Clustering MBTI Personalities With Graph Filters And Self Organizing Maps Over Pinecone

Georgios Drakopoulos
Department of Informatics
Ionian University
0000-0002-0975-1877

Phivos Mylonas
ICE Department
University of West Attica
0000-0002-6916-3129

***Abstract*—Self organizing maps (SOMs) or cognitive maps are designed to preserve major topological attributes of manifolds in a higher dimensionality data space to corresponding projections thereof in a low dimensional coordinate space. This is performed by mapping neighborhoods and the distances contained therein from the data space to ones in the coordinate space. Thus, SOM functionality relies heavily on the geometrical properties of both spaces. Topologically flexible data space distance metrics are constructed by combining tensors with graph filters, the latter coming from graph signal processing. The power of these distance metrics comes from naturally expressing the higher order relationships between points of the data space. This paves the way for addressing engineering scenarios involving a large number of densely interrelated attributes. One such case is discerning personalities from textual information based on the Myers-Briggs taxonomy indicator (MBTI), a framework of archetypal personalities derived from Jungian psychodynamic theory. Various graph filters were tested on a benchmark Kaggle dataset with ground truth with comparisons assessed in terms of topological error, cluster purity, average inter-cluster distance, and cluster curvature variability. Data points were stored in Pinecone, a recent vector database, with Python integration.**

***Index Terms*—SOMs, graph filters, graph signal processing, graph topology, tensor distance metrics, geometric analytics, cluster curvature, dimensionality reduction, MBTI, Pinecone**

I. Introduction

Self organizing maps (SOMs) or cognitive maps are unsupervised neural network architectures designed to approximate a manifold in a high dimensional data space $\mathcal{V}$ by constructing an approximately continuous projection thereof, a *topological map*, in a lower dimensional coordinate space $\mathcal{C}$ [1]. Geometry is crucial in shaping said projection as the latter preserves up to a point the primary topological properties of the original manifold [2]. Dimensionality reduction is thus achieved, which may be an intermediate step for clustering or visualization.

Human personality and the evolving dynamics underpinning their interplay have been for quite a number of years central research points across domains such as psychology [3], social network analysis [4], online marketing and digital campaign design [5], computational neurosciences [6], and sociology [7]. Gaining insight into team dynamics allows informed decisions in tasks so diverse as workgroup formulation, mentor matching, designing unobviated and unobtrusive communication, and task delegation. The basic building block of any such approach is a methodological and systematic framework for characterizing human personalities based on fundamental psychological traits [8]. One such model is the Myers-Briggs Type Indicator (MBTI) which is derived from the Jungian psychoanalytical framework [9]. The latter is a personality taxonomy based on four independent variables corresponding to cognitive functions including information collection and decision making in the face of incomplete data.

TABLE I
Notation Synopsis.

Symbol	Meaning	First in
$\triangleq$	Definition or equality by definition	Eq. (1)
$\{s_1, \ldots, s_n\}$	Set with elements $s_1, \ldots, s_n$	Eq. (1)
$\lvert\cdot\rvert$	Set cardinality	Eq. (14)
$\mathrm{loc}\,(\cdot)$	Location function for data points	Eq. (1)
$\mathrm{invloc}\,(\cdot)$	Inverse location relationship for neurons	Eq. (1)
$\mathrm{weight}\,(u)$	Synaptic weights of neuron u	Eq. (3)
$\Gamma\,(u)$	Neighborhood of neuron u	Eq. (2)
$\Delta\,(u;\eta_0)$	Cover of neuron u with threshold η_0	Eq. (2)
$\star$	Application of spatial filter to graph	Eq. (7)
$\lVert\cdot\rVert_F$	Frobenius tensor norm	Eq. (7)
$f^{(k)}\,(x)$	k-th derivative of function f	Eq. (23)
$\mathrm{prob}\,\{\Omega\}$	Probability of event Ω occurring	Eq. (16)

The primary research objective of this conference paper is the description of an SOM architecture with a graph filter for clustering personalities of the MBTI taxonomy using linguistic attributes extracted from written text. As a secondary objective the role of geometry to the overall learning process of the SOMs is explored through the proposal of the cluster curvature variability metric. Additionally, the capabilities of Pinecone[1], a new vector database, driven over Python are explored.

The remainder of this conference paper is structured as follows. In section II the recent scientific literature regarding tensor metrics and SOMs is reviewed. In section III the proposed methodology is described, while in section IV the results obtained are discussed. Possible research directions are given in section V. Bold capital letters denote tensors and matrices, boldface lowercase vectors, and lowercase scalars. Technical acronyms are explained the first time they appear in text. Finally, the notation is summarized in table I.

978-8-xxxx-xxxx-xx/24/$31.00 ©2024 IEEE

[1]https://www.pinecone.io

II. Previous Work

As stated earlier SOMs are unsupervised neural network architectures [10] explicitly designed to perform dimensionality reduction by building a projection in a low dimensional space of a manifold from a high dimensional one [11]. Since topological attributes like distances and neighborhoods are maintained in the projection [12], the latter maintains a degree of fidelity to the original manifold [13]. A general theory for topological maps is described in [14]. The potential of message passing mechanism similar to that of graph convolutional neural networks (GCNNs) is explored in [15]. Other architectures which have been applied to humanistic data processing include tensor stack networks (TSNs) for social graph decompression [16], multi-view clustering [17], and large language models (LLMs) [18] as well as graph neural networks (GNNs) for affective social network analysis [19], few-shot image classification [20], and audio tagging [21], and Twitter community discovery [22]. Graph signal processing (GSP) is an emerging field with applications to matrix approximation [23], biosignal analysis [24], and graph filters [25]. A space efficient data structure for evolving graphs is proposed in [26]. It is conjectured that hippocampus constructs a low dimensional representation of the physical world in a similar way to the SOM training process [27] [28].

