



HAL
open science

The Coincidence Similarity Index under Rotation

Luciano da Fontoura Costa

► **To cite this version:**

| Luciano da Fontoura Costa. The Coincidence Similarity Index under Rotation. 2023. <hal-03967181>

HAL Id: hal-03967181

<https://hal.science/hal-03967181v1>

Preprint submitted on 1 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

The Coincidence Similarity Index under Rotation

Luciano da F. Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

30th Jan. 2023

Abstract

Operations aimed at quantifying the similarity or distance between two given vectors play an especially important role in science and technology, especially in artificial intelligence, pattern recognition, and deep learning. Two families of vector comparison operators can be identified, corresponding to those that are bilinear and those that are non-linear. While the former is mostly founded on the inner product, the latter involves non-linear operations such as maximum, minimum, absolute values, etc. Provided the original data, as well as its pre-processing – aimed at normalization, filtering and feature extraction – it becomes interesting to consider the invariance of the comparison operations to specific types of transformations (e.g. rotation). Though the bilinear comparison operators are often invariant to rotation, the non-linear comparison approaches are almost invariably influenced by rotation. Suggested recently, the coincidence similarity index presents several interesting properties that make it particularly effective for several applications. However, though it is less influenced by rotation when the two compared vectors are similar, this similarity index is not invariant to rotation. In the present work, a procedure is suggested that is capable of making the coincidence similarity index mostly invariant to rotations. The potential of the proposed methodology is illustrated respectively to two benchmarking procedures, with encouraging results that complement also the aspect of rotational invariance among the several interesting features of the coincidence similarity index for performing strict comparisons between vectors.

1 Introduction

Several concepts and methods in science and technology rely on some measurement of similarity or distance between data properties organized as *feature vectors* (e.g. [1, 2, 3, 4]). Similarity (or dissimilarity) and distance (or proximity) are important because they allow entities to be compared and ordered, providing the basis for several concepts and methods in artificial intelligence, pattern recognition and deep learning (e.g. [5, 6, 7]).

Several comparison measurements have relied on the inner product, which is a bilinear operation involving products between the coordinates of the compared vectors. Examples of similarity/distance measurements that are based on the inner product include the Euclidean distance, the cosine similarity, and the Pearson correlation coefficient.

Another approach that has also been considered consists of using non-linear similarities and distances such as the Manhattan distance, Hamming distance, Minkowski distance, Sørensen-Dice similarity, Jaccard similarity index, coincidence similarity index, and overlap index, among many other possibilities.

These two great families of approaches will henceforth be understood as specific types of *comparison operators*.

A typical difference between the two great families of similarity/distance measurements based on the inner product (bilinear) and on non-linear operations concerns their *variance* (or *invariance*) to specific types of transformations. For instance, the Euclidean distance is invariant to rotations, while the Manhattan distance is not.

Interestingly, the issue of invariance of comparison measurements is not only related to the intrinsic geometric properties of the operators, but also strongly dependent on the type of data to be compared, which often involve preliminary modifications aimed at normalizing, filtering, or changing in some specific manner the original data.

Actually, the frequently adopted derivation of additional features from the original orthonormal data can also be understood as a specific type of data transformation. Examples of measurements obtained from original data include the entropy of orientations or gray-levels, color contrast, lengths, spatial lacunarity, and curvatures, among many other possibilities. Each of these measurements may or not be intrinsically invariant to specific transformations.

All in all, the subject of invariance to specific transformations depend on three main factors: (a) the original type of data and respective spaces; (b) preliminary handling of these data; and (c) intrinsic invariance of the comparison operations. These three aspects can be understood to correspond to successive necessary conditions, in the sense that (b) requires (a), and (c) requires (b), which can be summarized as:

$$\boxed{\text{data inv.} \implies \text{transformed data inv.} \implies \text{operation inv.}}$$

In other words, the consideration of the intrinsic invariance (or not) of given comparison operators assumes that the data, and their normalization, are invariant to the same transformations. Interestingly, this is not always the case because, as discussed in [8], it is often impossible, or very difficult, to establish the type of space underlying the original data. For instance, how can data of distinct origin including size, weight, age, color, etc. be decided to be orthogonal or even orthonormal? Even so, these data are often understood and treated as being orthonormal (the *Cartesian surmise* [8]).

Interestingly, only in relatively rare circumstances we can with confidence assume that the original data corresponds to vectors in an orthonormal (Cartesian) system, as is the case in Euclidean spaces. One typical example of this type of data consists in the 2D or 3D coordinates of objects from the geometrical space obtained by measurement devices adopting an orthonormal framework.

Even when the original data do correspond to vectors in orthonormal spaces, transformations are often applied in order to normalize the magnitudes, filter noise and outliers, etc. Normalization is particularly frequent because often the problems of interest require *relative* comparison. For instance, comparisons of length differences between insects and between large mammals typically require relative measurements. These normalizations typically imply changes in the data spaces, often resulting in non-orthonormal representations.

