



# Feature dimensionality reduction via homological properties of observability

Marcello Trovati<sup>1</sup> · Eslam Farsimadan<sup>2</sup>

Received: 12 June 2023 / Accepted: 6 August 2023  
© The Author(s) 2023

## Abstract

Feature selection and its subsequent dimensionality reduction are significant problems in machine learning and it is at the core of several data science techniques. The ‘shape’ of data, or in other words its related topological properties, can provide crucial insights into the corresponding data types and sources and it enables the identification of general properties that facilitate its analysis and assessment. In this article, we discuss an information theoretic approach combined with data homological properties to assess dimensionality reduction, which can be applied to semantic feature selection.

**Keywords** Information theory · Applied homology · Applied algebraic topology · Dimensionality reduction

## 1 Introduction

Data science has become an extensively investigated discipline, where diverse methods are applied to identify and extract actionable information.

Mathematical areas including Network Theory, Dynamical Systems, Topology, as well as sophisticated algebraic approaches have been introduced, which have demonstrated their potential in addressing data complexity and their applications have led to the design of effective algorithms (Trovati et al. 2020; Shao et al. 2017; Ray and Trovati 2018; Xu et al. 2019; Ray et al. 2018; Trovati et al. 2022).

In this work, an approach based on the topological and information theoretic properties of networks is introduced, which aims to assess and analyse semantic features and their reduction. The main motivation follows from the recent advance in Topological Data Analysis (TDA) which has been shown to identify invariant data properties that can be used to assess and analyse large quantities of unstructured data,

otherwise difficult to investigate via standard approaches (Carlsson 2008).

We define a *knowledge system* as a set of concepts and mutually connecting semantic properties, which naturally define semantic networks. More specifically, they are defined as networks  $G = G(V, E)$ , where  $V = \{v_i\}_{i=1}^n$  is the *node-set* and  $E = \{e(v_i, v_j)\}_{i \neq j=1}^n$  is the *edge set*; each edge can be represented as a multi-dimensional vector based on a predefined set of (semantic) relations. In other words, multiple relations typically exist between two concepts. Furthermore, each node is embedded into an multi-dimensional feature space. It is important to specify that features are assumed to be semantic, that is well defined relational concepts, which can be quantified based on their attributes. These properties define *semantic features*, which are the central part to this work. Loosely speaking, they refer to semantic properties, which are linked to (potentially multiple) edges, or embedded into the nodes. In particular, we will assume that each component of every edge can be evaluated within the interval  $[0, 1]$  via a weight  $w(e(v_i, v_j)) : E \mapsto [0, 1]$  of an edge  $e(v_i, v_j) \in E$ .

As discussed above, each edge contains specific information embedded in an  $n$ -dimensional *feature space*  $\mathcal{R} \subset [0, 1]^n$ . These features are typically captured by multiple edges between nodes, which can be flattened into a single edge. We will assume that such evaluation and assessment can be carried out independently from the approach introduced in this work, and as part of any suitable data pre-processing process. Furthermore, it is also assumed that

✉ Marcello Trovati  
trovatim@edgehill.ac.uk  
Eslam Farsimadan  
efarsimadan@unisa.it

<sup>1</sup> Department of Computer Science, Edge Hill University, St Helens Road, Ormskirk, UK

<sup>2</sup> Department of Computer Science, University of Salerno, Via Giovanni Paolo II, Fisciano, Salerno, Italy

an appropriate measure to capture the potentially multiple semantic links between pairs of nodes has been carried out.

As a consequence, each node  $v_i \in V$  can be represented as an element of  $[0, 1]^n$ . Furthermore, every edge is associated with a map  $R_k : x \rightarrow y$  for the  $k$ -th feature in  $\mathcal{R}$  linking the nodes  $x$  and  $y \in V$ . Finally, let

$$n(x \leftrightarrow y) = \{R_1, \dots, R_n\} \subset \mathcal{R}, \quad (1)$$

be the set of relations shared by the nodes  $x$  and  $y$ .