The MBTI framework defines a total of sixteen distinct personalities [29]. Such taxonomies constitute a considerable improvement compared to the various emotion models such as those proposed by Plutchik [30] or Eckman [31] since they go far beyond individual reponses elicited by specific stimuli [32]. MBTI has been applied among others to teaching [8] and online social networks [33]. Moreover, it has been used in conjunction with signal processing methodologies [34]. Human emotion estimators include gait [35], face [36], text attributes [37] [38], and more recently indicators defined on a broad spectrum of cognitive tasks [39] [40].

III. Proposed Methodology

A. Approach Overview

The algorithmic approach proposed here consists of the following steps. These can be altered depending on the underlying domain, forming thus the basis of a GSP framework.

- The dataset, consisting of data vectors along tagged with the respective real class, is stored in a Pinecone instance.
- The vector set $V \subseteq \mathcal{V}$ is clustered with an SOM using a distance metric based on graph filters and tensor metrics.
- Clustering is evaluated over the $|V|$ vectors through a combination of algebraic and geometric indicators.

The following subsections outline the above steps in detail.

B. MBTI Taxonomy

The MBTI personality taxonomy [9] is a framework for classifying personalities and it a tool for a wide range of tasks including team building, negotation, and conflict resolution. Perhaps the most tangible and recognazible result of MBTI is the set of sixteen archetypes resulting from the combinations of the four binary variables shown in table II. Each of these fundamental personalities is symbolized with a unique four letter acronym[2] denoting their cognitive functionality. The latter includes the balance between thought and action, the type of information considered, the objectives prioritized, and the general decision making process. For example, ISTJ stands for Introversion, Sensing, Thinking, Judging and similarly ENFP means Extraversion, Intuition, Feeling, and Perceiving.

TABLE II
MBTI Variables [9].

Variable	Meaning
I vs E	Response to events and interactionb with the world
Introvert	Mental work and introspection, indirect cues
Extrovert	Frequent and open communication, direct feedback
S vs N	Information collection and gathering
Sensing	Emphasis on current state of affairs
Intuition	Highlights on possible evolutionary paths
T vs F	Decision making process pylons
Thinking	Impersonal and objective indicators towards results
Feeling	Emotional taking into account consequences on people
P vs J	Strategy for dealing with the outer world
Perceiving	Flexible, adjustive, and open ended
Judging	Structured through a chain of smaller decisions

Each individual contributes in their unique way to a team or workgroup. For instance, S-type individuals relying on sensory observe the outside world and can report body language nuances or group collective dynamics to N-type individuals who can understand their meaning. Furthermore, J-type persons can implement abstract guidelines set forth by P-type individuals.

TABLE III
MBTI Personality Probabilites ($p_0 = 1/16$).

MBTI	ISFJ	ESFJ	ISTJ	INTJ	ENTJ	INFJ
Prob	$2.19p_0$	$1.92p_0$	$1.86p_0$	$0.33p_0$	$0.29p_0$	$0.25p_0$

Substantial evidence corroborates that the distriubtion of the MBTI personalities is not uniform. Assuming an average uniform probability of p_0, the three most common MBTI personalities are ISFJ, ESFJ, and ISTJ, while the three rarest ones are INTJ, ENTJ, and INFJ. The approximate probabbilities for these personalities as a fraction of p_0 are given in table III [8]. Therefore, ISFJ is roughly ten times more probable than INFJ. There have been many explanations in the scientific literature about this particular gap as well as about the character variation in general including social, educational, geographic, and neurobiological factors [9].

C. Self Organizing Maps

SOMs operate by progressively connecting points from the following two distinct spaces of different dimensionality and semantics through learning based on a modified Hebbian learning rule. The framework is shown in algorithm 1. Additionally, the projection from $\mathcal{V}$ to $\mathcal{C}$ is shown conceptually in figure 1.

- **Data space:** The high dimensional space $\mathcal{V}$ contains the manifold to be approximated. Data points are selected from there based on a selection policy.

[2]www.16personalities.com

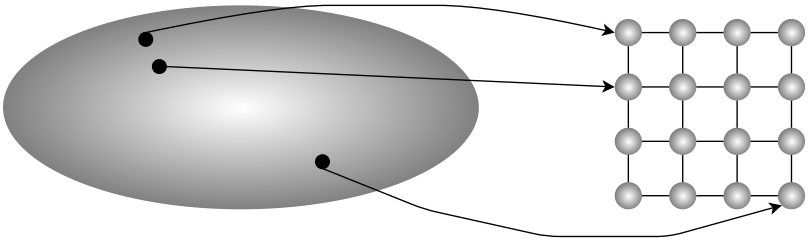

Fig. 1. SOM projection from data space to a two dimensional coordinate space.

- **Coordinate space:** The low dimensional space $\mathcal{C}$ is the grid where each point is a neuron with adjustable weights.

Algorithm 1 SOM training.