Thus, the characterization of the invariance of comparison operations is of particular interest provided the data supplied to the operations are invariant themselves. These situations are typically characterized by geometric data obtained from orthonormal spaces that have undergone pre-processing that retained the geometrical invariance of particular interest in each case.

Being a member of the non-linear comparison operators family, the Jaccard, interiority and coincidence similarity indices are not invariant to rotations. The present work aims at describing a linear-invariant coincidence similarity index,

We start by briefly reviewing the main concepts underlying the invariance of inner product-based measurements in bilinear comparison operators, and follows by present-

ing the coincidence similarity index and respectively derived complex networks, as well as it can be made invariant to rotations by using a pre-processing stage aimed at normalizing the orientation of the pair of vectors to be compared. Preliminary benchmark results are then presented that corroborated the potential of the approach for enhancing also this aspect of the coincidence similarity index.

2 The Inner Product

The inner product between two vectors has been frequently adopted as basis for defining several distance and similarity measurements (e.g. [9]). For instance, the distance between two vectors can be defined as corresponding to the inner product between the difference of the two vectors. The inner product is intrinsically invariant to linear transformations.

Given an N -dimensional vector space S with a non-necessarily orthonormal basis $\mathbf{b} = (\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N)$, the inner product between two vectors $u = (u_1, u_2, \dots, u_N)$ and $v = (v_1, v_2, \dots, v_N)$ from this space can be expressed as:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T G \mathbf{v} \quad (1)$$

where G is the *Gramian* matrix of the basis \mathbf{b} , defined as:

$$G = \begin{bmatrix} \langle \mathbf{b}_1, \mathbf{b}_1 \rangle & \langle \mathbf{b}_1, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{b}_1, \mathbf{b}_N \rangle \\ \langle \mathbf{b}_2, \mathbf{b}_1 \rangle & \langle \mathbf{b}_2, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{b}_2, \mathbf{b}_N \rangle \\ \dots & \dots & \dots & \dots \\ \langle \mathbf{b}_N, \mathbf{b}_1 \rangle & \langle \mathbf{b}_N, \mathbf{b}_2 \rangle & \dots & \langle \mathbf{b}_N, \mathbf{b}_N \rangle \end{bmatrix} \quad (2)$$

Observe that the Gramian matrix becomes the identity matrix in the case of orthonormal bases, giving rise to the dot product in those spaces.

However, in practice the *dot product* is often employed instead of the inner product, which assumes the vector spaces to be orthonormal. As such, the dot product will *not* be invariant to linear transformations, including rotations, in non-orthonormal systems of coordinates.

In case two vectors in a vector space S are linearly transformed by a matrix A , it can be shown that the inner product will be preserved by considering the *metric tensor* $M = A^T A$, which provides the “calibration” for compensating the geometric alterations implemented by the linear transformation.

In addition to underlying the basic invariances of the inner product and respectively derived measurements, the above considerations have implications also for non-linear comparison operations. For instance, in case the

basis vectors of the space underlying the data elements are known, the space can be transformed to a orthonormal counterpart so that rotation and shearing effects are avoided and invariant values are obtained. In a sense, this provides the basis for the method to be suggested in this work, in which the vectors to be compared are rotated into a normalized orthonormal space through a rotation determined, e.g., by the average of the two vectors.

3 The Coincidence Similarity Index

The coincidence similarity index is characterized by particularly strict quantification of the similarity between two non-zero vectors \mathbf{g}_i and \mathbf{g}_j .

Thanks to its interesting properties, the coincidence similarity index and the respective networks have been successfully applied to several situations, including non-linear filtering of 3D scalar fields [10], by template matching [11], image segmentation [12], relationships between city structures [13], defining the autorrelation and cross-relation of networks [14], automatic motifs identification [15], enzymes networks [16], neuromorphological networks [17], as well as finding patterns in bipartite networks [18].

Among the several interesting properties of the coincidence similarity index [11, 8], we have: (i) high selectivity and sensitivity; (ii) inherent normalization in the intervals $[0, 1]$ or $[-1, 1]$ without overlooking the magnitudes of the compared vectors, as is the case with the cosine similarity; (iii) marked tolerance to perturbations of some of the features, as well as substantial invariance to affine and other types of transformations.

In case all features are non-negative, the *Jaccard similarity* between two non-zero feature vectors \mathbf{g}_i and \mathbf{g}_j can be obtained as follows:

$$\mathcal{J}(\mathbf{g}_i, \mathbf{g}_j) = \frac{\sum_{k=1}^F \min(g_{i,k}, g_{j,k})}{\sum_{k=1}^F \max(g_{i,k}, g_{j,k})} \quad (3)$$

with $0 \leq \mathcal{J}(\mathbf{g}_i, \mathbf{g}_j) \leq 1$.

