Cognitive Processing
https://doi.org/10.1007/s10339-020-00985-5

RESEARCH ARTICLE

Informational content of cosine and other similarities calculated from high-dimensional Conceptual Property Norm data

Enrique Canessa[1,3] · Sergio E. Chaigneau[1,2] · Sebastián Moreno[3] · Rodrigo Lagos[4]

Received: 22 April 2019 / Accepted: 1 July 2020

Abstract

To study concepts that are coded in language, researchers often collect lists of conceptual properties produced by human subjects. From these data, different measures can be computed. In particular, inter-concept similarity is an important variable used in experimental studies. Among possible similarity measures, the cosine of conceptual property frequency vectors seems to be a de facto standard. However, there is a lack of comparative studies that test the merit of different similarity measures when computed from property frequency data. The current work compares four different similarity measures (cosine, correlation, Euclidean and Chebyshev) and five different types of data structures. To that end, we compared the informational content (i.e., entropy) delivered by each of those $4 \times 5 = 20$ combinations, and used a clustering procedure as a concrete example of how informational content affects statistical analyses. Our results lead us to conclude that similarity measures computed from lower-dimensional data fare better than those calculated from higher-dimensional data, and suggest that researchers should be more aware of data sparseness and dimensionality, and their consequences for statistical analyses.

Keywords Cosine similarity · Euclidean distance · Chebyshev distance · Clustering · Conceptual properties

Introduction

Though concept similarity can be computed by other means (e.g., pairwise similarity ratings), a frequent procedure is to compute concept similarity from concept descriptions. This approach has been widely applied when studying concepts coded in language (e.g., the properties *barks*, *has four legs*, and *wags its tail*, describe the concept DOG). In the *property approach* (a.k.a. featural approach) to concept similarity, concepts are described by sets of properties, and similarity between concepts is a function of their properties' distributions (e.g., Shepard and Arabie 1979; Tversky 1977; Tversky and Hemenway 1984).

The procedure generally used to obtain property-based descriptions is the Property Listing Task (PLT, a.k.a. Feature Listing Task; e.g., Cree and McRae 2003; Hampton 1979; McRae et al. 2005; Rosch et al. 1976). This is a widely used task across psychology (e.g., in cognitive psychology, social psychology, cognitive neuroscience, neuropsychology, consumer psychology), in which subjects are asked to produce properties typically associated with a given concept, and their responses are coded into property types (i.e., responses with only superficial differences across subjects are coded as a single property). In Conceptual Property Norms (CPNs), the PLT is used to obtain descriptions for a large number of concepts (e.g., Devereux et al. 2014; Kremer and Baroni 2011; Lenci et al. 2013; McRae et al. 2005; Montefinese et al. 2013, 2015; Vivas et al. 2017). These norms can be represented

This article is part of the special topic 'Eliciting Semantic Properties: Methods and Applications' guest-edited by Barry Devereux, and Alessandro Lenci.

Handling editor: Alessandro Lenci; Reviewers: David Vinson (University College London), Cai Wingfield (Lancaster University).

Enrique Canessa
ecanessa@uai.cl

[1] Center for Cognition Research (CINCO), School of Psychology, Universidad Adolfo Ibáñez, Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile

[2] Center for Social and Cognitive Neuroscience, School of Psychology, Universidad Adolfo Ibáñez, Av. Presidente Errázuriz 3328, Las Condes, Santiago, Chile

[3] Faculty of Engineering and Science, Universidad Adolfo Ibáñez, Av. Padre Hurtado 750, Lote H, Viña del Mar, Chile

[4] Programa Magister en Bioestadística, Universidad de Chile, Santiago, Chile

Published online: 09 July 2020

as rectangular matrices containing different concepts with their respective properties' frequency distributions.

Researchers use CPN data in at least two different ways. First, CPNs may be used as a source of normed stimuli and of control variables for experiments (McRae et al. 1999; Bruffaerts et al. 2019). Second, CPNs provide information about the underlying semantic structure of a representative individual (e.g., showing that, on average, DOG and CAT are conceptually more similar to each other than either is to CUP), thus allowing researchers to test theories about the nature of concepts and conceptual content (e.g., Cree and McRae 2003; Rosch and Mervis 1975; Vigliocco et al. 2004; Wu and Barsalou 2009).