Require: SOM parameters as mentioned here
Ensure: SOM is smooth and partially preserves topology
1: **repeat**
2: **for all** data points $\mathbf{v}_j \in V$ **do**
3: select a data point $\mathbf{v}_j$ according to selection policy
4: find the winning neuron u^* as in (1) using (7)
5: update $\text{weight}\,(u)\,, u \in \Delta\,(u^*; \eta_0)$ as in (3)
6: **end for**
7: **until** SOM training has converged
8: **return** cognitive map

The SOM training process is as follows. First, let an *epoch* be defined as the time steps necessary for the enire data point set to be used once to train SOM. During a given epoch each point $\mathbf{v}_j$, selected at random exactly once, is assigned to the neuron u^* with the closest synaptic weights as determined by a metric distance $g\,(\cdot,\cdot)$. Subsequently u^*, termed the *winning neuron*, and the neurons in its cover set $\Delta\,(u^*; \eta_0)$ are rewarded in a way that reinforces the connection between $\mathbf{v}_j$ and them through their respective synaptic weights. Rewarding the cover set neurons besides u^* is central to smoothness of the topological map, as updating only the winning neuron or a narrow area around it leads to sparse and discontinuous maps.

The low dimensional projection of the original manifold is constructed in the coordinate space $\mathcal{C}$. This is accomplished by mapping through a distance metric data points from $\mathcal{V}$. The location $\text{loc}\,(u)$ of a neuron u is the vector containing its coordinates in the grid, with the number of components being the dimension of $\mathcal{C}$, whereas the $\text{weight}\,(u)$ is a vector in $\mathcal{V}$. Along a similar line of reasoning $\text{invloc}\,(u)$ is the set of points mapped to u determined by distance metric $g\,(\cdot,\cdot)$.

$$\text{invloc}\,(u) \triangleq \{\mathbf{v}_j \mid u = \text{argmin}\,[g\,(\mathbf{v}_j, \text{weight}\,(u_k))]\} \quad (1)$$

Two important sets for each neuron u are its neighborhood $\Gamma\,(u)$ and its cover set $\Delta\,(u; \eta_0)$. The former is dervied directly from the physical neuron placement, whereas the latter depends on the selection of proximity function $h\,(\cdot,\cdot)$ and a threhold η_0. Both are defined in equation (2).

$$\Gamma\,(u) \triangleq \{u' \mid u, u' \text{ are adjacent}\}$$

$$\Delta\,(u; \eta_0) \triangleq \{u' \mid h\,(u', u) \;\geq\; \eta_0\} \quad (2)$$

Once a neuron is selected during an epoch, it becomes the center of a cluster with the corresponding cover set determining its periphery. When the cardinality of $\text{invloc}\,(u)$ of a neuron u is zero after one or even more epochs and additionally u does not belong to a cluster periphery, then either u should be assigned to the closest cluster or its synaptic weights should be recalibrated based on cluster centers. For large data sets cardinality estimators [41] may be used.

The reward for the synaptic weights is given in equations (3) and (4) for every neuron u in the cover set $\Delta\,(u^*; \eta_0)$ of the cluster center u^*. What differentiates the cluster center from its periphery is the weight function $w\,(\cdot,\cdot)$, which is maximized at the center and it is usually symmetric with respect to it.

$$\text{weight}\,(u)^{[n]} \triangleq \text{weight}\,(u)^{[n-1]} + \delta\,(u)^{[n]} \quad (3)$$

The correction factor is computed as in equation (4). Observe that it depends on attributes taken both from $\mathcal{V}$ and $\mathcal{C}$.

$$\delta\,(u)^{[n]} \triangleq \mu^{[n]} w\,(u, u^*)\left(\text{weight}\,(u^*)^{[n-1]} - \mathbf{v}_j\right) \quad (4)$$

In (4) the learning rate during epoch m is as given in (5). Therein n_0 is a prespecified hyperparameter. The advantages of the specific selection is that it is smooth and it decays with approximately a quadratic rate, allowing thus an initial cluster formation and additionally a mild correction thereof in subsequent epochs. Thus, clusters are progressively built.

$$\mu^{[n]} \triangleq \begin{cases} \cos\left(\dfrac{\pi n}{2\,(1+n_0)}\right) & 0 \;\leq\; n \;\leq\; n_0 \\ \cos\left(\dfrac{\pi n_0}{2\,(1+n_0)}\right), & n > n_0 \end{cases} \quad (5)$$

The abovementioned smooth behavior of the cosine learning rate can be explained by the second order Taylor expansion of the cosine function as shown in equation (6) below.

$$\cos\vartheta \;=\; \sum_{k=0}^{+\infty} (-1)^k \frac{\vartheta^{2k}}{(2\vartheta)!} \;\approx\; 1 - \frac{\vartheta^2}{2} \quad (6)$$

Since the cosine function is analytic, like most trigonometric functions, it allows higher order polynomial approximations with more terms and higher accuracy. However, since these terms vanish more quickly than the quadratic term, they have little effect on it. Moreover, because of their alternating signs, they cancel each other up to an extent as ϑ grows.

D. Linguistic Attributes

The linguistic attributes used here are shown in table IV and they are related to the four MBTI variables of table II. Said attributes describe how cognitive functions are reflected in text and were extracted from the Kaggle *MBTI dataset*[3] which has a usability score of 8.82 with 93 distinct code segments developed over it. Since it contains ground truth, it allows for elaborate subsequent analysis. The natural language features pertaining to sentence and word length are related to the tendency for extroversion [32]. Moreover, the emotional terms and the punctuation tend to reveal the structure of the human thought process at a subconscious level [34].

