, BARTMAP learns a mapping between feature and observation partitions. Upon training, the clusters identified by the ARTa module are considered for labeling the observations.

For recovering significant biclusters, from all possible combinations of clusters identified by ARTa and ARTb, homogeneity of the bicluster is used as a criterion. The quantitative assessment for homogeneity is made using the average correlation computed between the observations. For a bicluster $\mathbf{B}_{M_B \times N_B} = (F_B, O_B) = ([f_1, \cdots, f_{M_B}], [o_1, \cdots, o_{N_B}])$

composed of N_B observations and M_B features, the average correlation is defined as [43]:

$$\delta(\mathbf{B}_{M_B \times N_B}) = \frac{1}{\binom{N_B}{2}} \sum_{p=1}^{N_B-1} \sum_{q=p+1}^{N_B} r_{o_p,o_q}, \tag{12}$$

where r_{o_p,o_q} is the Pearson correlation coefficient between observations $o_p, o_q \in O_B$ across F_B. A feature cluster is mapped to an observation cluster if and only if $\delta(\mathbf{B}_{M_B \times N_B})$ is above recovery correlation threshold, η_r. With each feature cluster several observation clusters might form biclusters that satisfy the aforesaid condition. However, each feature cluster is mapped only to the observation cluster that results in a bicluster with highest $\delta(\mathbf{B}_{M_B \times N_B})$. Therefore, BARTMAP identifies block-diagonal structure in the provided data. BARTMAP had been extended to perform hierarchical biclustering and supervised learning[43], respectively. Hierarchical BARTMAP variant is furnished with the ability to evaluate and pair the feature and observation clusters to form biclusters.

3. TOPOLOGICAL BICLUSTERING ARTMAP

Topological Biclustering ARTMAP (TopoBARTMAP) is a modified BARTMAP wherein TopoART modules supplant fuzzy ART modules to identify the topological associations within the presented expression data matrix, $\mathbf{G} = (F, O)$. TopoBARTMAP inherits the modular structure and two-phase learning process from BARTMAP; nonetheless, each module's learning procedure differs due to the usage of TopoART. Learning procedure results in the identification of gene and sample subspace topologies while associating them through identified biclusters. While usage of TopoART renders the method noise-robust, it also produces a graphical representation of the underlying space. In what follows next, the training procedure is delineated.

The first phase of training involves grouping features, or genes, that exhibit similar intensities over all experimental observations of a given expression data matrix $\mathbf{G}$, which is achieved by clustering features with TopoARTb module. Clustering results in the

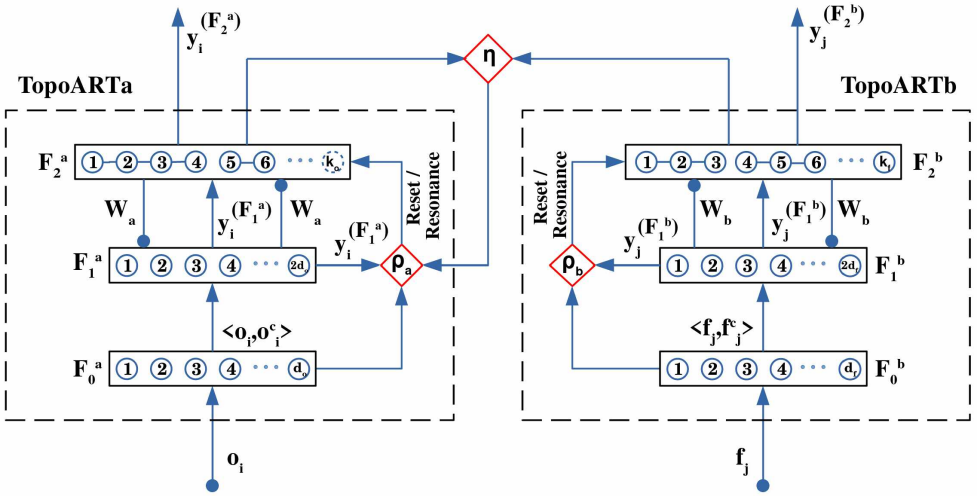


Figure 1. Schematic of TopoBARTMAP [22]. In construction, TopoBARTMAP contains two TopoART modules connected by the inter-ART module that relates clusters found by each module through the similarity test. The schematic shows the k_f prototypes identified by the TopoARTb module in layer $\mathbf{F}_2^b$. For TopoARTa, the layer $\mathbf{F}_2^b$ is shown to have k_o learned prototypes which include both permanent and temporary prototypes (indicated by a dashed circle). Connected nodes in the $\mathbf{F}_2$ layers of both modules indicate identified clusters/topologies.

identification of k_f prototypes that form k clusters. All the feature patterns that are associated with pruned prototypes are reclassified and associated with permanent prototypes using Eq. (6).

In the later training step, the TopoARTa module is trained on the observations O. During an observation presentation, the identified winner prototype is tested for match and associability using the vigilance test and similarity test, respectively. For similarity test, similarity can be measured using the Pearson Correlation coefficient [44] when the strength of linear association is desired to be measured. Nonetheless, when non-linear associations are desired to be identified, the similarity can be measured using Maximal Information Coefficient (MIC)[25], an information-theoretic-based measure. MIC is based on the premise that if the two variables are related, then the scatter plot or the ordered pairs of the two variables can be partitioned such that the relationship is preserved and summarized by the partitions.

To calculate the MIC between two variables $\mathbf{X} =< x_1, x_2, \cdots >$ and $\mathbf{Y} =< y_1, y_2, \cdots >$ of same size, the ordered pair $\mathbf{D} \mid \mathbf{D} = \{\{x_i, y_i\} \forall i \in \{1, 2, 3, \cdots\}\}$ is distributed into several bins using grids, with varying number of rows and columns, that divide the scatter plot. For n rows and m columns, several grids G are possible. For each of such combination of n and m, the maximum Mutual Information $I^*(\mathbf{D}, n, m)$ over all grids is found where

$$I^*(\mathbf{D}, n, m) = \max_G \left[I(\mathbf{D} \mid_G) \right], \tag{13}$$

here $I(\mathbf{D} \mid_G)$ denotes the mutual information computed for the grid G. With I^* computed for all combinations of n and m, a characteristic matrix is then defined with entries

$$M(\mathbf{D})_{n,m} = \frac{I^*(\mathbf{D}, n, m)}{log(min\{n, m\})}, \tag{14}$$

the division with $log(min\{n.m\})$ scales the values to between 0 and 1 as for any grid G, $0 \leq I(\mathbf{D} \mid_G) \leq 1$. Using the characteristic matrix, MIC of two variable $\mathbf{X}$ and $\mathbf{Y}$ with samples size s and grid size less than $B(s)$ is determined by

$$MIC = \max_{nm<B(s)} \left[M(\mathbf{D})_{n,m} \right], \tag{15}$$

where $B(s)$ defines the upper bound for the grid size [25].

If the input observation demonstrates the same behavior as the other observations summarized by the winner across at least one identified feature cluster, then the input is assigned to the winner. However, if significant similarity can not be identified, match tracking suppresses the current winner. Like BARTMAP, the match tracking resultant search proceeds to find an associable prototype until all the prototypes are suppressed, and a new cluster is formed. If the winner is an existing prototype, then a second-best prototype is recognized using WTA and tested for the match. If the considered second-best prototype meets the vigilance criterion, then the similarity test is conducted to verify the

observation's associability. If a significant similarity between the input observation and observations summarized by the second winner across at least one feature cluster is detected, then an edge between the first and the second winners is established. If the similarity test detects no significant similarity, then the current second-best is suppressed and the next winner is searched. The search continues until either a prototype qualifying both vigilance and similarity test is identified or all the prototypes are shutdown. Upon identification of qualifying second-best prototype, the weights are updated using Eq. (3) with a lower learning rate. The pruning mechanism and reclassification of the corresponding observations are unchanged from TopoART.

Unlike BARTMAP, where each observation prototype represents a bicluster; in case of TopoBARTMAP, the complex-shaped clusters formed by connected observation prototypes (each forming a topological subspace[3]) represent biclusters. Such representation aids in uncovering the intra-bicluster interactions. Further, if the input to TopoBARTMAP is an expression data, clustering genes with TopoARTb module generates a graph analogous to Gene Coexpression Networks (GCN) [45]. The nodes in the generated graph represent gene modules rather than individual genes, and links represent the interactions between these modules. Moreover, the learning in TopoARTa module associates these nodes to specific subsets of experimental conditions, facilitating further analysis of misregulation in genes; therefore, the generated network is called as *Topological Gene Bicluster Association Network*.

The similarity condition imposed for recognizing the second-best prototype during TopoARTa learning leads to the identification of gene prototypes for which the observations represented by both best and second-best prototypes show high similarity. However, each prototype's represented observation might demonstrate homogeneity for more than one of

[3]In topological space defined on $[0, 1]^{2d}$, from which the scaled complimented coded input data originates

the k_f gene prototypes. When applied to time series gene expression data, four relationships emerge between the gene clusters (call F_b and F_{sb}) associated with best and second best observation prototypes that represent an observation. And they are:

- $F_b = F_{sb}$: This indicates the continuity of co-expression through the conditions.
- $F_b \subset F_{sb}$: This indicates a possible down regulation and inhibition of the genes that are in the set $\{F_{sb} \setminus F_b\}$ while the genes in the set F_b continue to be up regulated.
- $F_b \supset F_{sb}$: This indicates possible triggered up regulation of the genes that are in the set $\{F_b \setminus F_{sb}\}$ due to the continued up regulation of genes in the set F_b.
- $F_b \neq F_{sb}$: This indicates the end of the upregulation of the genes in the set $F_b \setminus (F_b \cap F_{sb})$ and start of the up regulation of the $F_{sb} \setminus (F_b \cap F_{sb})$ in the overlap region.

4. EXPERIMENTATION

4.1. DATA

This work examines and compares the clustering performance of TopoBARTMAP with other methods on an extensive collection of thirty-five benchmark cancer data-sets [46] used in several other biclustering studies. Table 1 summarizes the pertinent information on the number of features, observations, and classes per data-set.

To study and understand the relevance of the topological graphs produced by TopoBARTMAP, TopoBARTMAP was run on the data-set made available by [26] on National Center for Biotechnology Information (NCBI) Gene Expression Omnibus with the accession number: GSE89116. The data-set consists of unnormalized expression levels of 39,326 genes collected for 38 tissue samples (observations) along with *p-values*. The samples consist of tissue collected from 12 Early-onset breast tumors (with age $\leq$ 40 yrs), 17

Late-onset breast tumors (with age $\geq$ 55 yrs), and adjacent normal tissues from 4 Early-onset and 5 Late-onset breast cancer patients as controls. The raw gene expression values without any normalization[47] were used to experiment. The information pertinent to tissue type was used to group the observations into four categories, and this partitioning is used to measure the clustering performance.

4.2. CLUSTERING PERFORMANCE

In all the experiments conducted, clustering is done to achieve hard partitioning; i.e., each sample is assigned to a single cluster. Cluster validity indices (CVIs) are used as a proxy for the quality of data partitions and can be categorized as internal or external. Internal CVIs are used in the absence of labels, which is the case of true unsupervised learning scenarios. While these CVIs have been traditionally computed offline [33], online versions have been recently developed in [73, 74, 75]. Conversely, external CVIs are typically used when a reference partition is available, which is the case when benchmarking clustering algorithms. Therefore, in our experiments, the clustering performance is measured using the Adjusted Rand Index [76], an external CVI, which is defined as the following for two partitions **P** and **Q** of a set $\mathbf{D} = \{d_1, d_2, \cdots d_N\}$ with N data points:

$$ARI = \frac{\binom{N}{2}(p+q) - ((p+r)(p+s) + (q+r)(q+s))}{\binom{N}{2}^2 - ((p+r)(p+s) + (q+r)(q+s))}, \tag{16}$$

where

- p denotes total number of pairs of elements in **D** that belong to the same partition in both **P** and **Q**
- q denotes total number of pairs of elements in **D** that are assigned different partitions in both **P** and **Q**

Table 1. Thirty-five benchmark cancer data sets used for experimentation [4]. The Sl.no, N_o, N_f, and N_c denote respectively the serial number, $\#Observations$, $\#Features$, and $\#Classes$ in observations.