On those cases where researchers are interested in similarity between concepts, a specific similarity measure needs to be computed. Note here that similarity is intimately related to semantic distance (i.e., distance may be viewed as a measure of dissimilarity). In what follows, because our arguments apply to both similarities and distances, when we refer to distances, the reader should bear in mind that the same ideas could be discussed in terms of similarity (and vice versa).

Because cosine similarity is currently in wide use (e.g., Hutchison et al. 2008; Mandera et al. 2015; Recchia and Jones 2009; Simmons and Estes 2006), in particular when reporting results from large scale studies about concept meaning like the aforementioned CPN studies (e.g., Devereux et al. 2014; Kremer and Baroni 2011; Lenci et al. 2013; McRae et al. 2005; Montefinese et al. 2013, 2015; Vivas et al. 2017), cosine similarity features prominently in our analyses. In CPN studies, cosine similarity is typically computed from high-dimensionality property frequency vectors (as will become clear shortly).

Because other distance measures are possible, in the current work we compare different measures computed from matrices like those obtained in CPN studies. More specifically, we compare the relative merit of cosine, correlation, Euclidean and Chebyshev distances. Because those measures may all be computed from different types of vectors (as will become clear shortly), we use a variety of vectors to further our findings' generalizability. Importantly, we use high- and low-dimensionality vectors, and analyze the effect of dimensionality on the informational content (i.e., entropy, which we will formally introduce and define later on) of those data. Furthermore, we submit our different distance measures to clustering analysis and gauge their relative performances. Because clustering algorithms separate data into intuitively similar groups based on some distance measure, they are frequently used to analyze semantic meaning of concepts in cognitive research (e.g., Maki and Buchanan 2008; Verbeemen et al. 2007). For all our analyses, we resorted to the Centre for Speech, Language and the Brain concept property norms (from here and on, the CSLB norms; Devereux et al. 2014).

To foreshadow, our findings suggest that CPN researchers should be aware that their data's informational content (i.e., entropy) depends on a trade-off between their matrices' dimensionality and sparsity (i.e., high dimensionality comes at the expense of increased sparseness), and that informational content will impact the statistical usefulness of any distance or similarity measure that they prefer to compute. As noted by an anonymous reviewer, although we frame this work on CPN research, the conclusions may also apply more generally to other areas that use similar data (e.g., Latent Semantic Analysis, LSA; Landauer and Dumais 1997).