In the majority of existing approaches the distance metric $g(\cdot,\cdot)$ in $\mathcal{V}$ would rely on stacking the attributes of table IV to a single vector. However, such a scheme is oblivious to the fact that the features come from two distinct categories. Moreover, any connections between features of the same category or across categories are ignored. Furthermore, any weight matrices would be big. Instead, more information can be extracted through an *attribute graph* where the location of each available feature and its spatial replationships also play a role besides its numerical value. This is the case of the graph shown in figure 2 where the structural features s_i, the functional ones f_i, and the references r_i are placed in three separate lines interconnected as shown. This allows for higher order patterns to emerge. The attributes themselves were selected based on recommendations from [34] and [35], whereas their respective location on the attribute graph was determined primarily by the semantics of each feature. The form of the proposed graph is certainly not unique and more complex graphs can express higher order connections between features. Nevertheless, it clearly illustrates the advantages of the proposed strategy as well as the role of topology therein.

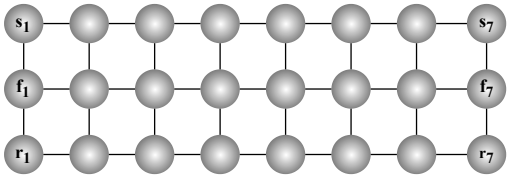

Fig. 2. Proposed attribute graph.

Observe that the attribute graph is regular with a rectangular configuration, which allows the application of image processing filters. The attribute graph has the following advantages:

- Compared to vector encoding, connectivity patterns such as degrees offer more information about the interdependencies between attributes.
- Existing knowledge or constraints about the underlying field can be incorporated into the graph with schemes similar to domain decomposition.
- Geometrical insight can yield distance metrics tailored for the data at hand, which can shed light into the true cluster structure, and graph filters can be applied.
- In clustering applications data points and clusters themselves can be represented with mutiple attribute sets, each with its own strengths.

Tensors are the primary algebraic tools for expressing graph topology and additionally they naturally handle matrices as data points. Moreover, tensor operations allow the modeling of graph filter operations such as spatial convolution. The latter is important as the graph of figure 2 can be regarded as a single data point which algebraically takes the form of a matrix with additional interdependencies between its entries. Equation (7) is the proposed tensor distance metric in unrolled form where I_c is the number of attribute categories, in this case three, and I_a is the maximum number of features per category, seven in this case. This is facilitated by both the rectangular form of the attribute graph and the equal number of features in either category. The $I_c \times I_a$ matrices $\mathbf{X}$ and $\mathbf{Y}$ encode directly the values of each attribute, normalized from 0 to 1.

$$g(\mathbf{X},\mathbf{Y}) \triangleq \frac{1}{I_c I_a} \left\| \sum_{i_a=1}^{I_a} \sum_{i_c=1}^{I_c} \mathbf{H}[i_a, i_c] \star (\mathbf{X}-\mathbf{Y}) \right\|_F \tag{7}$$

The Frobenius tensor norm of equation (7) is defined for a q-dimensional tensor $\mathbf{M}$ of dimensions $I_1 \times I_2 \times \ldots \times I_q$ as in equation (8). It is a common norm which is easy to implement in parallel or distributed computational environments and expresses the mean energy of the elements of the tensor.

$$\|\mathbf{M}\|_F \triangleq \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \ldots \sum_{i_q=1}^{I_q} |\mathbf{M}[i_1, i_2, \ldots, i_q]|^2} \tag{8}$$

The graph filter matrix $\mathbf{H}[i,j]$ is a 3×3 symmetric mask centered at graph coordinates (i,j). After the application of the graph filter at each point is obtained a weighted sum of the nearby differences akin to the way a two-dimensional filter is applied on an image. They take the elementwise forms of equations (9), (10), (11) and (12), and (13).

The Gaussian kernel of equation (9) is a common choice as it has many properties. Its decay allows the defintion of clusters with a reasonable number of points, whereas the boundary between two Gaussian kernels is linerly separable. Moreover, the Gaussian function is the eigenfunction of the Fourier operator, meaning that it can be also applied in the frequency domain unaleterd. Additionally, the Gaussian distribution has the maximum differential entropy among all continuous probability density functions of the same variance, meaning it can explain the broadest spectrum of probabilistic scenaria under the aforementioned variance constraint.

$$\mathbf{H}_g[i,j;\sigma_0] \triangleq \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{(i-j)^2}{2\sigma_0^2}\right) \tag{9}$$

The logistic kernel of equation (10) is frequently used in ML problems as it describes the evolution of a population over time. Moreover, its saturation property, monotonicity change,

[3]https://www.kaggle.com/datasets/datasnaek/mbti-type

TABLE IV
CHARACTER ATTRIBUTES.