This article is structured as follows: in Sect. 2, existing research and methods are discussed. Sections 3 and 4 introduce the main background information and the main results, respectively. Finally, Sect. 5 concludes the article and prompts to new directions.

## 2 Related works

Much of the research in this area is focused on the extraction of information that may be utilised to make decisions from unstructured data (Chazal and Michel 2017). In fact, research and industry organisations, as well as general human activities, generate massive amount of data, which poses significant challenges, as well as significant opportunities. A major problem is comprehending and evaluating the importance and worth of the useful information that is included within this abundance of data.

With an emphasis on the structure and *shape* of data in terms of the relationships between its components, new theoretical methods based on topological techniques have been created (Carlsson 2008). The capacity to categorise things based on their shared characteristics, or in other words, those displaying invariant qualities, is one property of topology. The geometrical characteristics of the data under investigation are crucial in many clustering and classification techniques. In reality, a large portion of these methods concentrate on the idea of distance, which must be described

in terms of the mathematical spaces in which the data is contained.

Persistent homology, which focuses on identifying the topological qualities that remain invariant (Edelsbrunner and Harer 2008), has been drawing increasing attention from the research community due to its applicability to AI, machine learning and data science in general. Simplicial complexes are the fundamental components in persistent topology. These are topological space triangulations consisting of coupled, non-overlapping polyhedra, as depicted in Fig. 1.

Image pixellation is a simple but instructive example of a space triangulation in which a genuine image is covered in pixels to provide an accurate approximation.

The fact that a simplicial complex provides an ‘approximation’ of the object which is covered via a *triangulation*. Triangulations include, but are not limited to Voronoi diagrams, Delaunay triangulations, Vietoris complexes, and Cech complexes. In general, they are defined in terms of a certain distance or by ball intersections, whose centres are the data points. Additionally, the adjacency graphs produced by these triangulations can provide a number of details about their invariant topological characteristics, making them pertinent to data science (Carlsson 2008). Refer to Edelsbrunner and Harer (2008) for further information on simplicial complexes.

## 3 Observability

In this section, the concept of observability is introduced. The main motivation is that information is inter-twinned with its observability. In other words, information must be observed. We define the following types of observability which, as explained above, are embedded in a knowledge system, or in other words, a suitably defined semantic network.

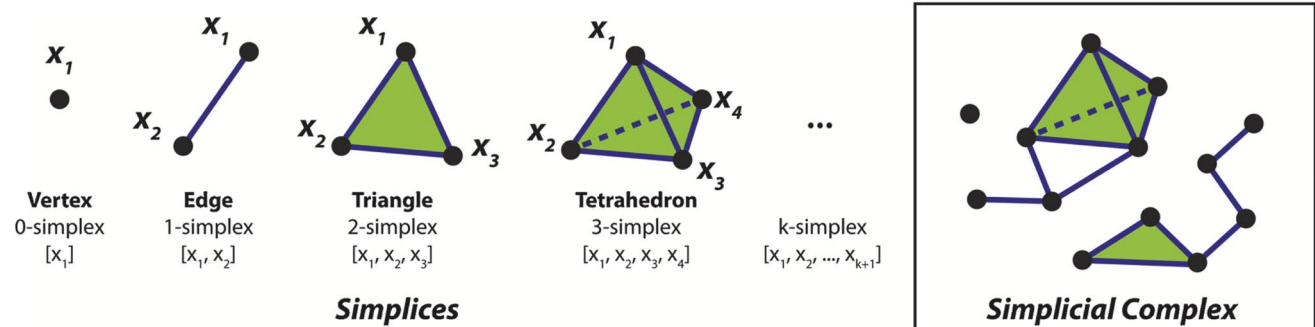


Fig. 1 The graphical representation of simplicial complexes as described in Zhang et al. (2020)

- **Single object observability:**  $O(x)$  measures the observability of  $x$  with respect to the other nodes. More specifically, we have
  - **Absolutely observability:** this is a single observability which covers the entire network
  - **Local observability:** similar to the above but covers only a subnetwork containing  $x$
- **Mutual observability:** this occurs between two (semantic) objects  $x$ , and  $y$ , and it is defined as  $O(x, y)$ .