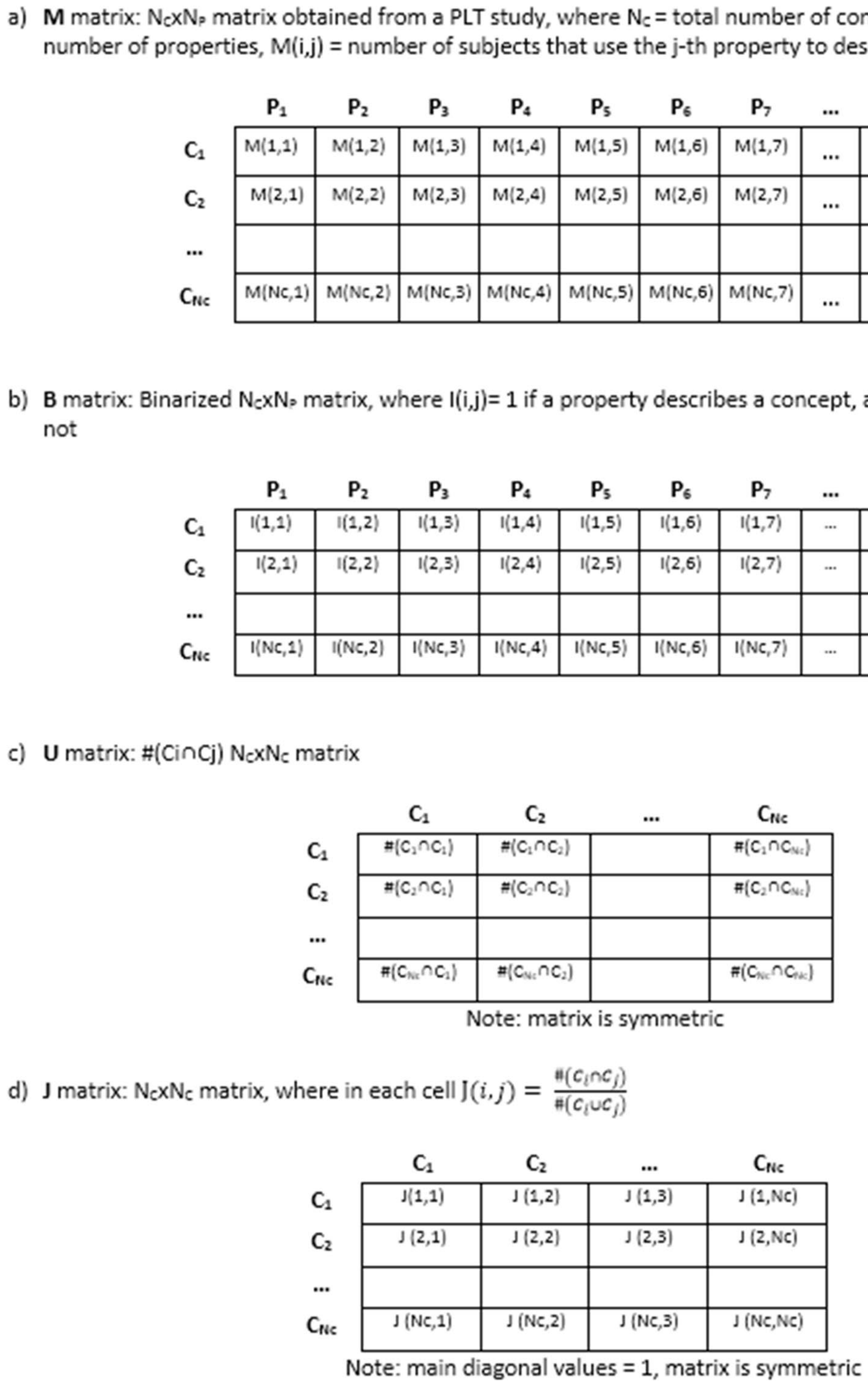
Computing concepts' distance

Data structure

As already mentioned, our data come from the CSLB norms (Devereux et al. 2014). In that study, semantic properties were collected for 638 concepts, thus leading to a 638 by 5542 unique properties matrix, with the properties' production frequencies in the matrix cells. More formally, data were arranged in a concept by property rectangular **M** matrix of size $N_C \times N_P$, where $N_C = 638$ and $N_P = 5542$. Values in the cells of that matrix reflected the frequency in which property P_a was mentioned for concept C_i by subjects in the sample (for similar data structures, see Cree and McRae 2003; De Deyne et al. 2008; McRae et al. 2005). In such a matrix, the i-th concept $\mathrm{C_i}$ is represented by a frequency vector, where $M(i,j)$ represents the number of subjects that use the j-th property to describe the i-th concept (see Fig. 1 panel a). For example, $M(3667) = 2$ corresponds to the third concept (ALLIGATOR) in the CSLB norms, where the conceptual property number 667 ("does live in swamps"), was mentioned by 2 subjects.

From an **M** matrix such as described above, researchers can compute a similarity or distance matrix for further analysis. Several methods exist to perform these computations from the original **M** matrix obtained from the PLT. As discussed above, an often used measure is the cosine of the **M** matrix, where concepts are treated as vectors based on their property production frequencies, and the cosine theta for each pair of vectors is used to build an $N_C \times N_C$ similarity matrix.

Consider, however, that other alternatives are possible. If frequencies are not of interest, original data in the **M** matrix could be binarized to produce a **B** matrix with a value of 1 if a property belongs to a concept, and a value of 0 if it does not (e.g., Brusco 2004; see Fig. 1 panel b). Because in a **B** matrix only property identity is of interest, the i-th concept $\mathrm{C_i}$ can be characterized by a set of



a) **M** matrix: N_CxN_P matrix obtained from a PLT study, where N_C = total number of concepts and N_P = total number of properties, M(i,j) = number of subjects that use the j-th property to describe the i-th concept