Name	Structure	Name	Function	Name	References
s_1	Total words	f_1	Positive terms	r_1	To self
s_2	Total characters	f_2	Negative terms	r_2	To persons
s_3	Words per sentence	f_3	Neutral terms	r_3	Other
s_4	Punctuation marks	f_4	Adjectives	r_4	Judging terms
s_5	Question marks	f_5	Nouns	r_5	Sensing terms
s_6	Exclamation points	f_6	Adverbs	r_6	Emotional terms
s_7	Two or more '.'	f_7	Connectors	r_7	Rational terms

and decay rates allow for the construction of smooth clusters of shapes controlled by two hyperparameters.

$$\mathbf{H}_l\left[i,j;\gamma_0,\gamma_1\right] \triangleq \frac{1+\gamma_1}{1+\gamma_1\exp\left(-\gamma_0\left|i-j\right|\right)} \tag{10}$$

The p-th degree Chebyshev polynomial of the first kind of equation (11) has a multitude of numerical and algorithmic properties as well as for being extensively employed in signal processing and ML applications. Only the degree of these polynomials needs to be defined, as its coefficients follow automatically from either (11) or (12).

$$\mathbf{H}_c\left[i,j;p\right] \triangleq \frac{2}{1+\cos\left(p\arccos\left(\frac{\left|i-j\right|}{2\pi}\right)\right)} \tag{11}$$

An alternative definition of Type I Chebyshev polynomials to the trigonometric one of equation (11) is the second order recursion shown in equation (12). The latter allows for algebraic understanding of this class of polynomials

$$T_p\left(x\right) \triangleq \begin{cases} 2xT_{p-1}\left(x\right)-T_{p-2}\left(x\right), p\geq 2 \\ x, p=1 \\ 1, p=0 \end{cases} \tag{12}$$

The Cauchy kernel of (13) is the frequency response of a first order Butterworth filter in the frequency domain and is extensively used in many ML applications like the t-SNE. Its positive hyperparameter β_0 is the relative weight of the squared difference term, essentially defining a filtering radius.

$$\mathbf{H}_y\left[i,j;\beta_0\right] \triangleq \frac{1}{1+\beta_0\left(i-j\right)^2} \tag{13}$$

From the above descriptions it follows that they have smooth and differentialbe shape which can be controlled by a small number of hyperparameters, which can be determine in some cases by heuristics. This is paramount for clustering.

IV. RESULTS

In table V the results for each of the seven test cases are shown. Specifically for each graph filter it has been recorded the number of epochs necessary for the SOM to converge according to the criterion of definition 1. The metrics evaluation are explained throughout this section. Each column has been normalized with respect to its maximum except the power law exponents in order to provide better insight.

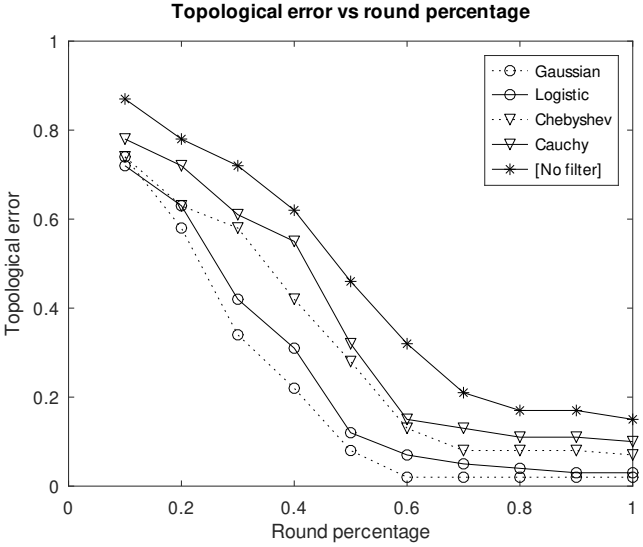

Fig. 3. Topological error vs round percentage.

The number of iterations requires the definition of a convergence criterion. One way to detect convergence is definition 1, assuming the cluster centroids do not cycle.

Definition 1 (SOM convergence): Training can stop when the centeroids have not moved for three successive epochs.

In section is assumed that SOM training results in a cluster set C. The latter consists of $|C|$ clusters in total as in (14).

$$C \triangleq \left\{c_1,\ldots,c_{|C|}\right\} \tag{14}$$

The topological error has been designed for SOMs and relies heavily on their geometrical properties as definition 2 shows.

Definition 2 (Topological error): During epoch n the topological error is the fraction of data vectors assigned neither to a cluster center nor to its neighboring neurons. Thus, they are assigned to cluster periphery. Formally, this is shown in (15):

$$e\left[n\right] \triangleq \frac{\left|\left\{\mathbf{v}_j\in V \mid \mathrm{loc}\left(\mathbf{v}_j\right)\notin\left\{u^*\cup\Gamma\left(u^*\right)\right\}\right\}\right|}{|V|} \tag{15}$$

Another clustering metric is cluster impurity, defined as the average number of data points assigned to the right cluster. Its computation implies knowledge of the ground truth.

Definition 3 (Cluster impurity): The cluster impurity Ω for $|C|$ clusters is the number of data points incorrectly assigned

TABLE V
GRAPH FILTER RESULTS (NORMALIZED).

Graph filter	Epochs	Topological	Impurity	Avg.	Max	Exponent*	Curvature
No graph filter	162	0.1533	0.3319	0.1655	0.1833	1.9982	0.2141
Gaussian $\left(\sigma_0^2 = 1\right)$	109	0.0411	0.1714	0.2475	0.2682	2.6711	0.4315
Gaussian $\left(\sigma_0^2 = 0.5\right)$	117	0.0733	0.2140	0.2333	0.2561	2.5333	0.4122
Logistic $(\gamma_0 = 1)$	128	0.1133	0.2355	0.2267	0.2399	2.2161	0.3618
Logistic $(\gamma_0 = 0.5)$	141	0.1267	0.2412	0.2271	0.2475	2.1967	0.3881
Chebyshev I $(n = 1)$	152	0.1392	0.2619	0.2222	0.2321	2.1166	0.3717
Chebyshev I $(n = 2)$	121	0.1033	0.2233	0.2282	0.2542	2.3333	0.3925
Chebyshev I $(n = 3)$	144	0.1285	0.2416	0.2267	0.2418	2.2175	0.3867
Cauchy $(\beta_0 = 0.5)$	127	0.1313	0.2391	0.2041	0.2350	2.1888	0.3612
Cauchy $(\beta_0 = 1)$	137	0.1366	0.2549	0.1773	0.2133	2.1331	0.3450