In the rest of this section, a formal definition of the relevant observabilities derived from the above instances, is introduced.

**Definition 1 (Observabilities)** Let  $G = G(V, E)$  be a (semantic) network (knowledge system) and let  $x, y \in V$  be two nodes, and  $e \in E$  be an edge. Assume that  $n(x)$  is the set of incidents edges with respect to  $x$  (or the degree of  $x$ ), so that  $n(x \leftrightarrow y)$  is the set of (undirected) edges connecting  $x$  with  $y$ . Let  $w(e) \in (0, 1]$  be the weight of the edge  $e$ . We define the *observability from  $x$  to  $y$*  as

$$O(x \rightarrow y) = -\log \left( \frac{w(x \leftrightarrow y)}{w(x)} \right), \tag{2}$$

where

$$w(x \leftrightarrow y) = \sum_{e_i \in n(x \leftrightarrow y)} w(e_i) \tag{3}$$

and

$$w(x) = \sum_{e_j \in n(x)} w(e_j). \tag{4}$$

In this article, a simplified version of observability, the *naïve observability of  $x$* , will be used, which is defined as

$$O_n(x) = -\log \left( \frac{w(x)}{\sum_{y \in V(\tilde{G})} w(y)} \right). \tag{5}$$

The difference between the naïve observability  $O_n(x)$  and those described above is that the former only consider the level of connectivity of  $x$  with its neighbours compared to all the other nodes in the network  $G$ . In other words, this type of observability is only affected by the immediate connections of  $x$ , as opposed to the overall topology of  $G$  and the corresponding connected components and paths. Despite this, it proves to provide a useful tool in the analysis of the results in the remaining of the article.

It is trivial to see that  $O(x \rightarrow y)$  is a *premetric* as in general it is not symmetric and the triangle inequality does not hold.

Let  $\underline{x} = x_1 x_2 \dots x_n$  be a path joining a set of incident nodes. We can easily see that the observability of such a path is

$$O(\underline{x}) = \sum_{i=1}^{n-1} O(x_i, x_{i+1}) \tag{6}$$

In particular, using Eq. 2 this can be written as

$$\begin{aligned} O(\underline{x}) &= -\log \left( \frac{w(x_1 \leftrightarrow x_2)}{w(x_1)} \dots \frac{w(x_{n-1} \leftrightarrow x_n)}{w(x_{n-1})} \right) \\ &= -\log \left( \prod_{i=1}^{n-1} \frac{w(x_i \leftrightarrow x_{i+1})}{w(x_i)} \right) \end{aligned} \tag{7}$$

### 3.1 Probability as a function of observability

Any probability measure defined over complex networks is closely related to their overall topology, which usually depends on the corresponding degree distribution.

Based on the above and on general network properties (Newman 2010), we have that

$$\begin{aligned} p(x) &= e^{-O_n(x)}, \\ p(x, y) &= e^{-O(x,y)}, \text{ and} \\ p(y | x) &= e^{-O(x \rightarrow y)}, \end{aligned} \tag{8}$$

where  $p(x)$ ,  $p(x, y)$  and  $p(y | x)$  are the probability of  $x$ , the joint probability of  $x$  and  $y$ , and the conditional probability of  $y$  given  $x$ , respectively.

Using the notation introduced in Definition 1, let  $x_1$  and  $x_k$  be two nodes. Define the path  $\underline{x} = x_1, \dots, x_k$  and let the probability  $p(\underline{x})$  of reaching  $x_k$  from  $x_1$  along the is

$$p(\underline{x}) = \prod_{i=1}^{k-1} \frac{w(x_i \rightarrow x_{i+1})}{w(x_i)} \tag{9}$$

Note that, if we define a path  $\tilde{x} = x_k, \dots, x_1$ , that is the path going in the opposite direction of  $\underline{x}$ , we have that

$$p(\tilde{x}) = \prod_{i=1}^{k-1} \frac{w(x_{k-i} \rightarrow x_{k-i-1})}{w(x_{k-i})} \tag{10}$$

and  $p(x) = \frac{w(x)}{\sum_{y \in V(\tilde{G})} w(y)}$  for any node  $x \in V$ .