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	...	P_{Np}
C_1	M(1,1)	M(1,2)	M(1,3)	M(1,4)	M(1,5)	M(1,6)	M(1,7)	...	M(1,Np)
C_2	M(2,1)	M(2,2)	M(2,3)	M(2,4)	M(2,5)	M(2,6)	M(2,7)	...	M(2,Np)
...									
C_{Nc}	M(Nc,1)	M(Nc,2)	M(Nc,3)	M(Nc,4)	M(Nc,5)	M(Nc,6)	M(Nc,7)	...	M(Nc,Np)

b) **B** matrix: Binarized N_CxN_P matrix, where I(i,j)= 1 if a property describes a concept, and I(i,j)= 0 if it does not

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	...	P_{Np}
C_1	I(1,1)	I(1,2)	I(1,3)	I(1,4)	I(1,5)	I(1,6)	I(1,7)	...	I(1,Np)
C_2	I(2,1)	I(2,2)	I(2,3)	I(2,4)	I(2,5)	I(2,6)	I(2,7)	...	I(2,Np)
...									
C_{Nc}	I(Nc,1)	I(Nc,2)	I(Nc,3)	I(Nc,4)	I(Nc,5)	I(Nc,6)	I(Nc,7)	...	I(Nc,Np)

c) **U** matrix: #(Ci∩Cj) N_CxN_C matrix

	C_1	C_2	...	C_{Nc}
C_1	$\#(C_1 \cap C_1)$	$\#(C_1 \cap C_2)$		$\#(C_1 \cap C_{Nc})$
C_2	$\#(C_2 \cap C_1)$	$\#(C_2 \cap C_2)$		$\#(C_2 \cap C_{Nc})$
...				
C_{Nc}	$\#(C_{Nc} \cap C_1)$	$\#(C_{Nc} \cap C_2)$		$\#(C_{Nc} \cap C_{Nc})$

Note: matrix is symmetric

d) **J** matrix: N_CxN_C matrix, where in each cell $J(i,j) = \frac{\#(C_i \cap C_j)}{\#(C_i \cup C_j)}$

	C_1	C_2	...	C_{Nc}
C_1	J(1,1)	J (1,2)	J (1,3)	J (1,Nc)
C_2	J (2,1)	J (2,2)	J (2,3)	J (2,Nc)
...				
C_{Nc}	J (Nc,1)	J (Nc,2)	J (Nc,3)	J (Nc,Nc)

Note: main diagonal values = 1, matrix is symmetric

Fig. 1 Different matrices from which similarities and distances may be computed

properties defined by $C_i = \{P_j: M_{ij} > 0$ and $1 \leq j \leq N_P\}$, i.e., we use C_i to stand for a concept label, which defines a set of properties $\{P_a, P_b, \ldots, P_g\}$ that describe its meaning. For instance, $C_3 = \{P_{24}, P_{34}, \ldots, P_{667}, \ldots, P_{5216}\}$ implies that the third concept, ALLIGATOR, is represented by the properties 24, 34, …, 667, …, 5216; where, as noted above, conceptual property 667 corresponds to "does live in swamps."

Relatedly, if a researcher believes that shared properties play a central role in concept similarity (e.g., as in Tversky's 1977 Contrast Model), yet other matrices are possible. Two concepts may be similar to the extent that they share their describing properties. An **U** matrix (see Fig. 1 panel c) is an $N_C \times N_C$, square matrix with the number of concepts' property intersections in its cells (i.e., values for $\#(C_i \cap C_j)$). Similarly, a **J** matrix is an $N_C \times N_C$, square matrix where intersections in the cells (i.e., the $\#(C_i \cap C_j)$ values) are normalized relative to the total number of properties involved in each comparison (i.e., $\#(C_i \cup C_j)$). This measure is called Jaccard similarity (Jaccard 1901), and is closely related to Tversky's (1977) Contrast Model (see Eq. (1) and Fig. 1 panel d).

$$J\left(C_i, C_j\right) = \frac{\#\left(C_i \cap C_j\right)}{\#\left(C_i \cup C_j\right)}. \tag{1}$$

Distance measures

As already mentioned in the "Introduction" section, similarity is intimately related to semantic distance. In what follows, some of the mathematical expressions we use are easier to express and understand in terms of distances rather than similarities. Hence, to be consistent, we will use distances for all our measures. Note that, strictly speaking, not all distances we compute satisfy the mathematical conditions to be metric distances (e.g., triangle inequality). However, this is not relevant for characterizing our measures' distributions, nor for the clustering analyses we report. As discussed by Harary et al. (1965), clustering may be used with non-metric distances.