The *interiority index* (also known as overlap, e.g. [19]) can be obtained as:

$$\mathcal{I}(\mathbf{g}_i, \mathbf{g}_j) = \frac{\sum_{k=1}^F \min(g_{i,k}, g_{j,k})}{\sum_{k=1}^F \min(g_{i,k}, g_{j,k})} \quad (4)$$

with $0 \leq \mathcal{I}(\mathbf{g}_i, \mathbf{g}_j) \leq 1$.

The *coincidence similarity* between two non-zero feature vectors can then be written as:

$$\mathcal{C}(\mathbf{g}_i, \mathbf{g}_j) = [\mathcal{J}(\mathbf{g}_i, \mathbf{g}_j)]^K [\mathcal{I}(\mathbf{g}_i, \mathbf{g}_j)] \quad (5)$$

with $0 \leq \mathcal{C}(\mathbf{g}_i, \mathbf{g}_j) \leq 1$, and $D \in \mathbb{R}$.

The higher the value of the parameter D , the more strict (selective and sensitive) the comparison of the similarity between the matrices becomes.

Respectively generalized versions of the above indices, to be used in case the feature vectors can have negative, can be obtained as ([11, 12, 20, 21]):

$$\mathcal{J}(\mathbf{g}_i, \mathbf{g}_j) = \frac{\sum_{k=1}^F s_{xy} \min(|g_{i,k}|, |g_{j,k}|)}{\sum_{k=1}^F \max(|g_{i,k}|, |g_{j,k}|)} \quad (6)$$

$$\mathcal{I}(\mathbf{g}_i, \mathbf{g}_j) = \frac{\sum_{k=1}^F \min(|g_{i,k}|, |g_{j,k}|)}{\sum_{k=1}^F \min(|g_{i,k}|, |g_{j,k}|)} \quad (7)$$

$$\text{where: } s_{xy} = \text{sign}(g_{i,k}) \text{sign}(g_{j,k}) \quad (8)$$

with $-1 \leq \mathcal{J}(\mathbf{g}_i, \mathbf{g}_j), \mathcal{I}(\mathbf{g}_i, \mathbf{g}_j), \mathcal{C}(\mathbf{g}_i, \mathbf{g}_j) \leq 1$. Additional approaches related the Jaccard for negative values, in the context of L1 norm, have also been described [22, 23].

4 Rotational Invariant Coincidence Similarity Index

The coincidence similarity index can be readily made invariant to rotations by applying the methodology described in this section, which involves a pre-processing aimed at normalizing the orientation of the pair of vectors to be compared. This methodology assumes that the vectors are represented in an orthonormal system, otherwise the transformation into such a system (discussed in Section 2) can be considered.

We start by addressing a 2D feature space. Given two vectors \mathbf{v} and \mathbf{r} , they can be orientation-normalized through the following procedure:

- Obtain the average of the two vectors, i.e.: $\mathbf{a} = (\mathbf{v} + \mathbf{r})/2$;
- Rotate the coordinate system so as to align its first axis with the average vector \mathbf{a} ;
- Calculate the coincidence similarity index in this new coordinate system and take it as result.

In the case of more than two dimensions, the respective principal component analysis (PCA, e.g. [24, 25]) can be taken as a reference for the alignment. However, it is also possible to proceed by taking the average vectors into successively reduced spaces. For instance, in the case of 3D, the first axis is aligned to the average vector, and the second axis is aligned to average of the vectors projected into its second and third axes.

The choice of the reference for the rotation normalization will define the reference point for the invariance

between the two supplied vectors. It is interesting to keep in mind that the average vectors constitute but one possibility as reference for normalization. Actually, it is also possible to take the vector with the largest, or smallest magnitude as a reference. Yet another possibility is to take a weighted average of the two vectors while considering their magnitude, or other fitness index, as respective weights. Each of these choices will be particularly suitable given specific datasets and research demands.

Yet other alternative possibilities for achieving rotation invariance include the determination of the coincidence values for a whole cycle (90°) and taking the average, minimum or maximum values. This possibility, however, requires substantially more computational resources than the other approaches suggested in this section.

5 Benchmarking

Two approaches are taken in the present section in order to quantify the effect of the above suggested normalization. First, we obtain the coincidence values while rotating systematically a given pair of vectors to be compared. Second, we adopt a geometrical benchmarking structure.

First, we illustrate the influence of rotation on the coincidence similarity values. Given two specific original vectors, respective coincidence similarity indices were obtained while successively rotating them by an angle θ varying from 0 to 360° .

Figure 1 depicts the results obtained without (red) and with (green) orientation normalization respective to three specific pairs of vectors. In the cases (a,b), which is an example of the particularly challenging situation of comparing two markedly similar vectors, little variance can be observed for the obtained coincidence similarity values. Even so, the suggested orientation normalization allowed complete invariance of the coincidence in this case.