Sl.no	Data set	N_o	N_f	N_c
22	Alizadeh-V1 [48]	42	1095	2
23	Alizadeh-V2 [48]	62	2093	3
24	Alizadeh-V3 [48]	62	2093	4
1	Armstrong-V1 [49]	72	1081	2
2	Armstrong-V2 [49]	72	2194	3
3	Bhattacharjee [50]	203	1543	5
25	Bittner [51]	38	2201	2
26	Bredel [52]	50	1739	3
27	Chen [53]	180	85	2
4	Chowdary [54]	104	182	2
5	Dyrskjot [55]	40	1203	3
28	Garber [56]	66	4553	4
6	Golub-V1 [57]	72	1877	2
7	Golub-V2 [57]	72	1877	3
8	Gordon [58]	181	1626	2
29	Khan [59]	83	1069	4
9	Laiho [60]	37	2202	2
30	Lapointe-V1 [61]	69	1625	3
31	Lapoint-V2 [61]	110	2496	4
32	Liang [62]	37	1411	3
10	Nutt-V1 [63]	50	1377	4
11	Nutt-V2 [63]	28	1070	2
12	Nutt-V3 [63]	22	1152	2
13	Pomeroy-V1 [64]	34	857	2
14	Pomeroy-V2 [64]	42	1379	5
15	Ramaswamy [65]	190	1363	14
33	Risinger [66]	42	1771	4
16	Shipp [67]	77	798	2
17	Singh [68]	102	339	2
18	Su [69]	174	1571	10
19	West [70]	49	1198	2
20	Yeoh-V1 [71]	248	2526	2
21	Yeoh-V2 [71]	248	2526	6
34	Tomlins-V1 [72]	104	2315	5
35	Tomlins-V2 [72]	92	1288	4

The data sets are available at `https://github.com/padilha/biclustlib`.

- r denotes total number of pairs of elements in **D** that are in same partition in **P** but different partition in **Q**

- s denotes total number of pairs of elements in **D** that are in different partitions in **P** but same partition in **Q**

4.3. METHODS

To compare and contrast the performance of TopoBARTMAP with other methods, BARTMAP, Fuzzy ART, Spectral biclustering, and TopoART were run on the same thirty-five data-sets. The choice of Spectral clustering is driven by the results reported in [4], where Spectral clustering was shown to perform better than sixteen other state-of-the-art biclustering methods when sample clustering accuracy, measured in terms of FARI [77] and 13 AGRI [78], was used as the performance measure. To run the Spectral biclustering, we used the Scikit-Learn [79] Python library. This study used MATLAB implementations of BARTMAP, Fuzzy ART, TopoART, and TopoBARTMAP, which are made available at the Applied Computational Intelligence Laboratory GitHub repository[4]. In conjunction with the MATLAB source code, the Cluster Validity Analysis Platform toolbox [80] and minepy[5] [81] (MIC tool) were used when needed.

To statistically analyze the differences in performance of multiple methods, comparisons were made using Statistical Comparison of Multiple Algorithms in Multiple Problems (SCMAMP) R-package [82]. In this study, the similarity in performance of two classifiers is ascertained using the Wilcoxon signed-rank test. For multiple methods, Friedman's rank-sum test [83] is used as an evaluation of statistical similarity in behavior. The assessment was made with Iman and Davenport omnibus test run on the observed performances. If the tests

[4]https://github.com/ACIL-Group/TopoBARTMAP
[5]https://minepy.readthedocs.io/en/latest/

resulted in a small *p-value*, an indicative that at least one method performs differently than the rest, then Friedman post-hoc test with Bergmann and Hommel's correction [83] was conducted to quantify the differences between multiple classifiers.

4.4. PARAMETER TUNING

To study the biclustering performance of TopoBARTMAP with Pearson Correlation and MIC on each of the thirty five data-sets, the parameters ρ_a, $\beta^a_{second-best}$, ϕ_a, $\tau^{\%}_a$, ρ_b, $\beta^b_{second-best}$, ϕ_b, $\tau^{\%}_b$, and η were optimized with Genetic Algorithm (GA) [84] for one epoch of data presentation. Here $\tau^{\%}_{\cdot}$ denotes the percentage of aggregate number of patterns input to each TopoART module. During the optimization the following constraints were imposed on the parameters:

$$\phi_a - \tau^{\%}_a N_o \geq 0, \tag{17}$$

and

$$\phi_b - \tau^{\%}_b N_f \geq 0, \tag{18}$$

where $N_o = \#\{observations\}$ and $N_f = \#\{features\}$. Note that $\tau_a = \lfloor \tau^{\%}_a N_o \rceil$ and $\tau_b = \lfloor \tau^{\%}_b N_f \rceil$. Throughout the experiments, the lower and upper bounds for the parameters, following the aforesaid order, were set to $< 0.0, 0.0, 0.0, 0.1, 0.0, 0.0, 0.0, 0.1, 0.0 >$ and $< 0.95, 1.0, 5.0, 0.3, 0.95, 1.0, 5.0, 0.3, 0.99 >$. Because many data-sets become sparse upon normalization, the correlation is computed between two samples over only those genes for which the normalized activity is non-zero [85]. The GA was run 10 times with a population size of 200 agents and for 25 generations. Further for all the ART based methods evaluated, the choice parameter α was set to 0.001, match tracking step size wherever applicable was set to 0.01, and best prototype learning rates were set to 1, thereby ensuring fast learning.

The Spectral biclustering method searches for the checkerboard pattern formed by the specified number of row and column clusters; thus, the number of row clusters and column clusters are considered as parameters for the method. Whilst the parameters

for other methods considered here take continuous values from a specified range, each aforementioned parameter of Spectral biclustering takes discrete values between 1 (resulting in a trivial cluster) to the number of rows or columns in the data; therefore, we did not use GA for optimization. In order to reduce the parameter search space, however, similar to experimentation in [4], the method received the ground truth number of classes, reported in Table 1, as the number of column clusters to be searched. Nonetheless, the input number of row clusters is varied from 2 to min{#rows,100}, where #rows is the number of rows in the dataset. Further, the scikit-learn library allows choosing among three different data normalization methods for Spectral biclustering - Independent row and column normalization, Log normalization, and Bistochastization; we ran experiments with each normalization method while setting other parameters related to Singular Value Decomposition to their default values.

To investigate the performance of BARTMAP, parameters $< \rho_a, \rho_b, \eta >$ were optimized with GA using $< 0.0, 0.0, 0.0 >$ and $< 0.95, 0.95, 0.99 >$ as lower and upper bounds, respectively. Similarly for studying TopoART, parameters $< \rho, \beta_{J_2}, \phi, \tau^{\%} >$ were optimized with GA using $< 0.0, 0.0, 0.0, 0.1 >$ and $< 0.95, 1.0, 5.0, 0.3 >$ as lower and upper bounds, respectively. These were also subjected to the following inequality constraints:

$$\phi - \tau^{\%} N_o \geq 0, \tag{19}$$

where $\tau^{\%}$ and τ are defined as for running TopoBARTMAP experiments. Similar to TopoBARTMAP, for data-sets that become sparse upon normalization, modified Pearson correlation was used.

We note that while TopoART can refine the identified cluster with multiple modules, no such refinement was considered during the experimentation. Further, like TopoBARTMAP, the identified complex shaped clusters formed by connected prototypes were used to measure the performance. Finally, because Fuzzy ART has single parameter which can be varied,

to maintain consistency with the number of fitness evaluations of the GA used to optimize other ART based methods, ρ was varied among 4960 values uniformly spaced from the range $[0, 1]$.

5. RESULTS AND DISCUSSION

This section presents a particularized account on the experimental observations while furnishing the qualitative and quantitative assessments made to evaluate the performance of TopoBARTMAP.

5.1. ORIGINAL DATA ORDERING EXPERIMENT

The Table 2 juxtaposes the performance of TopoBARTMAP with Pearson Correlation Coefficient and MIC, BARTMAP, Spectral Biclustering, TopoART, and Fuzzy ART in the experiments conducted on thirty-five benchmark data-sets. In these experiments, TopoBARTMAP with both Pearson Correlation Coefficient and MIC outperforms the rest four methods on eighteen data-sets. While BARTMAP outperforms other methods on five data-sets, it performs equally with TopoBARTMAP and TopoART on three data-sets. The topological non-bi-clustering method TopoART is observed to perform better than BARTMAP and equally to TopoBARTMAP on five data-sets while outperforming the other four methods on one data-set. Although such equal performance remains unexplained, we ascribe this to topological clustering and order of data presentation. For Spectral Biclustering, for each data-set, the Table 2 reports best-observed ARI value among three normalization procedures. While both Spectral Biclustering and Fuzzy ART are observed to perform poorly, these two methods perform better than other methods on one data-set each.

Due to the apparent similarity in performance between multiple algorithms over multiple data-sets, an assessment was made to check if the methods demonstrate similar behavior. The first test was made to check for the statistical difference in the behavior of both

Table 2. Biclustering results on thirty-five benchmark cancer data-sets. The best found Adjusted Rand Index value for each method is reported against the corresponding data-set. Best performances for each data-set are reported in bold. Columns TBM-P, TBM-MIC, and Spectral refer to the performances reported for TopoBARTMAP with Pearson Correlation, TopoBARTMAP with MIC, and Spectral Biclustering respectively

Sl.no	DataSet	TBM-P	TBM-MIC	BARTMAP	TopoART	Fuzzy ART	Spectral
1	Alizadeh-V1	0.4407	**0.5143**	0.4292	0.1899	0.0922	0.2222
2	Alizadeh-V2	**1.0000**	**1.0000**	0.8952	0.9186	0.8147	0.9471
3	Alizadeh-V3	0.5521	**0.5925**	0.5079	0.4852	0.4055	0.5526
4	Armstrong-V1	**1.0000**	**1.0000**	0.5565	**1.0000**	0.3379	0.2383
5	Armstrong-V2	**1.0000**	**1.0000**	0.8270	0.9583	0.6281	0.5879
6	Bhattacharjee	0.9827	0.9861	**0.9885**	0.9779	0.6449	0.5511
7	Bittner	0.7593	**0.7951**	0.5775	0.5931	0.3284	0.0085
8	Bredel	**0.8225**	0.8043	0.6631	0.2085	0.0569	0.3244
9	Chen	0.6765	0.615	0.4448	0.1998	0.0402	**0.7097**
10	Chowdary	**1.0000**	**1.0000**	0.9615	0.9418	0.8504	0.0409
11	Dyrskjot	0.7939	0.8206	**0.8391**	0.4235	0.2099	0.2279
12	Garber	0.4674	0.4075	**0.4836**	0.2284	0.1366	0.2344
13	Golub-V1	**0.9443**	0.9171	0.9295	0.8210	0.2666	0.7843
14	Golub-V2	0.7333	**0.856**	0.7873	0.6106	0.2974	0.4900
15	Gordon	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.2968	0.9469
16	Khan	0.7960	0.8725	**0.9095**	0.5869	0.4476	0.3930
17	Laiho	**1.0000**	**1.0000**	0.7608	**1.0000**	0.2480	0.3796
18	Lapointe-V1	0.3215	0.3621	**0.3996**	0.1629	0.0776	0.1583
19	Lapointe-V2	**0.3708**	0.3588	0.3639	0.0908	0.0255	0.2554
20	Liang	0.5699	**0.6108**	0.3619	0.2996	0.1123	0.1573
21	Nutt-V1	**0.9563**	0.91	0.7990	0.8970	0.4647	0.1726
22	Nutt-V2	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.6033	0.0023
23	Nutt-V3	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.1099	0.0487
24	Pomeroy-V1	**1.0000**	**1.0000**	**1.0000**	**1.0000**	0.3633	0.0203
25	Pomeroy-V2	**0.9600**	0.9516	0.8320	0.8494	0.7251	0.5451
26	Ramaswamy	0.6189	0.6421	0.6308	0.3989	**0.7048**	0.1685
27	Rising	**0.9594**	0.9428	0.7478	0.9428	0.4951	0.3318
28	Shipp-V1	**1.0000**	**1.0000**	0.9443	**1.0000**	0.1583	0.0161
29	Singh	**1.0000**	**1.0000**	0.9048	**1.0000**	0.5807	0.0404
30	Su	0.7774	**0.7891**	0.6773	0.5920	0.3782	0.3808
31	Tomlins	**0.9513**	0.8843	0.8539	0.8831	0.7246	0.1814
32	Tomlins-V2	**0.9718**	0.9446	0.8537	0.9562	0.7412	0.1359
33	West	**0.7652**	0.6938	0.7288	0.1911	0.1202	0.0562
34	Yeoh-V1	**1.0000**	**1.0000**	0.6369	0.9967	0.0626	0.9205
35	Yeoh-V2	0.8543	0.8507	0.6355	**0.8637**	0.2545	0.1423

TopoBARTMAP variants. The Wilcoxon statistical test for these two classifiers resulted in a *p-value* of 1, which indicates high statistical similarity. For later tests, TopoBARTMAP with Pearson correlation was used for assessment to maintain fairness in comparison with

BARTMAP, because the latter uses Pearson correlation coefficient. The omnibus test on results reported in Table 2 produced *p-value* < 2.2e-16. The observed small *p-value* indicates that at least one of the methods performs differently than the rest; thus, we proceeded to perform the Friedman post-hoc test with Bergmann and Hommel's correction. The corresponding pair-wise comparison *p-values* are reported in the Table 3 and the statistical difference plot is shown in the Figure 2. The statistical difference plot shows that TopoBARTMAP performs statistically different from the other algorithms. The test further shows while the BARTMAP and TopoART perform similarly, whereas Spectral Biclustering performs similar to Fuzzy ART.