* Not normalized

to a cluster c regardless of their true cluster. It approximates the misclassification probability as shown in (16).

$$\begin{aligned}\Omega &\triangleq \frac{1}{|C|}\sum_{c\in C}\frac{|\mathrm{loc}\left(\mathbf{v}\right)=c\wedge\mathbf{v}\notin c|}{|c|}\\ &\approx \mathrm{prob}\left\{\mathrm{loc}\left(\mathbf{v}\right)=c\wedge\mathbf{v}\notin c\right\}\end{aligned} \tag{16}$$

A more general approach is that of the average inter-cluster distance which requires only knowledge of the final result. Although it is not directly interpretable, comparing its values for different clustering schemes may well yield useful insight.

Definition 4 (Average inter-cluster distance): The average inter-cluster quality $\bar{d}$ for p clusters is the pairwise distance between clusters. The latter is computed as the average distance between every possible pair of data points belonging to either cluster. It is a measure of how discernible clusters are.

$$\bar{d} \triangleq \frac{1}{\binom{|C|}{2}}\sum_{(c,c')}\sum_{(\mathbf{v}\in c,\mathbf{v}'\in c')}\frac{g\left(\mathbf{v},\mathbf{v}'\right)}{|c|\,|c'|} \tag{17}$$

Along a similar line of reasoning, the average maximum inter-cluster distance of definition 5 is the averaged maximum distance taken over all points taken from a cluster pair.

Definition 5 (Max inter-cluster distance): The average maximum intercluster distance is the maximum distance taken over any two cluster data points and over all cluster pairs.

$$d \triangleq \frac{1}{\binom{|C|}{2}}\sum\max_{(\mathbf{v}\in c,\mathbf{v}'\in c')}\left[g\left(\mathbf{v},\mathbf{v}'\right)\right] \tag{18}$$

Computing the decay exponent as given in definition 6 gives insight into the velocity of the convergence in terms of topological error reduction. This requires a convergence model, in this case a power law. Such models are closely associated with learning on graphs and with human activity.

Definition 6 (Decay exponent): The power law model for the reduction of the topological error is given in (19). The exponent γ_0 determines the convergence rate in epoch n.

$$v\left[n\right] \triangleq \alpha_0 n^{\gamma_0} \tag{19}$$

One way to compute the exponent γ_0 of velocity model (19) is to linearize it as in equation (20) and fit a least squares model to it with α_0 as a nuisance parameter.

$$\ln v\left[n\right] = \ln\alpha_0 + \gamma_0 \ln n \tag{20}$$

The last metric is derived from the geometric properties of SOMs as it is based on averaging the curvature variance of the cluster boundaries, essentially evaluating cluster flexibility. First, for a given cluster the variance of the curvature of each peripheral point and then these variances are averaged.

Definition 7 (Cluster curvature variance): The average curvature variance for a cluster set C is the arithmetic mean over the number of clusters of the curvature variance of the points at the boundary of each cluster as shown in (21).

$$\begin{aligned}K &\triangleq \frac{1}{|C|}\sum_{c\in C}K_1\left(c\right)\\ K_1\left(c\right) &\triangleq \frac{1}{|\Pi\left(c\right)|-1}\sqrt{\sum_{u\in\Pi(c)}\left(\omega\left(u\right)-K_2\left(c\right)\right)^2}\\ K_2\left(c\right) &\triangleq \frac{1}{|\Pi\left(c\right)|}\sum_{u\in\Pi(c)}\omega\left(u\right)\end{aligned} \tag{21}$$

The periphery $\Pi\left(c\right)$ of a cluster c is defined as the set of its points with at least one neighbor assigned to another cluster. In equation (21) the last two branches compute the deterministic variance of $K\left(c\right)$ for a cluster c, whereas the first branch averages them over every cluster in cluster set C. The local approximation of curvature at a neuron u in the boundary of a cluster c is given in (22). It is a measure of elasticity at the location of u, which is a building block of cluster flexibility.

$$\omega\left(u\right) \triangleq \frac{\omega_2\left(u\right)}{\left(1+\left(\omega_1\left(u\right)\right)^2\right)^{3/2}} \tag{22}$$

The inspiration for (22) is the definition of curvature κ for a continuous and twice differentiable function $f\left(x\right)$ as shown in equation (23). As the SOM grid is discrete, the derivative approximations of (24) and (26) can be computed.

$$\kappa \triangleq \frac{f^{(2)}\left(x\right)}{\left(1+\left(f^{(1)}\left(x\right)\right)^2\right)^{3/2}} \tag{23}$$

The first derivative $f^{(1)}\left(x\right)$ of (23) at neuron u is approximated by the discrete quantity of equation (24).

$$\omega_1\left(u\right) \triangleq \frac{1}{|\Psi\left(u\right)|}\sum_{v\in\Psi(u)}g\left(\mathrm{weight}\left(u\right),\mathrm{weight}\left(v\right)\right) \tag{24}$$

The set of neurons $\Psi(u)$ is the neurons neighboring u which also belong to the same cluster.