Situations involving more substantially distinct pairs of vectors are shown in (c-e), which led to larger influence of the orientation on the obtained coincidence similarity indices, which was again completely eliminated by the application of the suggested orientation normalization procedure.

In order to complement the preliminary study of the effects of rotation, and respective methodology for achieving invariance, it is interesting to resource to a particular type of dataset illustrated in Figure 2. Basically, we have a lattice centered at a Cartesian system of coordinates. Two parameters are involved, the radius of the disk and the resolutions Δx and Δy , which are henceforth assumed to be equal.

Each point is understood as a data element with respective feature vector containing the x - and y - coordinates.

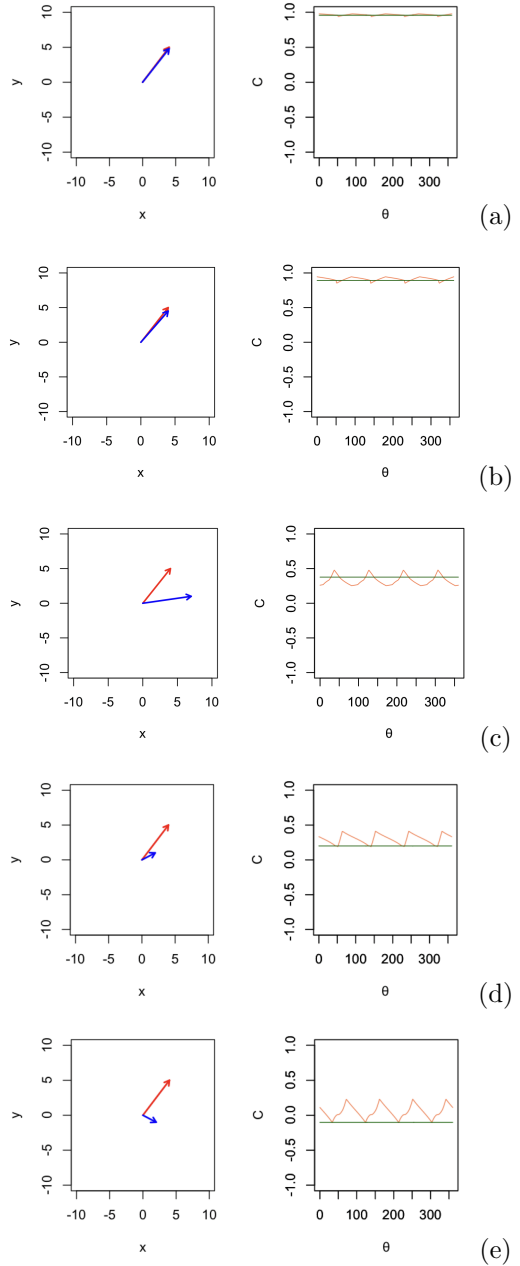


Figure 1: The influence of rotation by successive angles θ on the coincidence similarity index, considering four specific comparisons between progressively more distinct pairs of vectors (a-e), and adopting $D = 1$. When the two compared vectors are similar as in situations (a,b), implying increased demand on selectivity and sensitivity, the influence of the rotation on the coincidence similarity index values is smaller, increasing in cases where the two compared vectors are less similar. As observed in [8], the selectivity and sensitivity of the coincidence similarity will also be reduced in these cases. Also shown (in green) in the graphs are the constant coincidence similarity values obtained in each of the three situations by the invariance procedure suggested in the present work.

Points that are closer in the original space should be understood as being more similar.

Figures 3(a) and (b) depict the coincidence similar-

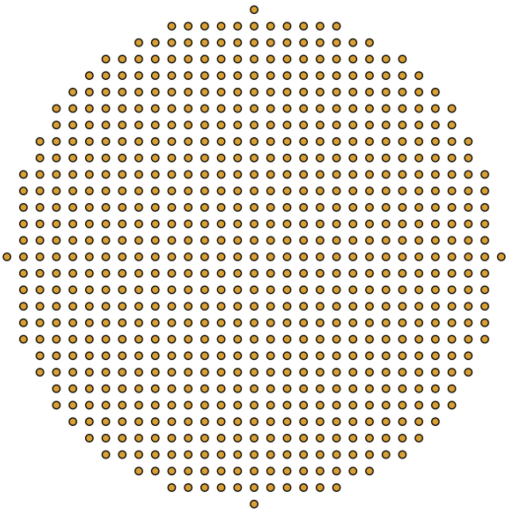
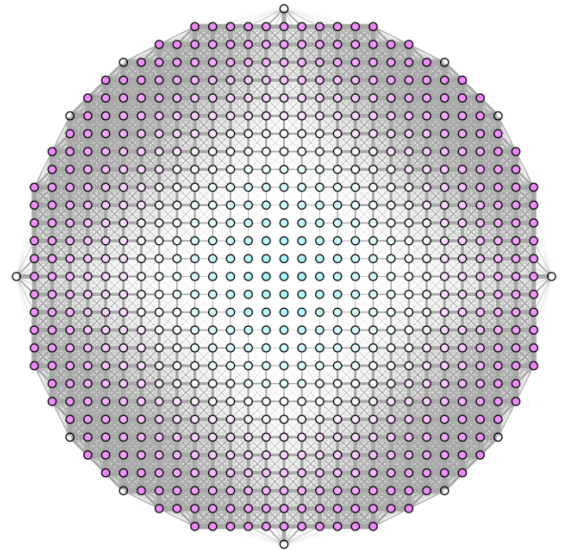