Table 3. Statistical comparison of different algorithms made using post-hoc test for observed biclustering results reported in Table 2. The *p-values* less than 0.05 are reported in bold.

	TopoBARTMAP	BARTMAP	TopoART	FuzzyART	SpectralBiclustering
TopoBARTMAP	n/a	**0.047**	**0.008**	**0.000**	**0.000**
BARTMAP	**0.047**	n/a	0.729	**0.000**	**0.000**
TopoART	**0.008**	0.729	n/a	**0.000**	**0.000**
FuzzyART	**0.000**	**0.000**	**0.000**	n/a	0.729
SpectralBiclustering	**0.000**	**0.000**	**0.000**	0.729	n/a

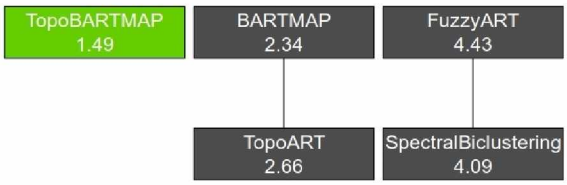


Figure 2. Statistical difference plot for the performance of each method on original data ordering experiments. These plots are produced using the Friedman post-hoc test with Bergmann and Hommel's correction. The omnibus test conducted to verify performance difference produced a *p-value* = 2.2e-16, indicative of significant difference. Each method is shown with its average rank in the block. The connections between blocks indicates statistical similarity in performance between corresponding methods.

5.2. RANDOMISED DATA ORDERING EXPERIMENT

To study the effect of randomization in presentation order on the performance of ART based methods, we presented randomized data to BARTMAP, TopoART, and TopoBARTMAP. The presentation order randomization in each of the thirty-five data-sets is achieved by permuting rows followed by columns. While similar experiments were conducted in our earlier work [22]; the permutations were limited to either rows or columns affecting the performance of one of the two ART modules of biclustering methods. By permuting rows followed by columns, both the ART modules are considered for the test, and hence a more challenging problem. With performance measure by Adjusted Rand Index, for each of the randomized data-set, BARTMAP, TopoART, and TopoBARTMAP were run with GA optimization with aforesaid settings. To run experiments with TopoBARTMAP, only Pearson Correlation Coefficient is used as a similarity measure due to the computational demand of running MIC [86] and statistical similarity in performance of both variants. Further, since permutations in data would only permute the left and right singular vectors of data matrix, Spectral Biclustering was not chosen for this study.

Table 4 summarizes the results of experimentation on randomized data-sets for ART-based agglomerative methods. In contrast to experiments with original data-sets, the TopoBARTMAP outperforms all the compared methods in twenty-six data-sets. While TopoBARTMAP and TopoART perform equally on eight data-sets, BARTMAP, while outperforming other methods on one data-set, performs equally with other compared methods on four data-sets. This similar performance is observed on the same set of data-sets as reported in Table 2. Further, it is observed that each biclustering method compared has a significant change in the best ARI values found. Whilst BARTMAP's performance is observed to degrade overall, where best ARI values were noticed to decline for 15 data-sets and improve for 10 data-sets; the performance of TopoBARTMAP remains relatively same, where best ARI values were observed to decline for 11 data-sets and improve for 12 data-sets than those reported in Table 2. Statistical tests were conducted to assess similarity in

the results, the omnibus test resulted in a *p-value* of 5.623e-10. The registered small *p-value* indicates that at least one of the algorithms performs differently from the others. Following omnibus test, Friedman post-hoc test with Bergmann and Hommel's correction was conducted to make a pair-wise comparison. The corresponding *p-value* and statistical differences plot are presented as Table 6 and Figure 3 respectively.

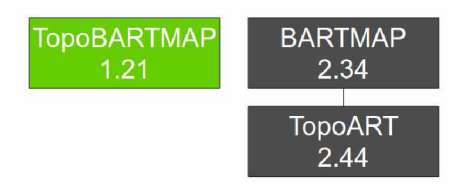


Figure 3. Statistical difference plot produced for the performance of each compared method on the randomized data-sets. The statistical significance is tested with the Friedman post-hoc test with Bergmann and Hommel's correction. The omnibus test conducted before post-hoc test produced a p-value = 5.623e-10, indicative of significant difference in performance of at least one method. In the plot, each method is shown with its average rank in the respective block. The connections between blocks indicate statistical similarity in performance between corresponding methods.

5.3. NCBI GEO GSE89116 GENE EXPRESSION STUDY-CASE

The experimentation with TopoBARTMAP to study the *Topological Gene Bicluster Association Network* on Gene Expression Omnibus GSE89116 data-set resulted in the best ARI of 0.8436 at the following parameter combination $\rho_a = 0.5127$, $\beta^a_{second-best} = 0.9913$, $\phi_a = 0.6025$, $\tau^{\%}_a = 0.1071$, $\rho_b = 0.9500$, $\beta^b_{second-best} = 0.0914$, $\phi_b = 2.7127$, $\tau^{\%}_b = 0.1617$, and $\eta = 0.5848$. For the mentioned parameter combination, TopoBARTMAP identified 577 gene prototypes. Further, 8 observation prototypes were created in the TopoARTa module which formed 4 clusters. The topological graph of the identified observation clusters, with prototypes colored according to the cluster, is shown in the Figure 4. The clustering misclassified two Late-normal observations as Early-tumors and one Late-normal sample as Late-tumor.

Table 4. Observed biclustering results on randomized data-sets. The optimal Adjusted Rand Index found for each method with GA are reported against the data-set. Best performances for each data-set are reported in bold. For these test TopoBARTMAP with Pearson Correlation was used.

Sl.no	DataSet	TopoBARTMAP	BARTMAP	TopoART
1	Alizadeh-V1	**0.6467**	0.5515	0.1899
2	Alizadeh-V2	**0.9799**	0.8952	0.9186
3	Alizadeh-V3	**0.5678**	0.5607	0.4852
4	Armstrong-V1	**1.0000**	0.5435	**1.0000**
5	Armstrong-V2	**1.0000**	0.8406	0.9583
6	Bhattacharjee	**1.0000**	0.9838	0.9779
7	Bittner	**0.7766**	0.5775	0.5931
8	Bredel	**0.7242**	0.6007	0.2085
9	Chen	**0.5761**	0.4630	0.1857
10	Chowdary	**1.0000**	0.9615	0.9418
11	Dyrskjot	**0.8692**	0.8183	0.4235
12	Garber	**0.6257**	0.6155	0.2213
13	Golub-V1	**0.9811**	0.8894	0.8234
14	Golub1-V2	**0.8861**	0.6772	0.6106
15	Gordon	**1.0000**	**1.0000**	**1.0000**
16	Khan	**0.8885**	0.7911	0.5869
17	Laiho	**1.0000**	0.6534	**1.0000**
18	Lapointe-V1	**0.3735**	0.3728	0.2232
19	Lapointe-V2	0.3450	**0.3789**	0.1143
20	Liang	**0.5290**	0.3863	0.2996
21	Nutt-V1	**0.9169**	0.7990	0.8970
22	Nutt-V2	**1.0000**	**1.0000**	**1.0000**
23	Nutt-V3	**1.0000**	**1.0000**	**1.0000**
24	Pomeroy-V1	**1.0000**	**1.0000**	**1.0000**
25	Pomeroy-V2	**0.9302**	0.8651	0.8494
26	Ramaswamy	**0.6052**	0.5919	0.3989
27	Rising	**0.9428**	0.7479	0.9428
28	Shipp-V1	**1.0000**	0.9443	**1.0000**
29	Singh	**1.0000**	0.8877	**1.0000**
30	Su	**0.6844**	0.6826	0.5920
31	Tomlins-V1	**0.9206**	0.8120	0.8831
32	Tomlins-V2	**0.9718**	0.8714	0.9562
33	West	**0.9183**	0.6607	0.2238
34	Yeoh-V1	**1.0000**	0.6421	0.9967
35	Yeoh-V2	**0.8859**	0.6171	0.8449

Table 5. Post-hoc test results for observed performances reported in Table 4. The *p-values* less than 0.05 are reported in bold, which shows that TopoBARTMAP performs differently from the rest.

	TopoBARTMAP	BARTMAP	TopoART
TopoBARTMAP	n/a	**0.000**	**0.000**
BARTMAP	**0.000**	n/a	0.676
TopoART	**0.000**	0.676	n/a

For each observation class,to recover biclusters formed by the associated observation prototypes with gene prototypes, average correlation defined in Eq. (12) was used as a coherence measure [87]. Biclusters with average correlation above a minimum of 0.5848, the optimal correlation level identified during optimization, were considered as homogeneous. During bicluster recovery, it is possible to assign gene prototypes with multiple sample prototypes to form biclusters. However, for this study, each gene prototype is assigned to the sample prototype to results in a bicluster with the highest $\delta(\mathbf{B}_{M\times N})$.

The topological graph of gene prototypes is refined to depict the bicluster associations using the recovered biclusters for each observation class. The refinement is achieved by associating graph nodes to the corresponding observation classes and color-coding them with the class colors used for observation clusters in Figure 4. Since each gene prototype is a multi-dimensional vector, the t-Distributed Stochastic Neighbor Embedding (t-SNE) [88] was used to reduce the dimensionality and embed each prototype in a 2-D space. The resultant network is shown in Figure 5. The produced graph is named as *topological gene bicluster association network.*

By varying the recovery correlation threshold, η_r, it was observed that for a particular observation prototype, a subset of gene prototypes are uniquely associable; they form bicluster only with this prototype and not with other prototypes in the observation class. Further, not all gene prototypes were observed to form biclusters, even at the optimal correlation level of 0.5848 found by GA as shown in the Figure 5. In the produced topological graphs, moreover, it was noticed that there are no direct links between nodes corresponding to Early

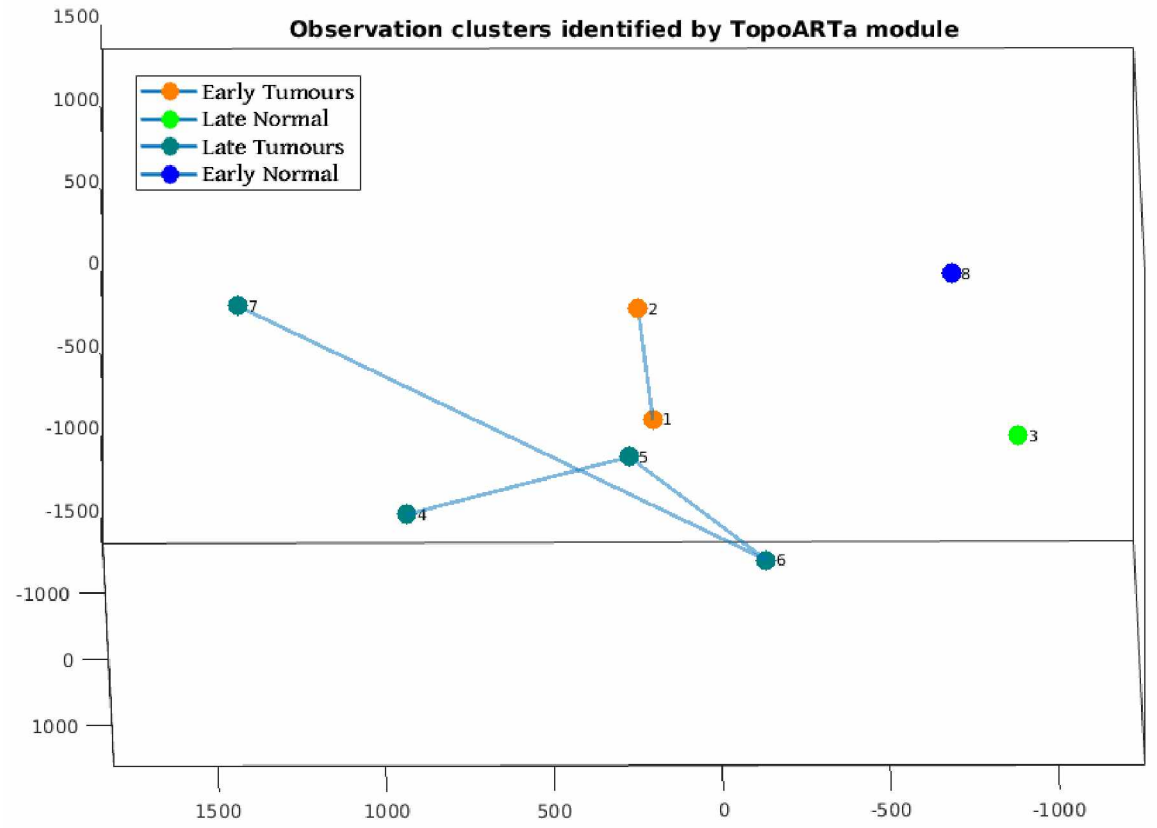


Figure 4. The sample classes identified by TopoARTa module of TopoBARTMAP for GSE89116 data-set. Each colored dot with corresponding number represents a prototype, with its association with other prototypes established by an edge.