The above shows that, in general,  $p(x) \neq p(\tilde{x})$ . Again, we can easily see that  $p(\underline{x})$  can be written in terms of  $O(\underline{x})$  as

$$p(\underline{x}) = e^{-\sum_{i=1}^{k-1} O(x_i \rightarrow x_{i+1})} = e^{-O(\underline{x})}. \tag{11}$$

The above observations and remarks lead to the following result

**Lemma 1** Let  $O(x, y)$  be the joint observability of  $y$  and  $x$ . Then

$$O(x, y) = O(x \rightarrow y) + O(x) \tag{12}$$

**Proof** Note that

$$p(y | x) = \frac{p(x, y)}{p(x)}.$$

Since

$$p(y | x) = e^{-O(x \rightarrow y)},$$

then we have that

$$e^{-O(x \rightarrow y)} = \frac{e^{-O(x, y)}}{e^{-O(x)}} = e^{O(x) - O(x, y)}. \tag{13}$$

The result follows. □

### 3.2 Mutual information

Mutual information plays an important role in assessing the feature space and in particular, in reducing its dimensionality (Shadvar 2012). It is defined as

$$I(x; y) = p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right) \tag{14}$$

Mutual information can provide useful insights into feature inter-dependencies. In particular, an optimal level of features is linked with a maximum value of  $I(x; y)$ . Similarly to the above, we can write  $I(x; y)$  in terms of observabilities using Eq. 14, namely

$$I(x; y) = e^{-O(x, y)}(O(x) + O(y) - O(x, y)) \tag{15}$$

Eq. 15 can be used to find a lower bound of  $I(x; y)$  with respect to the observabilities. Recall that

$$\begin{aligned} p(x, y) &= p(y | x)p(x) \\ &= e^{-(O(x \rightarrow y) + O(x))} \end{aligned} \tag{16}$$

Hence

$$I(x; y) = e^{-O(x \rightarrow y) + O(x)}(O(y) - O(x \rightarrow y)) \tag{17}$$

Since  $p(x, y)$  can also be written as  $p(x | y)p(y)$ , or in other words as  $e^{-(O(y \rightarrow x) + O(y))}$ , using a similar argument we have that

$$I(x; y) = e^{-O(y \rightarrow x) + O(y)}(O(x) - O(y \rightarrow x)) \tag{18}$$

Recall from Definition 1,  $O(y \rightarrow x) = O(x \rightarrow y) - \log \frac{w(x)}{w(y)}$ .

This implies that, if no extra neighbouring nodes are added to either  $x$  or  $y$ , then if  $O(y \rightarrow x)$  increases,  $O(x \rightarrow y)$  will

increase as well, and vice-versa. This leads to the following result.

**Lemma 2** Let  $I(x; y)$  be the mutual information between  $x$  and  $y$ . If  $O(y \rightarrow x)$ , or  $O(x \rightarrow y)$ , are minimised then  $I(x; y)$  will increase towards its maximum value.

It is important to emphasise that Lemma 2 does not provide an exact evaluation of the behaviour of mutual information, but only an estimation. However, it provides a tool to identify which features should be removed likely to increase mutual information.

### 3.3 Entropy and its variations

From Eq. 17, we can easily calculate the conditional entropy as

$$H(x_2 | x_1) = -e^{(O(x_1 \rightarrow x_2) + O(x_1))}(O(x_1 \rightarrow x_2) + O(x_1)) \tag{19}$$

If we consider such (semantic) networks as dynamical networks, which change with respect to a (discrete) ‘time’ parameter, Eq. 2 can be written to emphasise the iteration time  $k$ , as

$$O_k(x_1 \rightarrow x_2) = -\log \left( \frac{w_k(x_1 \leftrightarrow x_2)}{w_k(x_1)} \right), \tag{20}$$

Assume we remove one semantic feature from  $n(x_1, x_2)$  and let the corresponding weight of such feature be  $\tilde{w}$ . Equation 20 can be re-written as:

$$\log(w_k(x \leftrightarrow y)) = \log(w_k(x)) - O_k(x \rightarrow y). \tag{21}$$