Figure 2: The complementary benchmark considered in the present work in order to study the effects of the rotation, and the respective invariance methodology, on the interconnections between vectors. Each data element is a lattice point in an orthonormal coordinate system centered at the origin and with a given radius. The configuration adopted in the present work considers radius equal to 10 and $\Delta x = \Delta y = 2/3$.

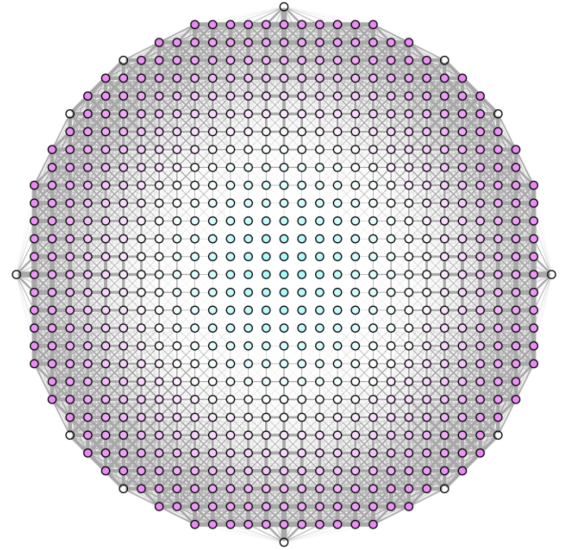
ity complex networks obtained without and with rotation normalization, respectively. The parameter configuration adopted $D = 15$, therefore meaning a relatively strict similarity comparison, implying only vectors corresponding to close points in the original space to result in significant coincidence similarity values. That is why the obtained networks mostly reflect the proximity between the original nodes. Less strict comparisons can be readily achieved by reducing the value of the D parameter.

By comparing the two obtained networks in Figure 3, the differences are not substantial, which confirms the relative tolerance of the coincidence similarity to rotations while comparing similar vectors. However, it can also be readily verified that a more isotropic network has been obtained in the network in Figure 3(b), obtained by employing the suggested orientation normalization procedure. As an additional benefit, several of the connections implied by orientation variations, originally present in the network shown in (a), were effectively removed by the orientation invariance approach, yielding a less cluttered and accurate network that better reflects the spatial relationships between the original points.

Another effect worth observing regards the less intense interconnectivity obtained in both cases respectively to the more central points (cyan). This is a direct consequence of the fact that the coincidence similarity index is normalized, so that the angular difference implied when comparing vectors with smaller magnitudes have a much greater effect on the similarity than when comparing vec-



(a)



(b)

Figure 3: The coincidence similarity complex networks obtained for the benchmark dataset in Fig. 2 without (a) and with (b) the rotation normalization procedure. In both cases, $D = 15$. The width of the obtained edges is proportional to the respectively obtained coincidence similarity values. The network obtained with the rotation normalization is not only more uniform and isotropic (compare with the respective unnormalized counterpart), but also less densely interconnected. The original coordinates of the considered data elements were used as the layout for visualization of the obtained coincidence similarity complex networks. The colors, progressing from cyan to magenta, indicate the distance of the original nodes to the origin of the coordinates system. Only the positive coincidence values are shown in both plates for the sake of not cluttering the visualization.

tors with larger magnitudes. This feature actually constitutes an asset in several situations in which the similarity needs to be taken in relative, normalized terms.

Figure 4 illustrates another benchmark, completely analogous to the previous one, but now considering polar coordinates. The effectiveness of the proposed rotational invariance methodology for the coincidence similarity index is corroborated also in this case, as could be expected.