Tumors (Orange color) and Early Normal tissue (Blue color). Nodes associated with Early Tumors are directly connected to Late Tumors (Blue-Green color) or Late Normal (Green color) nodes; the latter could be due to misclassification. To verify the significance of disease-specific biclusters, genes associated with these biclusters were compared with the differentially expressed genes reported by [26]. Of the identified genes, 195 were reported to be differentially expressed by [26], presented in Table 6. Because the presented results in this section were observations based on clustering, we can not vouch for the medical interpretation.

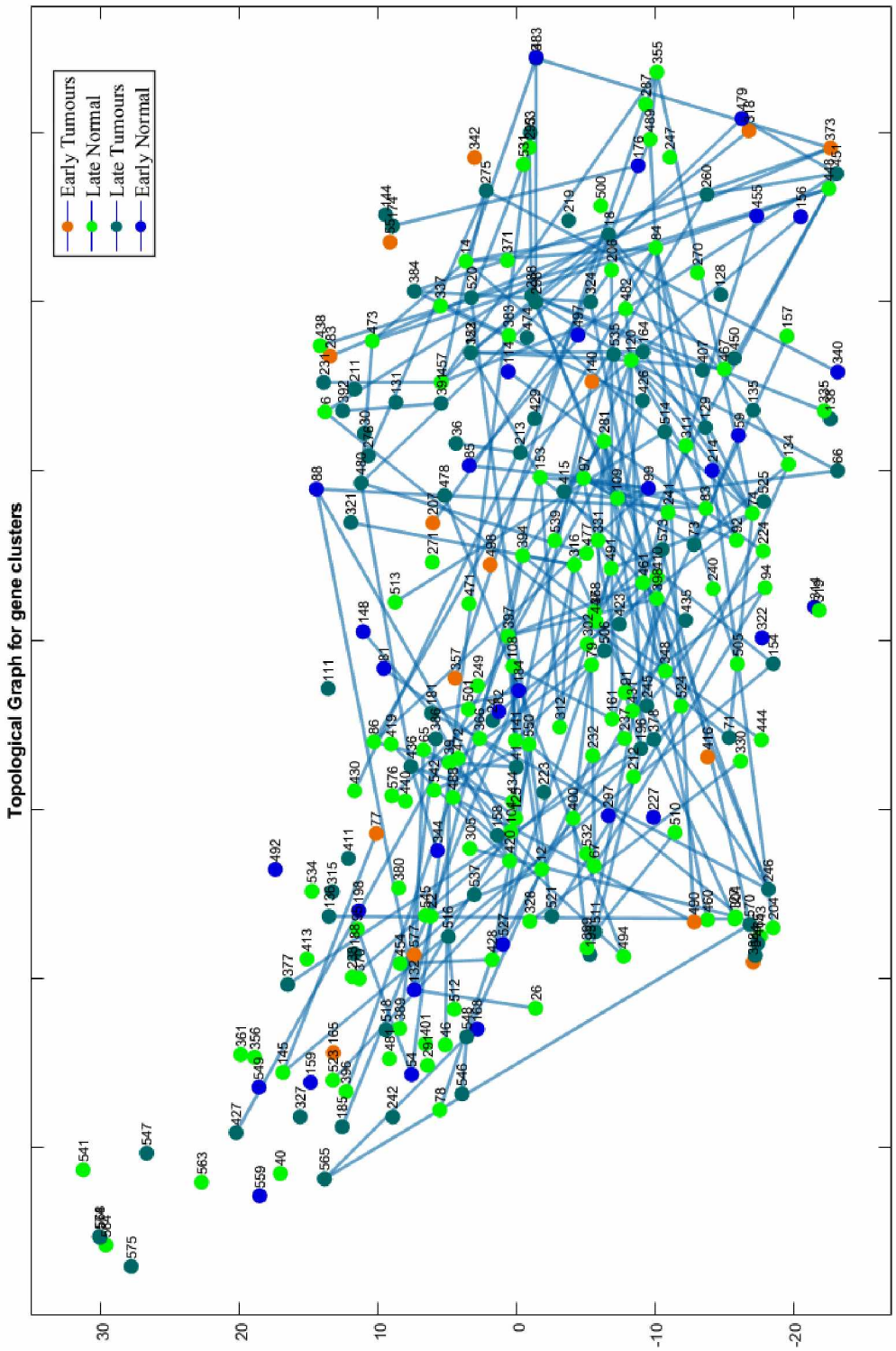


Figure 5. The *Topological Gene Bi-cluster Association Network* at correlation level of 0.5848. The node colors indicate the associated bi-cluster identified by the TopoBARTMAP.

Table 6. Identified highly active differentially expressed genes with corresponding observation prototypes and tumor categories.

Observation Prototypes	Category	Genes
1 and 2	Early tumor	AXUD1, B3GALNT1, C6, CDO1, CRY2, LEP, LEPR, LIPE, LMOD3, NCALD, PCOLCE2, RDH5, SLC19A3, THRSP, TIMP4
4, 5, 6, and 7	Late tumor	ABCA8, ADCY4, ADRB2, AGPAT2, AHNAK, AK5, AKAP12, AKAP13, ALDH18A1, ALDH1A2, ALDH1L1, ALDH2, ALDOA, ALPP, AMACR, AMFR, AMMECR1, AMOTL2, ANGPT1, ANGPTL1, ANGPTL4, ANGPTL7, ANK2, ANKRD29, APBB3, APOB, APOLD1, AQP7P2, ARID4B, ARL16, AURKA, BCHE, BLZF1, BOLA2, BPNT1, BRI3BP, BTNL9, C12ORF39, C14ORF85, C17ORF75, CA3, CALB2, CAT, CCNL1, CD248, CDC20, CDCA4, CEBPA, CEP350, CIDEC, CKS2, CLEC3B, CNN1, COL10A1, COL11A1, COPE, CORIN, CORO2A, CPXM2, CRB3, CSN1S1, DNAJB9, DSP, DTL, ECHDC3, ECM2, EHBP1, EID2B, EIF1B, ENG, ENPP2, EPB41L2, FAM102A, FHL1, FLAD1, FLJ21986, FLNC, FREM1, G0S2, G3BP1, GDF10, GJA4, GJB2, GPAM, GRK5, GRRP1, GSN, GSTM5, GYPC, HIST1H2BD, HIST1H2BJ, HIST2H2AA3, HIST2H2AC, HMGCS1, HPRT1, HS.163752, HSD17B13, HSD17B6, ICAM2, ID1, IFI6, IGF2, IL11RA, IL17B, KIF25, KIT, KLF11, KLF4, KRTAP21-2, LAMA4, LDB2, LMO2, LOC440348, LOC440359, LOC644322, LOC651816, LOC90586, LPL, LRAP, MFAP4, MMP1, MMP10, MMP13, MMP3, MOSC1, MRAP, MRPL12, NDRG2, NMT2, PALM, PECAM1, PELI1, PENK, PFKFB4, PGAM4, PI16, PLEKHM1, PLTP, PNPLA2, PPAP2A, PPAP2B, PPARG, RAD21, RARRES2, RBM6, RBP4, RBPMS2, RCP9, RIPK2, ROPN1, ROPN1B, RSPRY1, RXRA, SALL4, SDPR, SEMA3E, SHANK3, SLC38A1, SLCO2A1, SPRY1, SPTBN1, SRPX, SSPN, ST6GALNAC6, TDO2, TESC, TLE4, TMEM79, TMEM9, TOP2A, TSPAN7, TSPAN8, TUBB3, TYMS, UBE2T, VTI1B, WT1, YIF1A, ZMYND8, ZNF14

6. CONCLUSION

This work presents TopoBARTMAP, which uses TopoART modules and a maximal information coefficient or Pearson Correlation test to detect the similarity between samples during training. The experimental results on a collection of datasets demonstrated the statistical superiority of TopoBARTMAP's performance compared to BARTMAP, TopoART, Fuzzy ART and Spectral Biclustering. However, our experimentation demonstrates that it

is not always the case that the detection of nonlinear relationships improves the efficiency of clustering gene expression data. We suggest using the Pearson Correlation coefficient whenever the problem can be adequately characterized by linear relationships.

The experiments on randomized datasets demonstrate the relative robustness of the method to the order of presentation. We suggest using GA optimization to maximize average performance evaluated on multiple permutations of data.

The exploratory analysis on the GSE89116 data-set using a topological network produced by TopoBARTMAP showcases the method's utility as a tool for identifying modular interactions. With the method's ability to utilize topological information to uncover complex-shaped biclusters coupled with the ability to serve as a TDA tool, we posit that TopoBARTMAP can be used to discover intricate structure within data.

ACKNOWLEDGMENT

This research was sponsored by the Missouri University of Science and Technology Mary K. Finley Endowment and Intelligent Systems Center; the Army Research Laboratory (ARL) and the Lifelong Learning Machines program from the DARPA / Microsystems Technology Office, and it was accomplished under Cooperative Agreement Number W911NF-18-2-0260. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. This material is based on work partially supported by the National Science Foundation. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] O'brien, M., Costin, B., and Miles, M., 'Using genome-wide expression profiling to define gene networks relevant to the study of complex traits: from rna integrity to network topology,' in 'International review of neurobiology,' volume 104, pp. 91–133, Elsevier, 2012.

[2] Madeira, S. C. and Oliveira, A. L., 'Biclustering algorithms for biological data analysis: a survey,' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2004, **1**(1), pp. 24–45.

[3] Pontes, B., GirÃ¡ldez, R., and Aguilar-Ruiz, J. S., 'Biclustering on expression data: A review,' Journal of Biomedical Informatics, 2015, **57**, pp. 163 – 180, ISSN 1532-0464, doi:https://doi.org/10.1016/j.jbi.2015.06.028.

[4] Padilha, V. A. and Campello, R. J., 'A systematic comparative evaluation of biclustering techniques,' BMC bioinformatics, 2017, **18**(1), p. 55.

[5] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W., 'From molecular to modular cell biology,' Nature, Dec 1999, **402**(6761), pp. C47–C52, ISSN 1476-4687, doi:10.1038/35011540.

[6] Mitra, K., Carvunis, A.-R., Ramesh, S. K., and Ideker, T., 'Integrative approaches for finding modular structure in biological networks,' Nature Reviews Genetics, Oct 2013, **14**(10), pp. 719–732, ISSN 1471-0064, doi:10.1038/nrg3552.

[7] Singh, A. J., Ramsey, S. A., Filtz, T. M., and Kioussi, C., 'Differential gene regulatory networks in development and disease,' Cellular and Molecular Life Sciences, Mar 2018, **75**(6), pp. 1013–1025, ISSN 1420-9071, doi:10.1007/s00018-017-2679-6.

[8] Saelens, W., Cannoodt, R., and Saeys, Y., 'A comprehensive evaluation of module detection methods for gene expression data,' Nature Communications, Mar 2018, **9**(1), p. 1090, ISSN 2041-1723, doi:10.1038/s41467-018-03424-4.