A popular strategy being currently used to measure conceptual distance from data structured in **M** matrices, is to consider each property as a dimension (i.e., an N_P-dimensional space) and each concept $\mathrm{C_i}$ as a vector. Then, distance can be thought of as one minus the cosine of those vectors, see Eq. (2) (conversely, similarity is just the cosine of the vectors). Thus, two concepts are maximally distant if the frequency ratio of their corresponding properties is totally different (i.e., $1 - \cos\,\Theta = 0$). Equation (2) shows this distance measure, which potentially ranges from 2 (opposite direction vectors) to 0 (identical direction vectors), with 1 indicating orthogonal vectors. Note that in frequency data (**M** matrices) there are no negative vectors, so $1 - \cos\,\Theta$ in fact ranges from 0 to 1.

$$d_{\mathrm{Cos}}\left(C_i, C_j\right) = 1 - \frac{\sum_{k=1}^{N_p} C(i,k) \cdot C(j,k)}{\sqrt{\sum_{k=1}^{N_p} C(i,k)^2}\sqrt{\sum_{k=1}^{N_p} C(j,k)^2}} \tag{2}$$

A related distance measure that we will include in our analyses is the vector correlation measure in Eq. (3) (see Dry and Storms 2009). This measure corresponds to one minus the statistical correlation between two random variables, potentially ranging from 0 to 2.

$$d_r\left(C_i, C_j\right) = 1 - \frac{\sum_{k=1}^{N_p}\left(C(i,k) - \overline{C_i}\right) \cdot \left(C(j,k) - \overline{C_j}\right)}{\sqrt{\sum_{k=1}^{N_p}\left(C(i,k) - \overline{C_i}\right)^2 \sum_{k=1}^{N_p}\left(C(j,k) - \overline{C_j}\right)^2}} \tag{3}$$

To their advantage, these measures are easy to compute and make few theoretical commitments. Note here that cosine and correlation distances can be also computed from **B**, **U** and **J** matrices, the difference being that, in the two latter cases spaces are N_C-dimensional (i.e., they typically have a lower dimensionality, given that the common case is that $N_C \ll N_P$). This issue of dimensionality will be important in our discussions further ahead.

Because, by using the cosine one appeals to a spatial view of semantic structure, other spatially based computations could also be used to express distance. Two popular measures used are Euclidian and Chebyshev distances. In the current work, we will use them to allow greater generality in our discussions.

In an N-dimensional space, two concepts' distances can be expressed as their Euclidean distance in the corresponding space. For two concepts C_i and C_j that have N_p attributes, the Euclidean distance is the square root of the sum of the squared difference in the respective N_p attributes, as Eq. (4) shows:

$$d_E\left(C_i, C_j\right) = \sqrt{\sum_{k=1}^{N_p} (C(i,k) - C(j,k))^2} \tag{4}$$

Because high dimensionality, as typically found in **M** matrices, can be problematic (as will be discussed shortly), we also computed Chebyshev distances. Chebyshev distance handles multidimensionality by using only the dimension in which two concepts are maximally different. Formally, the Chebyshev distance is:

$$d_{\mathrm{CHE}}\left(C_i, C_j\right) = \underset{k}{\mathrm{Max}}(|C(i,k) - C(j,k)|) \tag{5}$$

The interrelated problems of dimensionality and sparseness

The overall message we convey in this section is that the nature or quality of the data one uses will limit the analyses one can carry out. In our analyses, data quality is measured by the informational content found in those data's distributions. Furthermore, the informational content cannot increase indefinitely simply by increasing the number of variables being measured (i.e., this is the issue of dimensionality). Aside from the intrinsic problems of dimensionality that are discussed next, in the case of co-occurrence matrices, high dimensionality leads to sparse data. As we will discuss in the current section, though high-dimensional data seems to be intuitively more informative, there is a trade-off between dimensionality and sparseness, such that an upper limit to informational content exists. In vectors with many parameters, the likely case is that increasing the number of parameters does not necessarily increase their informational content.