Since

$$\begin{aligned} w_{k-1}(x \leftrightarrow y) &= w_k(x \leftrightarrow y) - \tilde{w} \\ w_{k-1}(x) &= w_k(x) - \tilde{w} \end{aligned} \tag{22}$$

From the above observations, for a path  $\underline{x} = x_1, \dots, x_k$ , we have that

$$\log \frac{O_k(\underline{x}) - O_{k-1}(\underline{x})}{\prod_{i=1}^n \tilde{w}(x_i)} \geq \frac{O_k(\underline{x}) - O_{k-1}(\underline{x})}{(w_k(x_i) - \tilde{w})(w_k(x_n) - \tilde{w}) \prod_{i=2}^{n-1} w_k(x_i - 2\tilde{w})} \tag{23}$$

Let

$$\Delta_k(H(x_2 | x_1)) = |H_k(x_2 | x_1) - H_{k-1}(x_2 | x_1)| \tag{24}$$

From the above equations, we can derive that

$$\begin{aligned} \Delta_k(H(x_2 | x_1)) &\leq (O_k(x_1 \rightarrow x_2) - O_{k-1}(x_1 \rightarrow x_2)) \frac{\tilde{w}}{w_k(x_1)} \\ &\leq \Delta_k(O_k(x_1 \rightarrow x_2)) \end{aligned} \tag{25}$$

using the notation in (24), and we can also derive that

$$\Delta_k(H(x_2 | x_1)) \geq \sum_{i=1}^{n-1} \Delta_k(O_k(x_i \rightarrow x_{i+1})). \tag{26}$$

If we use  $H_k(x_i | x_j)$  for two adjacent nodes  $x_i$  and  $x_j$  at time iteration  $k$ , as a general entropy ‘indicator’, we can consider its overall contribution based on all the nodes in the network  $G_k(V, E)$ . In other words,

$$H_k(G) = \frac{1}{2} \sum_{i \in V} \sum_{j_i \in n_k(i)} H_k(x_{j_i} | x_i). \tag{27}$$

This will allow to evaluate

$$\Delta_k(H(G)) = \frac{1}{2} \sum_{i \in V} \left( \sum_{j_i \in n_k(i)} H_k(x_{j_i} | x_i) - \sum_{j_i \in n_{k-1}(i)} H_{k-1}(x_{j_i} | x_i) \right). \tag{28}$$

Following a similar approach as above, we can also see that

$$\Delta_k(H(G)) \leq \frac{1}{2} \sum_{i \in V} \sum_{j_i \in n_k(i)} \Delta_k(O(x_i \rightarrow x_{j_i})) \frac{\tilde{w}}{w_k(x_i)}, \text{ and} \tag{29}$$

$$\Delta_k(H(G)) \geq \frac{1}{2} \sum_{i \in V} \sum_{j_i \in n_k(i)} \Delta_k(O(x_i \rightarrow x_{j_i})) \frac{\tilde{w}}{w_k(x_i) - w_k(z)}. \tag{30}$$

Recall that we are removing one semantic relation across all the edges. Therefore, Eq. (29), can be arranged as follows

$$\begin{aligned} \Delta_k(H(G)) &\leq \frac{\tilde{w}}{2} \sum_{i \in V} \sum_{j_i \in n_k(i)} \frac{\Delta_k(O(x_i \rightarrow x_{j_i}))}{w_k(x_i)} \\ &= \frac{\tilde{w}}{2} \sum_{i \in V} \frac{\Delta_k(O_n(x_i))}{w_k(x_i)}. \end{aligned} \tag{31}$$

The above calculation lead to the following result

**Lemma 3** *The entropy variation  $\Delta_k(H(G))$  of the network  $G$  at the time iteration  $k$ , can be minimised by minimising*

$$\sum_{i \in V} \frac{\Delta_k(O_n(x_i))}{w_k(x_i)}$$

This lemma further demonstrates the role of observability in the dynamical and topological properties of the corresponding networks. Although it would be equivalent to maximise  $I(x; y)$  directly, this approach strengthen the motivation of using observability to achieve this, which can be used to further investigate and analyse the topological and stochastic properties of the corresponding network.