[9] Chazal, F. and Michel, B., 'An introduction to topological data analysis: fundamental and practical aspects for data scientists,' arXiv preprint arXiv:1710.04019, 2017.

[10] Patania, A., Vaccarino, F., and Petri, G., 'Topological analysis of data,' EPJ Data Science, 2017, **6**(1), p. 7.

[11] Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G., 'Extracting insights from the shape of complex data using topology,' Scientific reports, 2013, **3**, p. 1236.

[12] Carlsson, G., 'Topology and data,' Bulletin of the American Mathematical Society, 2009, **46**(2), pp. 255–308.

[13] Lockwood, S. and Krishnamoorthy, B., 'Topological features in cancer gene expression data,' in 'Pacific Symposium on Biocomputing Co-Chairs,' World Scientific, 2014 pp. 108–119.

[14] Snášel, V., Nowaková, J., Xhafa, F., and Barolli, L., 'Geometrical and topological approaches to big data,' Future Generation Computer Systems, 2017, **67**, pp. 286–296.

[15] Edelsbrunner, H. and Harer, J., *Computational Topology: An Introduction*, American Mathematical Society, 01 2010, ISBN 978-0-8218-4925-5, doi:10.1007/978-3-540-33259-6_7.

[16] Singh, G., Mémoli, F., and Carlsson, G. E., 'Topological methods for the analysis of high dimensional data sets and 3d object recognition.' in 'SPBG,' 2007 pp. 91–100.

[17] Ma, X., Gao, L., and Tan, K., 'Modeling disease progression using dynamics of pathway connectivity,' Bioinformatics, 04 2014, **30**(16), pp. 2343–2350, ISSN 1367-4803.

[18] Nielson, J. L., Paquette, J., Liu, A. W., Guandique, C. F., Tovar, C. A., Inoue, T., Irvine, K.-A., Gensel, J. C., Kloke, J., Petrossian, T. C., Lum, P. Y., Carlsson, G. E., Manley, G. T., Young, W., Beattie, M. S., Bresnahan, J. C., and Ferguson, A. R., 'Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury,' Nature Communications, Oct 2015, **6**(1), p. 8581, ISSN 2041-1723, doi:10.1038/ncomms9581.

[19] Nicolau, M., Levine, A. J., and Carlsson, G., 'Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,' Proceedings of the National Academy of Sciences, 2011, **108**(17), pp. 7265–7270.

[20] Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., and Rabadan, R., 'Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development,' Nature biotechnology, 2017, **35**(6), p. 551.

[21] Perkins, M. and Daniels, K., 'Visualizing dynamic gene interactions to reverse engineer gene regulatory networks using topological data analysis,' in '2017 21st International Conference Information Visualisation (IV),' ISSN 2375-0138, July 2017 pp. 384–389.

[22] Yelugam, R., Brito da Silva, L. E., and Wunsch, D. C., 'Topobartmap: Biclustering artmap with or without topological methods in a blood cancer case study,' in '2020 International Joint Conference on Neural Networks (IJCNN),' 2020 pp. 1–8.

[23] Xu, R. and Wunsch II, D. C., 'BARTMAP: A viable structure for biclustering,' Neural Networks, 2011, **24**(7), pp. 709–716.

[24] Tscherepanow, M., 'TopoART: A topology learning hierarchical ART network,' in 'International Conference on Artificial Neural Networks,' Springer, 2010 pp. 157–167.

[25] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C., 'Detecting novel associations in large data sets,' Science, 2011, **334**(6062), pp. 1518–1524, ISSN 0036-8075, doi:10.1126/science.1205438.

[26] Malvia, S., Bagadi, S. A. R., Pradhan, D., Chintamani, C., Bhatnagar, A., Arora, D., Sarin, R., and Saxena, S., 'Study of gene expression profiles of breast cancers in indian women,' Scientific Reports, Jul 2019, **9**(1), p. 10018, ISSN 2045-2322, doi:10.1038/s41598-019-46261-1.

[27] Carpenter, G. A. and Grossberg, S., 'A massively parallel architecture for a self-organizing neural pattern recognition machine,' Computer vision, graphics, and image processing, 1987, **37**(1), pp. 54–115.

[28] Carpetner, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns,' Technical report, Boston University Center for Adaptive Systems and Department of Cognitive . . . , 1991.

[29] Carpenter, G. A. and Grossberg, S., *Adaptive resonance theory*, Springer, 2010.

[30] Grossberg, S., 'Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world,' Neural Networks, 2013, **37**, pp. 1–47.

[31] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'A survey of adaptive resonance theory neural network models for engineering applications,' Neural Networks, 2019, **120**, pp. 167 – 203.

[32] Wunsch II, D. C., 'Admiring the Great Mountain: A Celebration Special Issue in Honor of Stephen Grossberg's 80th Birthday,' Neural Networks, 2019, **120**, pp. 1 – 4, special Issue in Honor of the 80th Birthday of Stephen Grossberg.

[33] Xu, R. and Wunsch, D., *Clustering*, volume 10, John Wiley & Sons, 2008.

[34] Carpenter, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system,' Neural Networks, 1991, **4**(6), pp. 759 – 771, ISSN 0893-6080.

[35] Carpenter, G. A., Grossberg, S., and Reynolds, J. H., 'Artmap: supervised real-time learning and classification of nonstationary data by a self-organizing neural network,' in '[1991 Proceedings] IEEE Conference on Neural Networks for Ocean Engineering,' 1991 pp. 341–342.

[36] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B., 'Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps,' IEEE Transactions on Neural Networks, Sep. 1992, **3**(5), pp. 698–713, ISSN 1941-0093.

[37] Bartfai, G., 'Hierarchical clustering with ART neural networks,' in 'Proc. IEEE International Conference on Neural Networks (ICNN),' volume 2, Jun 1994 pp. 940–944.

[38] Brito da Silva, L. E., Rayapati, N., and Wunsch II, D. C., 'iCVI-ARTMAP: Accelerating and improving clustering using adaptive resonance theory predictive mapping and incremental cluster validity indices,' arXiv, 2020, arXiv:2008.09903v1 [cs.LG].

[39] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D., 'Cluster analysis and display of genome-wide expression patterns,' Proceedings of the National Academy of Sciences, 1998, **95**(25), pp. 14863–14868, ISSN 0027-8424.

[40] Song, L., Langfelder, P., and Horvath, S., 'Comparison of co-expression measures: mutual information, correlation, and model based indices,' BMC Bioinformatics, Dec 2012, **13**(1), p. 328, ISSN 1471-2105, doi:10.1186/1471-2105-13-328.

[41] Bain, L. J. and Engelhardt, M., *Introduction to Probability and Mathematical Statistics*, Brooks/Cole, Cengage Learning, 2 edition, 1992.

[42] Jaskowiak, P. A., Campello, R. J. G. B., and Costa, I. G., 'Proximity measures for clustering gene expression microarray data: A validation methodology and a comparative analysis,' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, **10**(4), pp. 845–857, doi:10.1109/TCBB.2013.9.

[43] Kim, S., *Novel approaches to clustering , biclustering algorithms based on adaptive resonance theory and intelligent control*, Ph.D. thesis, Missouri University of Science and Technology, 2016.

[44] Rodgers, J. L. and Nicewander, W. A., 'Thirteen ways to look at the correlation coefficient,' The American Statistician, 1988, **42**(1), pp. 59–66, doi:10.1080/00031305.1988.10475524.

[45] Weirauch, M. T., *Gene Coexpression Networks for the Analysis of DNA Microarray Data*, chapter 11, pp. 215–250, John Wiley & Sons, Ltd, ISBN 9783527638079, 2011, doi:https://doi.org/10.1002/9783527638079.ch11.

[46] de Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermir, T. B., and Schliep, A., 'Clustering cancer gene expression data: a comparative study,' BMC bioinformatics, 2008, **9**(1), p. 497.

[47] Quackenbush, J., 'Computational analysis of microarray data,' Nature Reviews Genetics, Jun 2001, **2**(6), pp. 418–427, ISSN 1471-0064, doi:10.1038/35076576.

[48] Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M., 'Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling,' Nature, Feb 2000, **403**(6769), pp. 503–511, ISSN 1476-4687, doi:10.1038/35000501.

[49] Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J., 'Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia,' Nature Genetics, Jan 2002, **30**(1), pp. 41–47, ISSN 1546-1718, doi:10.1038/ng765.

[50] Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M., 'Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses,' Proceedings of the National Academy of Sciences, 2001, **98**(24), pp. 13790–13795, ISSN 0027-8424, doi:10.1073/pnas.191502998.

[51] Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V., Hayward, N., and Trent, J., 'Molecular classification of cutaneous malignant melanoma by gene expression profiling,' Nature, Aug 2000, **406**(6795), pp. 536–540, ISSN 1476-4687, doi:10.1038/35020115.

[52] Bredel, M., Bredel, C., Juric, D., Harsh, G. R., Vogel, H., Recht, L. D., and Sikic, B. I., 'Functional network analysis reveals extended gliomagenesis pathway maps and three novel myc-interacting genes in human gliomas,' Cancer Research, 2005, **65**(19), pp. 8679–8689, ISSN 0008-5472, doi:10.1158/0008-5472.CAN-05-1204.

[53] Chen, X., Cheung, S. T., So, S., Fan, S. T., Barry, C., Higgins, J., Lai, K.-M., Ji, J., Dudoit, S., Ng, I. O. L., Van De Rijn, M., Botstein, D., and Brown, P. O., 'Gene expression patterns in human liver cancers,' Molecular biology of the cell, Jun 2002, **13**(6), pp. 1929–1939, ISSN 1059-1524, doi:10.1091/mbc.02-02-0023, 12058060[pmid].

[54] Chowdary, D., Lathrop, J., Skelton, J., Curtin, K., Briggs, T., Zhang, Y., Yu, J., Wang, Y., and Mazumder, A., 'Prognostic gene expression signatures can be measured in tissues collected in rnalater preservative,' The Journal of Molecular Diagnostics, 2006, **8**(1), pp. 31 – 39, ISSN 1525-1578, doi:https://doi.org/10.2353/jmoldx.2006.050056.

[55] Dyrskjã t, L., Thykjaer, T., KruhÃffer, M., Jensen, J. L., Marcussen, N., Hamilton-Dutoit, S., Wolf, H., and Ãrntoft, T. F., 'Identifying distinct classes of bladder carcinoma using microarrays,' Nature Genetics, Jan 2003, **33**(1), pp. 90–96, ISSN 1546-1718, doi:10.1038/ng1061.

[56] Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. I., Altman, R. B., Brown, P. O., Botstein, D., and Petersen, I., 'Diversity of gene expression in adenocarcinoma of the lung,' Proceedings of the National Academy of Sciences, 2001, **98**(24), pp. 13784–13789, ISSN 0027-8424, doi:10.1073/pnas.241500798.

[57] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S., 'Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring,' Science, 1999, **286**(5439), pp. 531–537, ISSN 0036-8075, doi:10.1126/science.286.5439.531.

[58] Gordon, G. J., Jensen, R. V., Hsiao, L.-L., Gullans, S. R., Blumenstock, J. E., Ramaswamy, S., Richards, W. G., Sugarbaker, D. J., and Bueno, R., 'Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma,' Cancer Research, 2002, **62**(17), pp. 4963–4967, ISSN 0008-5472.

[59] Khan, J., Wei, J. S., RingnÃ©r, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S., 'Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks,' Nature Medicine, Jun 2001, **7**(6), pp. 673–679, ISSN 1546-170X, doi:10.1038/89044.

[60] Laiho, P., Kokko, A., Vanharanta, S., Salovaara, R., Sammalkorpi, H., JÃrvinen, H., Mecklin, J.-P., Karttunen, T. J., Tuppurainen, K., Davalos, V., Schwartz, S., Arango, D., MÃkinen, M. J., and Aaltonen, L. A., 'Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis,' Oncogene, Jan 2007, **26**(2), pp. 312–320, ISSN 1476-5594, doi:10.1038/sj.onc.1209778.

[61] Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R., 'Gene expression profiling identifies clinically relevant subtypes of prostate cancer,' Proceedings of the National Academy of Sciences, 2004, **101**(3), pp. 811–816, ISSN 0027-8424, doi:10.1073/pnas.0304146101.

[62] Liang, Y., Diehn, M., Watson, N., Bollen, A. W., Aldape, K. D., Nicholas, M. K., Lamborn, K. R., Berger, M. S., Botstein, D., Brown, P. O., and Israel, M. A., 'Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme,' Proceedings of the National Academy of Sciences, 2005, **102**(16), pp. 5814–5819, ISSN 0027-8424, doi:10.1073/pnas.0402870102.