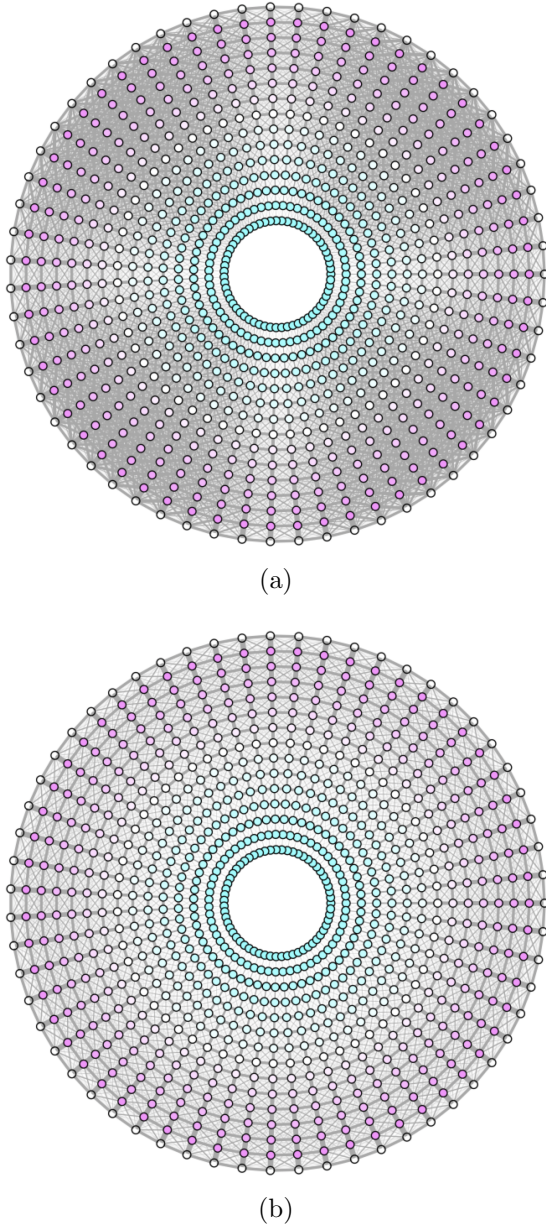


Figure 4: A study analogous to that in Fig. 3, but taking into account points distributed uniformly in a polar system of coordinates. The effectiveness of the rotation invariance procedure suggested in the present work can be again verified: the coincidence similarity network shown in (a) was obtained without rotation invariance pre-processing, while the network depicted in (b) incorporated that invariance methodology.

To conclude this section, we present in Figure 5 the effects of rotation on a geometrical network consisting of 100 points with coordinates drawn with uniform probability in the interval $[0, 1]$. Each point becomes a node of

the network, while the pairwise links correspond to the respective coincidence similarity indices. The pair of networks shown in (a) and (b) refers, respectively, to the original points and the same points rotated by $\theta = 30^\circ$. Though the results look very similar, some differences can be observed that are a consequence of the variance of the coincidence similarity to rotations. The networks in (c) and (d) were obtained in analogous manner, but now considering the invariance methodology described in this work. The results are practically identical, confirming the effectiveness of the proposed methodology.

6 Concluding Remarks

Comparing vectors in terms of respective similarity or distance constitute an important and frequently employed operation in science and technology, and especially in artificial intelligence, pattern recognition, and deep learning. Provided the data and respective pre-processing concern vectors in an orthonormal space, which is not always the case in practice, it becomes of particular interest to consider the invariance of the comparison operations respectively to geometric transformations such as rotations.

Two large families of comparison approaches can be identified, respectively involving bilinear or non-linear operators. The former type, which typically considers the inner product as the basic foundation for measuring similarity or distance, frequently results invariant to rotations. The latter type, involving non-linear operations, is often non-invariant to rotations. The coincidence similarity index, among several other indices and distances, is influenced by rotation when the two compared vectors are not markedly similar.

Interestingly, the coincidence index results almost invariant to rotations when the compared vectors are markedly similar. Comparing similar vectors is a particularly challenging task, given that enhanced selectivity, sensitivity and robustness to local data perturbations are required. This is precisely the case where the coincidence similarity is more effective, therefore implying relative little influence of rotations. Indeed, comparisons between markedly distinct vectors will not, in general, require especially selective or sensitive comparison indices.

However, in order to enhance also the robustness of the coincidence similarity index to rotations when the compared vectors are markedly distinct, it is possible to implement a pre-normalization of the involved vectors. Describing and preliminary quantifying the effect of this rotation normalization constituted the main objectives of the present work.

After presenting the basic approach, which involves rotating the original vectors to be compared into some ori-