$$\Psi(u) \triangleq \{v \mid v \in \Gamma(u) \wedge v \in c\} \tag{25}$$

Along a similar line of reasoning the second derivative $f^{(2)}(x)$ in equation (23) is approximated as in equation (26) by the second order differences taken over all possible trajectories in the SOM grid leading to neuron u.

$$\omega_2(u) \triangleq \frac{1}{2|\Psi(u)|} \sum_{v \in \Psi(u)} \sum_{s \in \Psi(v)} \frac{\omega_1(u,v) - \omega_1(v,s)}{|\Psi(v)|} \tag{26}$$

From the entries of table V the following conclusions can be drawn. The Gussian graph filters achieve the best performance in the sense that they quickly generate the most compact clusters. This can be attributed to the more flexible the Gaussian kernel has plus its increased smoothness. Concerning the inverse polynomials, the quadratic ones seem to outperform the cubic one which in turn are better to the linear one. Among the inverse quadratic kernels the Chebyshev type I attains better scores. This can be attributed to the fact that they are derived from optimality criteria, whereas the Cauchy kernel was designed for robustness. The logistic kernels achieve an intermediate performance between the quadratic polynomials.

Regarding the role of geometry, different combinations of neighborhood and weight functions can lead to various cluster shapes in $\mathcal{C}$. To this end, shapes like squares and triangles are generally not preferred as they are less smooth, whereas circles and Gaussian shapes are more common as long as the corresponding parameters are such that cluster boundaries are continuous.

V. Conclusions And Future Work

This conference paper focused on the development of graph filters for clustering graphs with self organizing maps (SOMs) where each such graph is a data point containing structural, functional, and psychological attributes describing human personalities, as they are reflected in writing. Clustering took place according to the MBTI taxonomy, which is based on Jungian personality theory. The novelty of this work is that each personality is not represented by a vector but by a graph where vertex proximity implies semantic one. This allows for greater flexibility when discovering clusters in the data set. The SOMs were selected because geometry plays a central role through the respective topological properties of both the original data space and the neuron grid. The results, obtained from *MBTI dataset* posted on Kaggle indicate the superior performance of the Gaussian graph filter.

The methodology proposed here can be extended in a number of ways. First and foremost, the topology of the graphs representing personalities can be changed to include more attributes or new interactions between them. Especially the latter can imply new edges, possibly weighted. Additionally, graphs are by definition an irregular domain in contrast to, for instance, images. Graph filters take this into account by adjusting their structure accordingly, which may well be another research direction. Finally, larger datasets can reveal possible scalability issues of the proposed methodology but they can also hint at ways to circumvent this since graph filters can be applied locally in parallel.

Acknowledgment

This conference paper is part of Project 451, a research initiative whose primary research objective is the development of novel, scalable, and interpretable tensor analytics.

References

[1] G. Drakopoulos, I. Giannoukou, P. Mylonas, and S. Sioutas, "On tensor distances for self organizing maps: Clustering cognitive tasks," in *DEXA part II*, vol. 12392. Springer, 2020, pp. 195–210.

[2] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and G. Lagani, "Hebbian learning meets deep convolutional neural networks," in *International Conference on Image Analysis and Processing*. Springer, 2019, pp. 324–334.

[3] J. Zimmermann, W. C. Woods, S. Ritter, M. Happel, O. Masuhr, U. Jaeger, C. Spitzer, and A. G. Wright, "Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a personality dynamics diary," *Psychological assessment*, vol. 31, no. 4, p. 516, 2019.

[4] K. A. Nisha, U. Kulsum, S. Rahman, M. F. Hossain, P. Chakraborty, and T. Choudhury, "A comparative analysis of machine learning approaches in personality prediction using MBTI," in *CIPR*. Springer, 2022.

[5] C. J. Lake, J. Carlson, A. Rose, and C. Chlevin-Thiele, "Trust in name brand assessments: The case of the Myers-Briggs type indicator," *The psychologist-manager journal*, vol. 22, no. 2, p. 91, 2019.

[6] V. Filimonov, N. Burmistrova, V. Chernyavskaya, and V. Malakhova, "Collective development of cognitive abilities using the "4c" approach," in *CSGB*. IEEE, 2021, pp. 60–63.

[7] A. M. Gordon and D. Jackson, "A balanced approach to ADHD and personality assessment: A Jungian model," *North American Journal of Psychology*, vol. 21, no. 3, 2019.

[8] R. Stein and A. B. Swan, "Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology," *Social and Personality Psychology Compass*, vol. 13, no. 2, 2019.

[9] S. Myers, "Myers-Briggs typology and Jungian individuation," *Journal of analytical psychology*, vol. 61, no. 3, pp. 289–308, 2016.

[10] G. Drakopoulos, I. Giannoukou, S. Sioutas, and P. Mylonas, "Self organizing maps for cultural content delivery," *NCAA*, vol. 31, no. 7, 2022.

[11] Z. Li, J. D. Wegner, and A. Lucchi, "Topological map extraction from overhead images," in *CVPR*, 2019, pp. 1715–1724.

[12] Z. Yang and J. Liu, "Learning of fuzzy cognitive maps using a niching-based multi-modal multi-agent genetic algorithm," *Applied Soft Computing*, vol. 74, pp. 356–367, 2019.

[13] S. Simhon and G. Dudek, "A global topological map formed by local metric maps," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 3. IEEE, 1998.