[63] Nutt, C. L., Mani, D. R., Betensky, R. A., Tamayo, P., Cairncross, J. G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M. E., Batchelor, T. T., Black, P. M., von Deimling, A., Pomeroy, S. L., Golub, T. R., and Louis, D. N., 'Gene expression-based classification of malignant gliomas correlates better with survival than histological classification,' Cancer Research, 2003, **63**(7), pp. 1602–1607, ISSN 0008-5472.

[64] Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y. H., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D. N., Mesirov, J. P., Lander, E. S., and Golub, T. R., 'Prediction of central nervous system embryonal tumour outcome based on gene expression,' Nature, Jan 2002, **415**(6870), pp. 436–442, ISSN 1476-4687, doi:10.1038/415436a.

[65] Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.-H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S., and Golub, T. R., 'Multiclass cancer diagnosis using tumor gene expression signatures,' Proceedings of the National Academy of Sciences, 2001, **98**(26), pp. 15149–15154, ISSN 0027-8424, doi:10.1073/pnas.211566398.

[66] Risinger, J. I., Maxwell, G. L., Chandramouli, G. V. R., Jazaeri, A., Aprelikova, O., Patterson, T., Berchuck, A., and Barrett, J. C., 'Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer,' Cancer Research, 2003, **63**(1), pp. 6–11, ISSN 0008-5472.

[67] Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. W., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C., and Golub, T. R., 'Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning,' Nature Medicine, Jan 2002, **8**(1), pp. 68–74, ISSN 1546-170X, doi:10.1038/nm0102-68.

[68] Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R., 'Gene expression correlates of clinical prostate cancer behavior,' Cancer Cell, 2002, **1**(2), pp. 203 – 209, ISSN 1535-6108, doi:https://doi.org/10.1016/S1535-6108(02)00030-2.

[69] Su, A. I., Welsh, J. B., Sapinoso, L. M., Kern, S. G., Dimitrov, P., Lapp, H., Schultz, P. G., Powell, S. M., Moskaluk, C. A., Frierson, H. F., and Hampton, G. M., 'Molecular classification of human carcinomas by use of gene expression signatures,' Cancer Research, 2001, **61**(20), pp. 7388–7393, ISSN 0008-5472.

[70] West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Marks, J. R., and Nevins, J. R., 'Predicting the clinical status of human breast cancer by using gene expression profiles,' Proceedings of the National Academy of Sciences, 2001, **98**(20), pp. 11462–11467, ISSN 0027-8424, doi:10.1073/pnas.201162998.

[71] Yeoh, E.-J., Ross, M. E., Shurtleff, S. A., Williams, W., Patel, D., Mahfouz, R., Behm, F. G., Raimondi, S. C., Relling, M. V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C.-H., Evans, W. E., Naeve, C., Wong, L., and Downing, J. R., 'Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling,' Cancer Cell, 2002, **1**(2), pp. 133 – 143, ISSN 1535-6108, doi:https://doi.org/10.1016/S1535-6108(02)00032-6.

[72] Tomlins, S. A., Mehra, R., Rhodes, D. R., Cao, X., Wang, L., Dhanasekaran, S. M., Kalyana-Sundaram, S., Wei, J. T., Rubin, M. A., Pienta, K. J., Shah, R. B., and Chinnaiyan, A. M., 'Integrative molecular concept modeling of prostate cancer progression,' Nature Genetics, Jan 2007, **39**(1), pp. 41–51, ISSN 1546-1718, doi:10.1038/ng1935.

[73] Moshtaghi, M., Bezdek, J. C., Erfani, S. M., Leckie, C., and Bailey, J., 'Online cluster validity indices for performance monitoring of streaming data clustering,' International Journal of Intelligent Systems, 2019, **34**(4), pp. 541–563.

[74] Ibrahim, O. A., Keller, J. M., and Bezdek, J. C., 'Evaluating Evolving Structure in Streaming Data With Modified Dunn's Indices,' IEEE Transactions on Emerging Topics in Computational Intelligence, 2019, pp. 1–12, ISSN 2471-285X.

[75] Brito Da Silva, L. E., Melton, N. M., and Wunsch II, D. C., 'Incremental Cluster Validity Indices for Online Learning of Hard Partitions: Extensions and Comparative Study,' IEEE Access, 2020, **8**, pp. 22025–22047, doi:10.1109/ACCESS.2020.2969849.

[76] Hubert, L. and Arabie, P., 'Comparing partitions,' J. Classification, 1985, **2**(1), pp. 193–218.

[77] Brouwer, R. K., 'Extending the rand, adjusted rand and jaccard indices to fuzzy partitions,' Journal of Intelligent Information Systems, 2009, **32**(3), pp. 213–235, ISSN 1573-7675, doi:10.1007/s10844-008-0054-7.

[78] Horta, D. and Campello, R. J., 'Comparing hard and overlapping clusterings,' Journal of Machine Learning Research, 2015, **16**(93), pp. 2949–2997.

[79] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 'Scikit-learn: Machine learning in Python,' Journal of Machine Learning Research, 2011, **12**, pp. 2825–2830.

[80] Wang, K., Wang, B., and Peng, L., 'CVAP: validation for cluster analyses,' Data Science Journal, 2009, pp. 0904220071–0904220071.

[81] Albanese, D., Filosi, M., Visintainer, R., Riccadonna, S., Jurman, G., and Furlanello, C., 'minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers,' Bioinformatics, 12 2012, **29**(3), pp. 407–408, ISSN 1367-4803, doi:10.1093/bioinformatics/bts707.

[82] Calvo, B. and Santafé Rodrigo, G., 'scmamp: Statistical comparison of multiple algorithms in multiple problems,' The R Journal, Vol. 8/1, Aug. 2016, 2016.

[83] Demšar, J., 'Statistical comparisons of classifiers over multiple data sets,' Journal of Machine Learning Research, 2006, **7**(1), pp. 1–30.

[84] Eiben, A. E. and Smith, J. E., *Introduction to Evolutionary Computing*, Springer Publishing Company, Incorporated, 2nd edition, 2015.

[85] Elnabarawy, I., Wunsch, D. C., and Abdelbar, A. M., 'Biclustering ARTMAP collaborative filtering recommender system,' in '2016 International Joint Conference on Neural Networks (IJCNN),' ISSN 2161-4407, July 2016 pp. 2986–2991.

[86] Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C., and Mitzenmacher, M., 'Measuring dependence powerfully and equitably,' The Journal of Machine Learning Research, 2016, **17**(1), p. 7406–7468, ISSN 1532-4435.

[87] Padilha, V. A. and de Carvalho, A. C. P. L. F., 'A study of biclustering coherence measures for gene expression data,' in '2018 7th Brazilian Conference on Intelligent Systems (BRACIS),' 2018 pp. 546–551, doi:10.1109/BRACIS.2018.00100.

[88] van der Maaten, L. and Hinton, G., 'Visualizing data using t-sne,' Journal of Machine Learning Research, 2008, **9**(86), pp. 2579–2605.

SECTION

2. CONCLUSIONS

The objective of this thesis is to develop a biclustering method to identify intra-cluster gene-functional relationships while associating functionally related genes that might not coexpress significantly. To meet this objective, a biclustering method, BARTMAP, was integrated with a topology learning clustering method, TopoART, to produce TopoBARTMAP. The rationale driving the integration is that identifying the spatial relations of the observations within a bicluster can associate gene subgroups within the same cluster. Furthermore, recovering the shape of the objects in the data can aid in recognizing arbitrarily shaped clusters. Along with the detection of local linear relationships between genes and observations, TopoBARTMAP was extended to detect non-linear relationships with the aid of Mutual Information Coefficient.

The experimental study on 35 benchmark data-sets demonstrated a statistically significant improvement, subject to the Adjusted Rand Index, in the biclustering accuracy. It was also observed that the detection of non-linear patterns by the method might not necessarily result in better biclusters. Moreover, TopoBARTMAP is less susceptible to noise and is more robust to presentation order than either TopoART or BARTMAP. Nonetheless, the experiments on the parameter sensitivity indicate that the optimal hyperparameters for one presentation order might not work for different presentation order. Due to the rigidness and computational costs, this thesis suggests using TopoBARTMAP on data with complex structures which are hard to identify. Otherwise, this thesis recommends using BARTMAP. The analysis on the breast cancer data-set (GSE89116) accentuates the utility of TopoBARTMAP as a tool for uncovering interactions between gene groups.

REFERENCES

[1] O'Connor, C. M., Adams, J. U., and Fairman, J., 'Essentials of cell biology,' Cambridge, MA: NPG Education, 2010, **1**, p. 54.

[2] Frodsham, A. J. and Hill, A. V., 'Genetics of infectious diseases,' Human Molecular Genetics, 2004, **13**(suppl_2), pp. R187–R194.

[3] Hubner, N., Wallace, C. A., Zimdahl, H., Petretto, E., Schulz, H., Maciver, F., Mueller, M., Hummel, O., Monti, J., Zidek, V., *et al.*, 'Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease,' Nature genetics, 2005, **37**(3), pp. 243–253.

[4] Casanova, J.-L. and Abel, L., 'Human genetics of infectious diseases: a unified theory,' The EMBO journal, 2007, **26**(4), pp. 915–922.

[5] Krebs, O., Schäfer, B., Wolff, T., Oesterle, D., Deml, E., Sund, M., and Favor, J., 'The dna damaging drug cyproterone acetate causes gene mutations and induces glutathione-s-transferase p in the liver of female big blue transgenic f344 rats.' Carcinogenesis, 1998, **19**(2), pp. 241–245.

[6] Sinclair, A., 'Genetics 101: detecting mutations in human genes,' Cmaj, 2002, **167**(3), pp. 275–279.

[7] Mahdieh, N. and Rabbani, B., 'An overview of mutation detection methods in genetic disorders,' Iranian journal of pediatrics, 2013, **23**(4), p. 375.

[8] Vnencak-Jones, C., Berger, M., and Pao, W., 'Types of molecular tumor testing,' My Cancer Genome, 2016.

[9] Frayling, I. M., 'Methods of molecular analysis: mutation detection in solid tumours,' Molecular Pathology, 2002, **55**(2), p. 73.

[10] Mullis, K. B. and Faloona, F. A., '[21] specific synthesis of dna in vitro via a polymerase-catalyzed chain reaction,' Methods in enzymology, 1987, **155**, pp. 335–350.

[11] TAUB, E., FLOYD, DeLEO, J. M., and Thompson, E. B., 'Sequential comparative hybridizations analyzed by computerized image processing can identify and quantitate regulated rnas,' Dna, 1983, **2**(4), pp. 309–327.

[12] Heller, M. J., 'Dna microarray technology: devices, systems, and applications,' Annual review of biomedical engineering, 2002, **4**(1), pp. 129–153.

[13] , M. K. e. a., 'Book: Computational Biology - Genomes, Networks, and Evolution (Kellis et al.),' 1 2021, [Online; accessed 2021-03-23].

[14] D'haeseleer, P., 'How does gene expression clustering work?' Nature biotechnology, 2005, **23**(12), pp. 1499–1501.

[15] Jiang, D., Tang, C., and Zhang, A., 'Cluster analysis for gene expression data: a survey,' IEEE Transactions on knowledge and data engineering, 2004, **16**(11), pp. 1370–1386.

[16] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D., 'Cluster analysis and display of genome-wide expression patterns,' Proceedings of the National Academy of Sciences, 1998, **95**(25), pp. 14863–14868, ISSN 0027-8424.

[17] Xu, R. and Wunsch, D., *Clustering*, volume 10, John Wiley & Sons, 2008.

[18] Chazal, F. and Michel, B., 'An introduction to topological data analysis: fundamental and practical aspects for data scientists,' arXiv preprint arXiv:1710.04019, 2017.

[19] Rosenlicht, M., *Introduction to analysis*, chapter III, pp. 33–61, Dover Publications, ISBN 9780486650388, 1985.

[20] Armstrong, M. A., *Basic topology*, Springer Science & Business Media, 2013.

[21] Edelsbrunner, H. and Harer, J., *Computational Topology: An Introduction*, American Mathematical Society, 01 2010, ISBN 978-0-8218-4925-5, doi:10.1007/978-3-540-33259-6_7.

[22] Edelsbrunner, H. and Shah, N. R., 'Triangulating topological spaces,' in 'Proceedings of the tenth annual symposium on Computational geometry,' 1994 pp. 285–292.