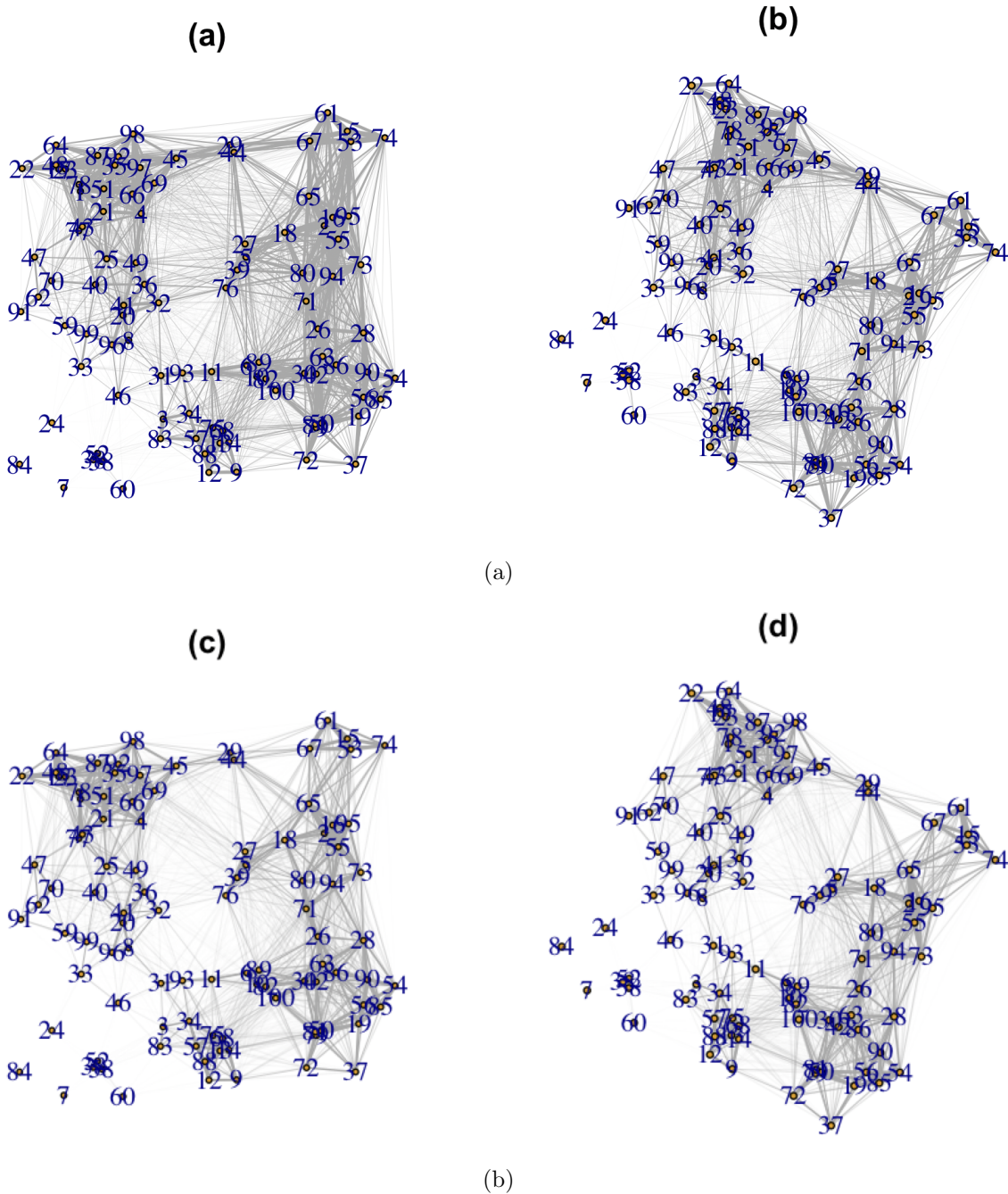


Figure 5: A random geometric network (a), defined by the coincidence similarity between the position of 100 points, with coordinates drawn with uniform probability in the interval $[0, 1]$, and the respective new network (b) obtained after rotating the positions of all original points by $\theta = 30^\circ$. The networks in (c) and (d) are analogous, but considering the rotation invariance approach suggested in this work. Though there is relatively little difference between the networks in (a) and (b), these differences are completely eliminated when applying the invariance approach for obtaining (c) and (d). The parameter D was set as $D = 9$ in both cases.

entational reference, two benchmarking approaches were considered for quantifying the effects of rotations on the obtained coincidence similarity networks, as well as the effectiveness of the suggested procedure for enhancing its stability respectively to vector rotations. The obtained preliminary results are encouraging, indicating that complete invariance is obtained that can contribute to obtaining more isotropic networks.

One interesting point is that it does not necessarily follow that the rotation invariance methodology suggested in this work should necessarily be applied to any type of dataset. This correction is of particular interest when the original data were derived from an orthonormal system of coordinates, and the pre-processing maintained that property. Otherwise, when the features do not necessarily correspond to geometrical quantities obtained from

Euclidean spaces, no or alternative invariance methodologies should be considered, depending on each specific type of data and application requirements. In particular, rotation invariance will have less effect when comparing vectors that are markedly similar, which constitutes one of the main applications of the coincidence similarity index, given its enhanced selectivity and sensitivity. Actually, the incorporation of rotational invariance to the coincidence similarity index will have most of its effects related to the similarity between substantially distinct pairs of vectors, which may not be of particular interest in many applications focused on sharp comparisons between similar data entities.