Bellman (1961) noted that high dimensionality creates a problem for measurement, which he called "the curse of dimensionality." The main difficulty arising from high dimensionality is that distances between data points tend to become uniform as dimensionality increases. This has been shown by Beyer et al. (1999) and can be expressed more formally in the following equation:

$$\lim_{D\to\infty} \frac{MaxDist - MinDist}{MinDist} = 0 \tag{6}$$

where D = a measure's dimensionality, $MaxDist$ = the maximum distance between two points, and $MinDist$ = the minimum distance between two points. Note that Eq. (6) implies that distances between any two points in high-dimensional spaces will be practically the same, which means that distributions of distances computed from high-dimensionality data will exhibit low variability, and this is likely to be problematic for many statistical analysis techniques, as we will explain later on. In particular, Steinbach et al. (2004) argued that dimensionality is theoretically problematic for clustering.

A related problem is that matrices with many dimensions tend also to be sparse matrices (i.e., matrices with many empty cells), which makes them relatively uninformative. Particularly for CPN/PLT data, as many more concepts and properties are considered, the less likely it will be that any two concepts share those properties (i.e., increased sparseness). Note that this problem will only become worse as more concepts are considered for the analysis. Though it is intuitive that increasing the number of dimensions should increase informational content, dimensionality comes at the price of sparsity. In other words, there is a trade-off between dimensionality and sparsity regarding how informative data can be (Sahlgren 2006).

Because statistical analysis methods need data which convey information, sparse and non-variable data will generally be inappropriate for statistical analysis. Take for example regression analysis. If many data points are near each other (i.e., exhibit low variability), the degrees of freedom of the regression term will be low, making the data matrix almost singular. In turn, that will render unstable and difficult to interpret regression coefficients (Kleinbaum et al. 1988). Note that the issue at stake here is not what informational content implies for the nature of similarity itself, but instead what it implies regarding the statistical properties of variables computed with the goal of learning something about similarity.

Characterizing informational content and the sparsity to dimensionality relation in CPN data

For the present work, note that cosine and correlation measures calculated from property production frequencies (the **M** matrix, see Fig. 1 panel a) are based on high-dimensional data (N_P = 5542 dimensions). Our binarized measure, computed from the **B** matrix (see Fig. 1 panel b) has the same dimensionality. In contrast, our **U** and **J** matrices (Fig. 1 panels c and d) are relatively low-dimensional data (in our data, 638 dimensions).

Because, as we discuss in the "The interrelated problems of dimensionality and sparseness" section, high-dimensionality poses problems for statistical analyses, a possible solution is to perform some form of dimensionality reduction (see, e.g., Latent Semantic Analysis, LSA; Landauer and Dumais 1997). Thus, to handle that problem we submitted our CSLB **M** matrix to Principal Components Analysis (PCA). Note that this is a transformation that researchers could easily perform on CPN data, which is why we deemed interesting to examine its effects. Using PCA with varimax rotation and extracting those dimensions with eigenvalues ≥ 1.0, we obtained 364 dimensions (down from 5542). The rotated dimensions explained 94% of the original data variance. Note that the relatively small number of extracted dimensions compared with the original number of dimensions, and the large proportion of explained variance, attests to the **M** matrix's sparsity.

Given that we had five different data matrices from which to compute distances and four different distance computation methods, we were able to characterize the distributions produced by each of the twenty combinations (5 data matrices times 4 distances). Figure 2 shows these distributions.

Visual inspection of the distributions in Fig. 2, allows some clear conclusions. First, it is rather obvious that distances computed from high-dimensional data exhibit, in general, highly skewed and low variability distributions. It is also obvious that distances calculated from low-dimensional data show more uniform distributions. Per our discussion above, the distributions in Fig. 2 imply that distances obtained from low-dimensional data have more informational content than those computed from high-dimensional data. As a way of quantifying this, we calculated each distribution's entropy (shown on top of each graph in Fig. 2). Higher entropy values indicate that a distribution conveys more informational content (i.e., that it exhibits more potentially useful variability). Entropy is given by Eq. (7):