[23] Cheng, Y. and Church, G. M., 'Biclustering of expression data.' in 'Ismb,' volume 8, 2000 pp. 93–103.

[24] Govaert, G. and Nadif, M., *Co-clustering: models, algorithms and applications*, John Wiley & Sons, 2013.

[25] Madeira, S. C. and Oliveira, A. L., 'Biclustering algorithms for biological data analysis: a survey,' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2004, **1**(1), pp. 24–45.

[26] Padilha, V. A. and Campello, R. J., 'A systematic comparative evaluation of biclustering techniques,' BMC bioinformatics, 2017, **18**(1), p. 55.

[27] Parsons, L., Haque, E., and Liu, H., 'Subspace clustering for high dimensional data: a review,' Acm Sigkdd Explorations Newsletter, 2004, **6**(1), pp. 90–105.

[28] Wang, H., Wang, W., Yang, J., and Yu, P. S., 'Clustering by pattern similarity in large data sets,' in 'Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data,' SIGMOD '02, Association for Computing Machinery, New York, NY, USA, ISBN 1581134975, 2002 p. 394–405.

[29] Chen, J., Ji, L., Hsu, W., Tan, K.-L., and Rhee, S. Y., 'Exploiting domain knowledge to improve biological significance of biclusters with key missing genes,' in '2009 IEEE 25th International Conference on Data Engineering,' IEEE, 2009 pp. 1219–1222.

[30] Patania, A., Vaccarino, F., and Petri, G., 'Topological analysis of data,' EPJ Data Science, 2017, **6**(1), p. 7.

[31] Carlsson, G., 'Topology and data,' Bulletin of the American Mathematical Society, 2009, **46**(2), pp. 255–308.

[32] Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., and Carlsson, G., 'Extracting insights from the shape of complex data using topology,' Scientific reports, 2013, **3**, p. 1236.

[33] Lockwood, S. and Krishnamoorthy, B., 'Topological features in cancer gene expression data,' in 'Pacific Symposium on Biocomputing Co-Chairs,' World Scientific, 2014 pp. 108–119.

[34] Singh, G., Mémoli, F., and Carlsson, G. E., 'Topological methods for the analysis of high dimensional data sets and 3d object recognition.' in 'SPBG,' 2007 pp. 91–100.

[35] Nicolau, M., Levine, A. J., and Carlsson, G., 'Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,' Proceedings of the National Academy of Sciences, 2011, **108**(17), pp. 7265–7270.

[36] Perkins, M. and Daniels, K., 'Visualizing dynamic gene interactions to reverse engineer gene regulatory networks using topological data analysis,' in '2017 21st International Conference Information Visualisation (IV),' ISSN 2375-0138, July 2017 pp. 384–389.

[37] Rizvi, A. H., Camara, P. G., Kandror, E. K., Roberts, T. J., Schieren, I., Maniatis, T., and Rabadan, R., 'Single-cell topological rna-seq analysis reveals insights into cellular differentiation and development,' Nature biotechnology, 2017, **35**(6), p. 551.

[38] Ma, X., Gao, L., and Tan, K., 'Modeling disease progression using dynamics of pathway connectivity,' Bioinformatics, 04 2014, **30**(16), pp. 2343–2350, ISSN 1367-4803.

[39] Xu, R. and Wunsch, D. C., 'Clustering algorithms in biomedical research: A review,' IEEE Reviews in Biomedical Engineering, 2010, **3**, pp. 120–154, ISSN 1941-1189.

[40] Xu, R. and Wunsch II, D. C., 'BARTMAP: A viable structure for biclustering,' Neural Networks, 2011, **24**(7), pp. 709–716.

[41] Tscherepanow, M., 'TopoART: A topology learning hierarchical ART network,' in 'International Conference on Artificial Neural Networks,' Springer, 2010 pp. 157–167.

[42] Chaibi, A., Lebbah, M., and Azzag, H., 'A new bi-clustering approach using topological maps,' in 'The 2013 International Joint Conference on Neural Networks (IJCNN),' IEEE, 2013 pp. 1–7.

[43] Sarazin, T., Lebbah, M., Azzag, H., and Chaibi, A., 'Feature group weighting and topological biclustering,' in 'International Conference on Neural Information Processing,' Springer, 2014 pp. 369–376.

[44] Carpenter, G. A. and Grossberg, S., 'A massively parallel architecture for a self-organizing neural pattern recognition machine,' Computer vision, graphics, and image processing, 1987, **37**(1), pp. 54–115.

[45] Kim, S., *Novel approaches to clustering , biclustering algorithms based on adaptive resonance theory and intelligent control*, Ph.D. thesis, Missouri University of Science and Technology, 2016.

[46] Elnabarawy, I., Wunsch, D. C., and Abdelbar, A. M., 'Biclustering ARTMAP collaborative filtering recommender system,' in '2016 International Joint Conference on Neural Networks (IJCNN),' ISSN 2161-4407, July 2016 pp. 2986–2991.

[47] Carpenter, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system,' Neural Networks, 1991, **4**(6), pp. 759 – 771, ISSN 0893-6080.

[48] Carpetner, G. A., Grossberg, S., and Rosen, D. B., 'Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns,' Technical report, Boston University Center for Adaptive Systems and Department of Cognitive . . . , 1991.

[49] Carpenter, G. A. and Grossberg, S., *Adaptive resonance theory*, Springer, 2010.

[50] Grossberg, S., 'Adaptive resonance theory: How a brain learns to consciously attend, learn, and recognize a changing world,' Neural Networks, 2013, **37**, pp. 1–47.

[51] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'A survey of adaptive resonance theory neural network models for engineering applications,' Neural Networks, 2019, **120**, pp. 167 – 203.

[52] Wunsch II, D. C., 'Admiring the Great Mountain: A Celebration Special Issue in Honor of Stephen Grossberg's 80th Birthday,' Neural Networks, 2019, **120**, pp. 1 – 4, special Issue in Honor of the 80th Birthday of Stephen Grossberg.

[53] Carpenter, G. A., Grossberg, S., and Reynolds, J. H., 'Artmap: supervised real-time learning and classification of nonstationary data by a self-organizing neural network,' in '[1991 Proceedings] IEEE Conference on Neural Networks for Ocean Engineering,' 1991 pp. 341–342.

[54] Carpenter, G. A., Grossberg, S., Markuzon, N., Reynolds, J. H., and Rosen, D. B., 'Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps,' IEEE Transactions on Neural Networks, Sep. 1992, **3**(5), pp. 698–713, ISSN 1941-0093.

[55] de Souto, M. C., Costa, I. G., de Araujo, D. S., Ludermir, T. B., and Schliep, A., 'Clustering cancer gene expression data: a comparative study,' BMC bioinformatics, 2008, **9**(1), p. 497.

[56] Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. D., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M., 'Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling,' Nature, Feb 2000, **403**(6769), pp. 503–511, ISSN 1476-4687, doi:10.1038/35000501.

[57] Wang, K., Wang, B., and Peng, L., 'CVAP: validation for cluster analyses,' Data Science Journal, 2009, pp. 0904220071–0904220071.

[58] Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J., 'Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia,' Nature Genetics, Jan 2002, **30**(1), pp. 41–47, ISSN 1546-1718, doi:10.1038/ng765.

[59] Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. W., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C., and Golub, T. R., 'Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning,' Nature Medicine, Jan 2002, **8**(1), pp. 68–74, ISSN 1546-170X, doi:10.1038/nm0102-68.

[60] Eiben, A. E. and Smith, J. E., *Introduction to Evolutionary Computing*, Springer Publishing Company, Incorporated, 2nd edition, 2015.

[61] Hubert, L. and Arabie, P., 'Comparing partitions,' J. Classification, 1985, **2**(1), pp. 193–218.

[62] Brito da Silva, L. E. and Wunsch II, D. C., 'A study on exploiting VAT to mitigate ordering effects in Fuzzy ART,' in 'Proc. IEEE International Joint Conference on Neural Networks (IJCNN),' 2018 pp. 2351–2358.

[63] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'Dual vigilance fuzzy adaptive resonance theory,' Neural Networks, 2019, **109**, pp. 1 – 5.

[64] Brito da Silva, L. E., Elnabarawy, I., and Wunsch II, D. C., 'Distributed dual vigilance fuzzy adaptive resonance theory learns online, retrieves arbitrarily-shaped clusters, and mitigates order dependence,' Neural Networks, 2020, **121**, pp. 208 – 228.

[65] O'brien, M., Costin, B., and Miles, M., 'Using genome-wide expression profiling to define gene networks relevant to the study of complex traits: from rna integrity to network topology,' in 'International review of neurobiology,' volume 104, pp. 91–133, Elsevier, 2012.

[66] Pontes, B., GirÃ¡ldez, R., and Aguilar-Ruiz, J. S., 'Biclustering on expression data: A review,' Journal of Biomedical Informatics, 2015, **57**, pp. 163 – 180, ISSN 1532-0464, doi:https://doi.org/10.1016/j.jbi.2015.06.028.

[67] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W., 'From molecular to modular cell biology,' Nature, Dec 1999, **402**(6761), pp. C47–C52, ISSN 1476-4687, doi:10.1038/35011540.

[68] Mitra, K., Carvunis, A.-R., Ramesh, S. K., and Ideker, T., 'Integrative approaches for finding modular structure in biological networks,' Nature Reviews Genetics, Oct 2013, **14**(10), pp. 719–732, ISSN 1471-0064, doi:10.1038/nrg3552.

[69] Singh, A. J., Ramsey, S. A., Filtz, T. M., and Kioussi, C., 'Differential gene regulatory networks in development and disease,' Cellular and Molecular Life Sciences, Mar 2018, **75**(6), pp. 1013–1025, ISSN 1420-9071, doi:10.1007/s00018-017-2679-6.

[70] Saelens, W., Cannoodt, R., and Saeys, Y., 'A comprehensive evaluation of module detection methods for gene expression data,' Nature Communications, Mar 2018, **9**(1), p. 1090, ISSN 2041-1723, doi:10.1038/s41467-018-03424-4.

[71] Snášel, V., Nowaková, J., Xhafa, F., and Barolli, L., 'Geometrical and topological approaches to big data,' Future Generation Computer Systems, 2017, **67**, pp. 286–296.

[72] Nielson, J. L., Paquette, J., Liu, A. W., Guandique, C. F., Tovar, C. A., Inoue, T., Irvine, K.-A., Gensel, J. C., Kloke, J., Petrossian, T. C., Lum, P. Y., Carlsson, G. E., Manley, G. T., Young, W., Beattie, M. S., Bresnahan, J. C., and Ferguson, A. R., 'Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury,' Nature Communications, Oct 2015, **6**(1), p. 8581, ISSN 2041-1723, doi:10.1038/ncomms9581.

[73] Yelugam, R., Brito da Silva, L. E., and Wunsch, D. C., 'Topobartmap: Biclustering artmap with or without topological methods in a blood cancer case study,' in '2020 International Joint Conference on Neural Networks (IJCNN),' 2020 pp. 1–8.

[74] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., and Sabeti, P. C., 'Detecting novel associations in large data sets,' Science, 2011, **334**(6062), pp. 1518–1524, ISSN 0036-8075, doi:10.1126/science.1205438.

[75] Malvia, S., Bagadi, S. A. R., Pradhan, D., Chintamani, C., Bhatnagar, A., Arora, D., Sarin, R., and Saxena, S., 'Study of gene expression profiles of breast cancers in indian women,' Scientific Reports, Jul 2019, **9**(1), p. 10018, ISSN 2045-2322, doi:10.1038/s41598-019-46261-1.

[76] Bartfai, G., 'Hierarchical clustering with ART neural networks,' in 'Proc. IEEE International Conference on Neural Networks (ICNN),' volume 2, Jun 1994 pp. 940–944.

[77] Brito da Silva, L. E., Rayapati, N., and Wunsch II, D. C., 'iCVI-ARTMAP: Accelerating and improving clustering using adaptive resonance theory predictive mapping and incremental cluster validity indices,' arXiv, 2020, arXiv:2008.09903v1 [cs.LG].

[78] Song, L., Langfelder, P., and Horvath, S., 'Comparison of co-expression measures: mutual information, correlation, and model based indices,' BMC Bioinformatics, Dec 2012, **13**(1), p. 328, ISSN 1471-2105, doi:10.1186/1471-2105-13-328.

[79] Bain, L. J. and Engelhardt, M., *Introduction to Probability and Mathematical Statistics*, Brooks/Cole, Cengage Learning, 2 edition, 1992.