Possible further developments include the incorporation of the orientation normalization into the own coincidence similarity index, instead of a pre-processing of the original data. In addition, it would be interesting to study the effect of choosing different normalization orientations, such as the maximum magnitude vector, or weighted averages of the two vectors to be compared. It would also be interesting to verify the effect of the suggested methodology respectively to several other types of data, especially in higher dimensions. Yet another interesting prospect is to study the invariance of the coincidence similarity index to more general linear transformations, including non-linear ones (e.g. [9]), and devising schemes for achieving respective invariance. With this respect, preliminary results reported in [12] indicate substantial robustness of this index to shearing transformations.

Acknowledgments

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

Observations

As all other preprints by the author, the present work contains preliminary work subject to further revision and validation. Respective modification, commercial use, or distribution of any of its parts are not possible, as this work has author copyright. Many of the preprints by the author are also available in Hal and/or arXiv. This work can also be cited by using the DOI number or article identification link. Thanks for reading.

References

- [1] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley Interscience, 2000.
- [2] K. Koutrombas and S. Theodoridis. *Pattern Recognition*. Academic Press, 2008.
- [3] K. Fukunaga. *Statistical Pattern Recognition*. Morgan Kaufmann, San Diego, 1990.
- [4] L. da F. Costa. *Shape Classification and Analysis: Theory and Practice*. CRC Press, Boca Raton, 2nd edition, 2009.
- [5] J. Schmidhuber. Deep learning in neural networks: an overview. *Neural networks*, 61:85–117, 2015.
- [6] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800, 2002.
- [7] H. F. de Arruda, A. Benatti, C. H. Comin, and L. da F. Costa. Learning deep learning. Researchgate, 2019. https://www.researchgate.net/publication/335798012_Learning_Deep_Learning_CDT-15. [Online; accessed 22-Dec-2019].
- [8] L. da F. Costa. Multiset neurons. https://www.researchgate.net/publication/356042155_Multiset_Neurons, 2021.
- [9] L. da F. Costa. Nonlinear vector fields: An interconnected approach. Researchgate, 2023. https://www.researchgate.net/publication/367339519_Nonlinear_Vector_Fields_An_Interconnected_Approach.
- [10] E. K. Tokuda, C. H. Comin, and L. da F. Costa. Impact of the topology of urban streets on mobility optimization. *J. Stat. Mech.*, 2022:103204, 2022.
- [11] L. da F. Costa. On similarity. *Physica A*, 599:127456, 2022.
- [12] L. da F. Costa. Multiset neurons. *Physica A*, 609:128318, 2023.
- [13] L. da F. Costa and E. K. Tokuda. A similarity approach to cities and features. *Eur. Phys. J. B*, 95:155, 2022.
- [14] L. da F. Costa. Autorrelation and cross-relation of graphs and networks. <https://iopscience.iop.org/article/10.1088/2632-072X/aca57c/meta>, 2022. *J. Phys.: Complexity*, (3): 045009.

- [15] E. K. Tokuda G. S. Domingues and L. da F. Costa. Identification of city motifs: a method based on modularity and similarity between hierarchical features of urban networks. <https://iopscience.iop.org/article/10.1088/2632-072X/ac9446>, 2022. *J. Phys.: Complexity*, (3): 045003.
- [16] R. dos Reis and L. da F. Costa. Enzyme similarity networks. https://www.researchgate.net/publication/360263872_Enzyme_Similarity_Networks, 2022.
- [17] A. Benatti, H. F. de Arruda, E. Tokuda, and L. da F. Costa. Neuromorphic networks as revealed by features similarity. <https://hal.archives-ouvertes.fr/hal-03689882v1>, 2022.
- [18] L. da F. Costa. Discovering patterns in bipartite networks. https://www.researchgate.net/publication/360408705_Discovering_Patterns_in_Bipartite_Networks, Apr. 2022.
- [19] M. K. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications*, 3(1):19–28, 2016.
- [20] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021. [Online; accessed 21-Aug-2021].
- [21] L. da F. Costa. Coincidence complex networks. <https://iopscience.iop.org/article/10.1088/2632-072X/ac54c3>, 2022. *J. Phys.: Complexity*, (3): 015012.
- [22] B. Mirkin. *Mathematical Classification and Clustering*. Kluwer Academic Publisher, Dordrecht, 1996.
- [23] C. E. Akbas, A. Bozkurt, M. T. Arslan, H. Aslanoglu, and A. E. Cetin. L1 norm based multiplication-free cosine similarity measures for big data analysis. In *IEEE Computational Intelligence for Multimedia Understanding (IWCIM)*, France, Nov. 2014.
- [24] R. A. Johnson and D.W. Wichern. *Applied multivariate analysis*. Prentice Hall, 2002.
- [25] F. Gewers, G. R. Ferreira, H. F. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa. Principal component analysis: A natural approach to data exploration. Researchgate, 2019. https://www.researchgate.net/publication/324454887_Principal_Component_Analysis_A_Natural_Approach_to_Data_Exploration. accessed 1-Oct-2020.