## 4 Feature dimensionality reduction

Lemma 2 provides a useful tool to assess a feature space based on the corresponding observabilities. This will be achieved by considering the (semantic) network, as defined above. This will be constructed as a simplicial complex as discussed in Sect. 4.1.

### 4.1 Simplicial complexes via observability

Observability, despite being a premetric, can naturally create a network structure, or more precisely a simplicial complex  $G = G(V, E)$ . More specifically, we assume that a link (or edge) exists between any two nodes in  $V$ , if their observability is less than a given (numerical) threshold  $\theta > 0$ , which will define a filtration of the simplicial complex (Chazal and Michel 2017).

More formally, we define the edge-set  $E$  of a simplicial complex  $G$  as

$$E = \{e(v_i, v_j) : O(v_i \rightarrow v_j), \text{ or } O(v_i \rightarrow v_j) \leq \theta\}. \tag{32}$$

Note that (32) implies that  $G$  is an undirected network. Furthermore, as discussed in Sect. 1, the nodes of  $G$  are  $n$ -dimensional so that they can be embedded in an  $n$ -dimensional space to satisfy the homology construction as per Eq. 33.

We can define the usual simplicial homology

$$G_n \xrightarrow{\partial_n} G_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_2} G_1 \xrightarrow{\partial_1} G_0 \xrightarrow{\partial_0} 0 \tag{33}$$

where  $\partial_m \circ \partial_{m+1} = 0$ . The  $m$ -th homology group is defined as  $H_m(G) := \ker(\partial_m) / \text{Im}(\partial_{m+1})$

The assessment of  $m$ -dimensional homology groups, will determine, loosely speaking, the number of  $m$ -dimensional holes in  $G$ , which can be associated with the corresponding Betti numbers  $b_k$ . Recall that a contractible space has trivial reduced homology groups, Hatcher (2002). Based on Alexander and Bishop (1990), and the fact that  $G$  is typically a compact space, a non-contractible space is not necessarily a uniquely geodesic space. Therefore,  $G$  might not be a uniquely geodesic space and so there are some points (associated with nodes in the simplicial complex) which cannot be connected via shortest path (from an observability point of view). This implies that mutual information cannot be in general maximised for all nodes in  $V$ . Therefore, by iteratively removing the appropriate features associated to the nodes in the node-set  $V$ , the feature space would be contractible. The above facts lead to the following result.

**Proposition 1** *Let  $G = G(V, E)$  be a simplicial complex as defined in Sect. 4.1. Its overall mutual information can be*



**Table 1** Details of the evaluation

$\theta$	Approx number of edges	Number of (reduced) feature dimensions
0.8	2400	6
0.6	970	8
0.4	350	7

maximised by removing those feature dimensions related to the nodes in  $V$ , that do not yield trivial reduced homology groups.

The above proposition can be written as the following algorithm

- Step 1: let  $G = G(V, E)$  be the simplicial complex as per Proposition 1.  
 Step 2: let  $p : G_n \rightarrow G_m^k$  be the projection from the  $n$ -dimensional space  $G$  onto the  $m$ -dimensional space  $G_m^k$ , by removing  $k$  features so that  $n = m + k$ .  
 Step 3: start from  $k = 0$   
 Step 4: if  $b_k \neq 0$  for  $G_m^k$ , then  $k = k - 1$  and repeat this step  
 Step 5: else, stop.

## 4.2 Evaluation

As part of this work, an evaluation of the above approach was carried out. The datasets introduced in Trovati et al. (2020) and Trovati et al. (2022) based on ConceptNet (Liu and Singh 2004) were used to create a selection of semantic networks (with approximately 150 nodes), based on different values of  $\theta$ . Each node had 10 features.

Table 1 depicts the main results for different  $\theta$  which generates different edges and from 10 different reductions

The data used in this evaluation is limited due to the extensive manual labelling required, which has been created by the authors as part of ongoing research (Trovati et al. 2022). However, more datasets are being designed to provide more comprehensive data-driven evaluations for future research and implementations.