[14] E. Remolina and B. Kuipers, "Towards a general theory of topological maps," *Artificial Intelligence*, vol. 152, no. 1, 2004.

[15] L. Pasa, N. Navarin, and A. Sperduti, "SOM-based aggregation for graph convolutional neural networks," *NCAA*, vol. 34, no. 1, 2022.

[16] G. Drakopoulos, E. Kafeza, P. Mylonas, and L. Iliadis, "Transform-based graph topology similarity metrics," *NCAA*, vol. 33, no. 23, pp. 16 363–16 375, 2021.

[17] Y. Liu, J. Chen, Y. Lu, W. Ou, Z. Long, and C. Zhu, "Adaptively topological tensor network for multi-view subspace clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[18] J. Zeng, C. Li, Z. Sun, Q. Zhao, and G. Zhou, "TNGPS: Discovering unknown tensor network structure search algorithms via large language models (LLMs)," in *ICML*. IMLS, 2024.

[19] M. Karavokyris and S. Sioutas, "Graph neural networks for affective social media: A comprehensive overview," in *CIKM 2022 Workshops*, ser. CEUR, G. Drakopoulos and E. Kafeza, Eds., vol. 3318. CEUR-WS.org, 2022.

[20] P. Ganesan, S. K. Jagatheesaperumal, M. M. Hassan, F. Pupo, and G. Fortino, "Few-shot image classification using graph neural network with fine-grained feature descriptors," *Neurocomputing*, vol. 610, 2024.

[21] S. Singh, C. J. Steinmetz, E. Benetos, H. Phan, and D. Stowell, "ATGNN: Audio tagging graph neural network," *IEEE Signal Processing Letters*, 2024.

[22] G. Drakopoulos, E. Kafeza, P. Mylonas, and S. Sioutas, "Approximate high dimensional graph mining with matrix polar factorization: A Twitter application," in *IEEE Big Data*. IEEE, 2021, pp. 4441–4449.

[23] G. Drakopoulos and P. Mylonas, "A genetic algorithm for Boolean semiring matrix factorization with applications to graph mining," in *Big Data*. IEEE, 2022.

[24] S. Goerttler, M. Wu, and F. He, "Understanding concepts in graph signal processing for neurophysiological signal analysis," in *Machine Learning Applications in Medicine and Biology*. Springer, 2024, pp. 1–41.

[25] E. Isufi, F. Gama, D. I. Shuman, and S. Segarra, "Graph filters for signal processing and machine learning on graphs," *IEEE Transactions on Signal Processing*, 2024.

[26] S. Kontopoulos and G. Drakopoulos, "A space efficient scheme for graph representation," in *ICTAI*. IEEE, 2014, pp. 299–303.

[27] N. Yu, Y. Zhai, Y. Yuan, and Z. Wang, "A bionic robot navigation algorithm based on cognitive mechanism of hippocampus," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1640–1652, 2019.

[28] A. Babichev, D. Morozov, and Y. Dabaghian, "Replays of spatial memories suppress topological fluctuations in cognitive map," *Network Neuroscience*, vol. 3, no. 3, pp. 707–724, 2019.

[29] A. Furnham, "Myers-Briggs type indicator (MBTI)," *Encyclopedia of personality and individual differences*, pp. 3059–3062, 2020.

[30] R. E. Plutchik and H. R. Conte, *Circumplex models of personality and emotions*. American Psychological Association, 1997.

[31] P. Ekman, "Darwin, deception, and facial expression," *Annals of the New York Academy of Sciences*, vol. 1000, no. 1, pp. 205–221, 2003.

[32] X. Sun, Z. Pei, C. Zhang, G. Li, and J. Tao, "Design and analysis of a human-machine interaction system for researching human's dynamic emotion," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2019.

[33] G. Drakopoulos and E. Kafeza, "Estimating the MBTI of Twitter accounts with graph neural networks over Neo4j," in *THECOG*. CEUR-WS, 2022.

[34] Y. Xie, R. Liang, Z. Liang, C. Huang, C. Zou, and B. Schuller, "Speech emotion classification using attention-based LSTM," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 11, pp. 1675–1685, 2019.

[35] Y. Kim, J. Moon, N.-J. Sung, and M. Hong, "Correlation between selected gait variables and emotion using virtual reality," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–8, 2019.

[36] T.-L. Nguyen, S. Kavuri, and M. Lee, "A multimodal convolutional neuro-fuzzy network for emotion understanding of movie clips," *Neural Networks*, vol. 118, pp. 208–219, 2019.

[37] S. Poria, N. Majumder, R. Mihalcea, and E. Hovy, "Emotion recognition in conversation: Research challenges, datasets, and recent advances," *IEEE Access*, vol. 7, pp. 100 943–100 953, 2019.

[38] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE Access*, vol. 7, pp. 111 866–111 878, 2019.

[39] R. Beliy, G. Gaziv, A. Hoogi, F. Strappini, T. Golan, and M. Irani, "From voxels to pixels and back: Self-supervision in natural-image reconstruction from fMRI," in *NIPS*, 2019, pp. 6517–6527.

[40] G. Sidhu, "Locally linear embedding and fMRI feature selection in psychiatric classification," *IEEE journal of translational engineering in health and medicine*, vol. 7, pp. 1–11, 2019.

[41] G. Drakopoulos, S. Kontopoulos, and C. Makris, "Eventually consistent cardinality estimation with applications in biodata mining," in *SAC*. ACM, 2016.