[80] Jaskowiak, P. A., Campello, R. J. G. B., and Costa, I. G., 'Proximity measures for clustering gene expression microarray data: A validation methodology and a comparative analysis,' IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, **10**(4), pp. 845–857, doi:10.1109/TCBB.2013.9.

[81] Rodgers, J. L. and Nicewander, W. A., 'Thirteen ways to look at the correlation coefficient,' The American Statistician, 1988, **42**(1), pp. 59–66, doi:10.1080/00031305.1988.10475524.

[82] Weirauch, M. T., *Gene Coexpression Networks for the Analysis of DNA Microarray Data*, chapter 11, pp. 215–250, John Wiley & Sons, Ltd, ISBN 9783527638079, 2011, doi:https://doi.org/10.1002/9783527638079.ch11.

[83] Quackenbush, J., 'Computational analysis of microarray data,' Nature Reviews Genetics, Jun 2001, **2**(6), pp. 418–427, ISSN 1471-0064, doi:10.1038/35076576.

[84] Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M., 'Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses,' Proceedings of the National Academy of Sciences, 2001, **98**(24), pp. 13790–13795, ISSN 0027-8424, doi:10.1073/pnas.191502998.

[85] Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V., Hayward, N., and Trent, J., 'Molecular classification of cutaneous malignant melanoma by gene expression profiling,' Nature, Aug 2000, **406**(6795), pp. 536–540, ISSN 1476-4687, doi:10.1038/35020115.

[86] Bredel, M., Bredel, C., Juric, D., Harsh, G. R., Vogel, H., Recht, L. D., and Sikic, B. I., 'Functional network analysis reveals extended gliomagenesis pathway maps and three novel myc-interacting genes in human gliomas,' Cancer Research, 2005, **65**(19), pp. 8679–8689, ISSN 0008-5472, doi:10.1158/0008-5472.CAN-05-1204.

[87] Chen, X., Cheung, S. T., So, S., Fan, S. T., Barry, C., Higgins, J., Lai, K.-M., Ji, J., Dudoit, S., Ng, I. O. L., Van De Rijn, M., Botstein, D., and Brown, P. O., 'Gene expression patterns in human liver cancers,' Molecular biology of the cell, Jun 2002, **13**(6), pp. 1929–1939, ISSN 1059-1524, doi:10.1091/mbc.02-02-0023, 12058060[pmid].

[88] Chowdary, D., Lathrop, J., Skelton, J., Curtin, K., Briggs, T., Zhang, Y., Yu, J., Wang, Y., and Mazumder, A., 'Prognostic gene expression signatures can be measured in tissues collected in rnalater preservative,' The Journal of Molecular Diagnostics, 2006, **8**(1), pp. 31 – 39, ISSN 1525-1578, doi:https://doi.org/10.2353/jmoldx.2006.050056.

[89] Dyrskjã t, L., Thykjaer, T., KruhÃffer, M., Jensen, J. L., Marcussen, N., Hamilton-Dutoit, S., Wolf, H., and Ãrntoft, T. F., 'Identifying distinct classes of bladder carcinoma using microarrays,' Nature Genetics, Jan 2003, **33**(1), pp. 90–96, ISSN 1546-1718, doi:10.1038/ng1061.

[90] Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. I., Altman, R. B., Brown, P. O., Botstein, D., and Petersen, I., 'Diversity of gene expression in adenocarcinoma of the lung,' Proceedings of the National Academy of Sciences, 2001, **98**(24), pp. 13784–13789, ISSN 0027-8424, doi:10.1073/pnas.241500798.

[91] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S., 'Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring,' Science, 1999, **286**(5439), pp. 531–537, ISSN 0036-8075, doi:10.1126/science.286.5439.531.

[92] Gordon, G. J., Jensen, R. V., Hsiao, L.-L., Gullans, S. R., Blumenstock, J. E., Ramaswamy, S., Richards, W. G., Sugarbaker, D. J., and Bueno, R., 'Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma,' Cancer Research, 2002, **62**(17), pp. 4963–4967, ISSN 0008-5472.

[93] Khan, J., Wei, J. S., RingnÃ©r, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S., 'Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks,' Nature Medicine, Jun 2001, **7**(6), pp. 673–679, ISSN 1546-170X, doi:10.1038/89044.

[94] Laiho, P., Kokko, A., Vanharanta, S., Salovaara, R., Sammalkorpi, H., JÃrvinen, H., Mecklin, J.-P., Karttunen, T. J., Tuppurainen, K., Davalos, V., Schwartz, S., Arango, D., MÃkinen, M. J., and Aaltonen, L. A., 'Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis,' Oncogene, Jan 2007, **26**(2), pp. 312–320, ISSN 1476-5594, doi:10.1038/sj.onc.1209778.

[95] Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R., 'Gene expression profiling identifies clinically relevant subtypes of prostate cancer,' Proceedings of the National Academy of Sciences, 2004, **101**(3), pp. 811–816, ISSN 0027-8424, doi:10.1073/pnas.0304146101.

[96] Liang, Y., Diehn, M., Watson, N., Bollen, A. W., Aldape, K. D., Nicholas, M. K., Lamborn, K. R., Berger, M. S., Botstein, D., Brown, P. O., and Israel, M. A., 'Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme,' Proceedings of the National Academy of Sciences, 2005, **102**(16), pp. 5814–5819, ISSN 0027-8424, doi:10.1073/pnas.0402870102.

[97] Nutt, C. L., Mani, D. R., Betensky, R. A., Tamayo, P., Cairncross, J. G., Ladd, C., Pohl, U., Hartmann, C., McLaughlin, M. E., Batchelor, T. T., Black, P. M., von Deimling, A., Pomeroy, S. L., Golub, T. R., and Louis, D. N., 'Gene expression-based classification of malignant gliomas correlates better with survival than histological classification,' Cancer Research, 2003, **63**(7), pp. 1602–1607, ISSN 0008-5472.

[98] Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y. H., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D. N., Mesirov, J. P., Lander, E. S., and Golub, T. R., 'Prediction of central nervous system embryonal tumour outcome based on gene expression,' Nature, Jan 2002, **415**(6870), pp. 436–442, ISSN 1476-4687, doi:10.1038/415436a.

[99] Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.-H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S., and Golub, T. R., 'Multiclass cancer diagnosis using tumor gene expression signatures,' Proceedings of the National Academy of Sciences, 2001, **98**(26), pp. 15149–15154, ISSN 0027-8424, doi:10.1073/pnas.211566398.

[100] Risinger, J. I., Maxwell, G. L., Chandramouli, G. V. R., Jazaeri, A., Aprelikova, O., Patterson, T., Berchuck, A., and Barrett, J. C., 'Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer,' Cancer Research, 2003, **63**(1), pp. 6–11, ISSN 0008-5472.

[101] Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R., 'Gene expression correlates of clinical prostate cancer behavior,' Cancer Cell, 2002, **1**(2), pp. 203 – 209, ISSN 1535-6108, doi:https://doi.org/10.1016/S1535-6108(02)00030-2.

[102] Su, A. I., Welsh, J. B., Sapinoso, L. M., Kern, S. G., Dimitrov, P., Lapp, H., Schultz, P. G., Powell, S. M., Moskaluk, C. A., Frierson, H. F., and Hampton, G. M., 'Molecular classification of human carcinomas by use of gene expression signatures,' Cancer Research, 2001, **61**(20), pp. 7388–7393, ISSN 0008-5472.

[103] West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Marks, J. R., and Nevins, J. R., 'Predicting the clinical status of human breast cancer by using gene expression profiles,' Proceedings of the National Academy of Sciences, 2001, **98**(20), pp. 11462–11467, ISSN 0027-8424, doi:10.1073/pnas.201162998.

[104] Yeoh, E.-J., Ross, M. E., Shurtleff, S. A., Williams, W., Patel, D., Mahfouz, R., Behm, F. G., Raimondi, S. C., Relling, M. V., Patel, A., Cheng, C., Campana, D., Wilkins, D., Zhou, X., Li, J., Liu, H., Pui, C.-H., Evans, W. E., Naeve, C., Wong, L., and Downing, J. R., 'Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling,' Cancer Cell, 2002, **1**(2), pp. 133 – 143, ISSN 1535-6108, doi:https://doi.org/10.1016/S1535-6108(02)00032-6.

[105] Tomlins, S. A., Mehra, R., Rhodes, D. R., Cao, X., Wang, L., Dhanasekaran, S. M., Kalyana-Sundaram, S., Wei, J. T., Rubin, M. A., Pienta, K. J., Shah, R. B., and Chinnaiyan, A. M., 'Integrative molecular concept modeling of prostate cancer progression,' Nature Genetics, Jan 2007, **39**(1), pp. 41–51, ISSN 1546-1718, doi: 10.1038/ng1935.

[106] Moshtaghi, M., Bezdek, J. C., Erfani, S. M., Leckie, C., and Bailey, J., 'Online cluster validity indices for performance monitoring of streaming data clustering,' International Journal of Intelligent Systems, 2019, **34**(4), pp. 541–563.

[107] Ibrahim, O. A., Keller, J. M., and Bezdek, J. C., 'Evaluating Evolving Structure in Streaming Data With Modified Dunn's Indices,' IEEE Transactions on Emerging Topics in Computational Intelligence, 2019, pp. 1–12, ISSN 2471-285X.

[108] Brito Da Silva, L. E., Melton, N. M., and Wunsch II, D. C., 'Incremental Cluster Validity Indices for Online Learning of Hard Partitions: Extensions and Comparative Study,' IEEE Access, 2020, **8**, pp. 22025–22047, doi:10.1109/ACCESS.2020.2969849.

[109] Brouwer, R. K., 'Extending the rand, adjusted rand and jaccard indices to fuzzy partitions,' Journal of Intelligent Information Systems, 2009, **32**(3), pp. 213–235, ISSN 1573-7675, doi:10.1007/s10844-008-0054-7.

[110] Horta, D. and Campello, R. J., 'Comparing hard and overlapping clusterings,' Journal of Machine Learning Research, 2015, **16**(93), pp. 2949–2997.

[111] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., 'Scikit-learn: Machine learning in Python,' Journal of Machine Learning Research, 2011, **12**, pp. 2825–2830.

[112] Albanese, D., Filosi, M., Visintainer, R., Riccadonna, S., Jurman, G., and Furlanello, C., 'minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers,' Bioinformatics, 12 2012, **29**(3), pp. 407–408, ISSN 1367-4803, doi:10.1093/bioinformatics/bts707.

[113] Calvo, B. and Santafé Rodrigo, G., 'scmamp: Statistical comparison of multiple algorithms in multiple problems,' The R Journal, Vol. 8/1, Aug. 2016, 2016.

[114] Demšar, J., 'Statistical comparisons of classifiers over multiple data sets,' Journal of Machine Learning Research, 2006, **7**(1), pp. 1–30.

[115] Reshef, Y. A., Reshef, D. N., Finucane, H. K., Sabeti, P. C., and Mitzenmacher, M., 'Measuring dependence powerfully and equitably,' The Journal of Machine Learning Research, 2016, **17**(1), p. 7406–7468, ISSN 1532-4435.

[116] Padilha, V. A. and de Carvalho, A. C. P. L. F., 'A study of biclustering coherence measures for gene expression data,' in '2018 7th Brazilian Conference on Intelligent Systems (BRACIS),' 2018 pp. 546–551, doi:10.1109/BRACIS.2018.00100.

[117] van der Maaten, L. and Hinton, G., 'Visualizing data using t-sne,' Journal of Machine Learning Research, 2008, **9**(86), pp. 2579–2605.

VITA

Raghu Yelugam is a graduate student at Missouri University of Science and Technology (Missouri S&T). He received his Bachelor's degree in Electrical Engineering from the National Institute of Technology - Warangal, India, in 2014. Upon graduation, he joined Reliance Industries Limited and worked at Jamnagar, India, as a Graduate Engineering Trainee till July 2015. He got involved with Reinforcement learning during an internship at Swayatt Robots, Bhopal - India, in August 2016. His interest in the subject motivated him to pursue a Master's degree at Missouri S&T. At Missouri S&T, he joined the RDDC laboratory led by Dr. Hamidreza Modres in 2017. Eventually, he joined the Applied Computational Intelligence Laboratory led by Dr. Donald C Wunsch II in 2018. His research interests included Adaptive Resonance Theory, Topology, and Model-free Reinforcement Learning. Raghu received his master's in Computer Engineering in December 2021 under the supervision of Dr. Donald C Wunsch II.