## 5 Conclusions

In this article, a method to identify the most likely feature dimensions to reduce based on homological properties of observability is introduced. Despite the limited experimental evaluation, the potential of this approach is clear, both from theoretical and implementation perspectives. In fact, the use of the concept of observability suggests important research areas to further investigate where the dynamical

and topological properties should be integrated to provide a more comprehensive understanding of the evolving nature of data. In particular, future steps include a deeper assessment of different network topology properties which can provide further insights.

**Data Availability** The datasets generated during and/or analysed during the current study are not publicly available due to being under review but are available from the corresponding author on reasonable request

## Declarations

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alexander S, Bishop R (1990) The hadamard-cartan theorem in locally convex metric spaces. *Enseign Math* 36:309–320. <https://doi.org/10.5169/seals-57911>
- Carlsson G (2008) Topology and data. Technical report
- Chazal F, Michel B (2017) An introduction to topological data analysis: fundamental and practical aspects for data scientists. *arXiv:1710.04019*
- Edelsbrunner H, Harer J (2008) Persistent Homology – a Survey. In: *Surveys on Discrete and Computational Geometry*, vol. 453, p. 257. Amer Mathematical Society
- Hatcher A (2002) *Algebraic Topology*. Cambridge University Press, Cambridge
- Liu H, Singh P (2004) Conceptnet – a practical commonsense reasoning tool-kit. *BT Technol J* 22(4):211–226. <https://doi.org/10.1023/B:BTJ.0000047600.45421.6d>
- Newman M (2010) *Networks: an Introduction*. Oxford University Press, Inc. [http://scholar.google.de/scholar.bib?q=info:WDP-LHCuDj4J:scholar.google.com/&output=citation&hl=de&as\\_sdt=0,5&ct=citation&cd=0](http://scholar.google.de/scholar.bib?q=info:WDP-LHCuDj4J:scholar.google.com/&output=citation&hl=de&as_sdt=0,5&ct=citation&cd=0)
- Ray J, Trovati M (2018) A survey of topological data analysis (tda) methods implemented in python. In: *Advances in Intelligent Networking and Collaborative Systems: The 9th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2017)*, pp. 594–600. Springer
- Ray J, Johnny O, Trovati M, Sotiriadis S, Bessis N (2018) The rise of big data science: a survey of techniques, methods and approaches in the field of natural language processing and network theory. *Big Data Cognit Comput* 2(3). <https://doi.org/10.3390/bdcc2030022>
- Shadvar A (2012) Dimension reduction by mutual information feature extraction. *CoRR arXiv:1207.3394*

- Shao Y, Trovati M, Shi Q, Angelopoulou O, Asimakopoulou E, Bessis N (2017) A hybrid spam detection method based on unstructured datasets. *Soft Comput-A Fusion Found Methodol Appl* 21(1):233–243
- Trovati M, Zhang H, Ray J, Xu X (2020) An entropy-based approach to real-time information extraction for industry 4.0. *IEEE Trans Ind Inf* 16(9):6033–6041
- Trovati M, Johnny O, Xu X, Polatidis N (2022) A new model for artificial intuition. In: *Artificial Neural Networks and Machine Learning - ICANN 2022: 31st International Conference on Artificial Neural Networks*, Bristol, UK, September 6-9, 2022, Proceedings, Part I, pp. 454–465. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-031-15919-0\\_38](https://doi.org/10.1007/978-3-031-15919-0_38)
- Xu X, Hu N, Li T, Trovati M, Palmieri F, Kontonatsios G, Castiglione A (2019) Distributed temporal link prediction algorithm based on label propagation. *Fut Gen Comput Syst* 93(C), 627–636. <https://doi.org/10.1016/j.future.2018.10.056>
- Zhang M, Kalies W, Kelso S, Tognoli E (2020) Topological portraits of multiscale coordination dynamics. *J Neurosci Methods* 339:108672. <https://doi.org/10.1016/j.jneumeth.2020.108672>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.