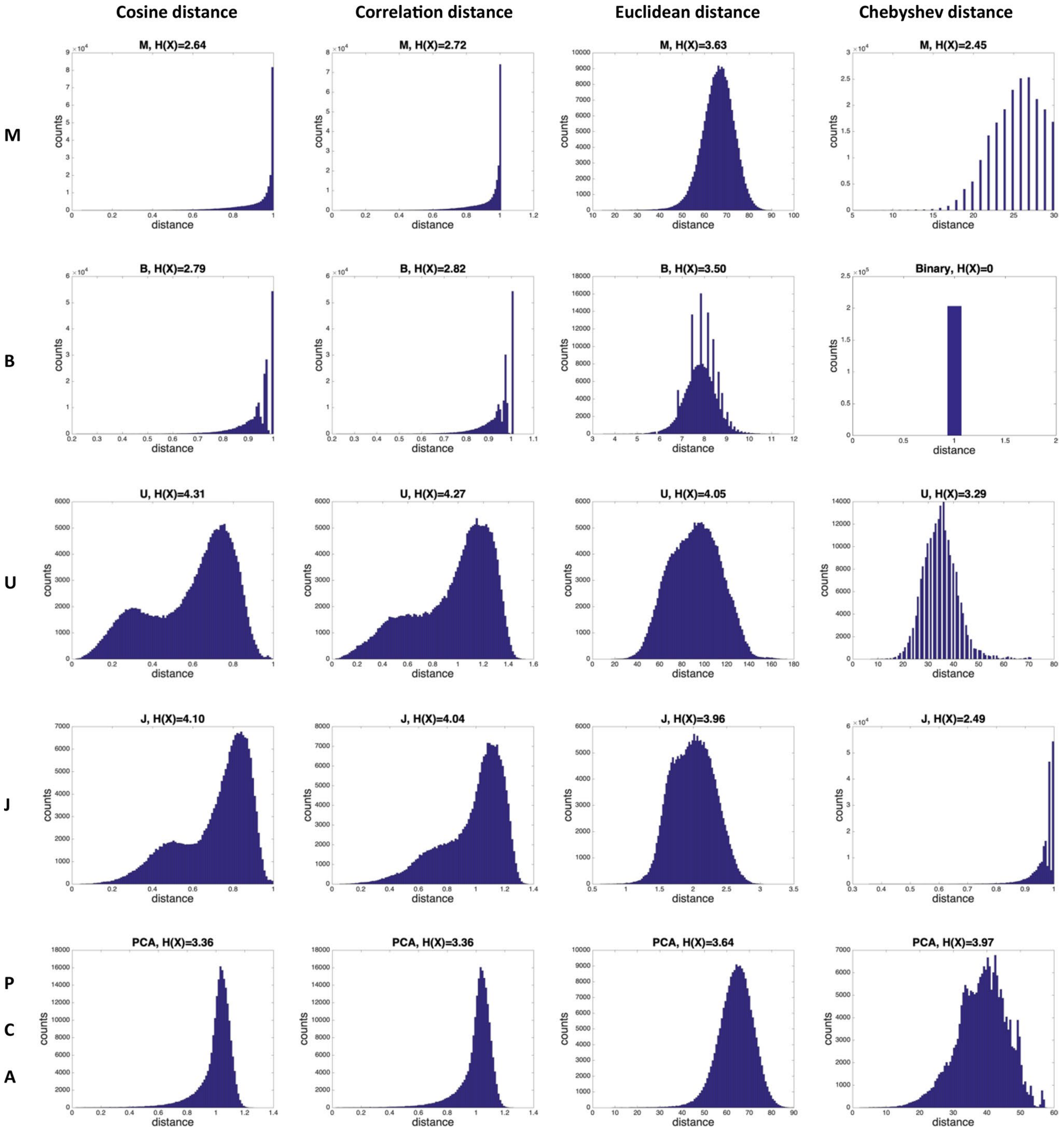


Fig. 2 Histograms showing distributions of four different distance measures applied to five different data matrices and their corresponding entropies. The *X*-axis represents distances computed from the respective data matrices

$$H = \sum_{k=1}^{m} p_k \log_m \left(\frac{1}{p_k} \right) \qquad (7)$$

where p_k is the probability associated to each of the m distances. As Eq. (7) shows, entropy is based on the probability of each possible value, instead of the value itself. Then, to make a fair comparison among all cases, each distribution was estimated using 100 possible values (the same number of bins for the histograms), i.e., $m = 100$ in Eq. (7).

Note that entropy values in Fig. 2 tell the same story as the visual inspection of the corresponding distributions shows. For all computed distances, there is a deleterious effect of dimensionality, such that the highest entropies (i.e., best distributions) are found for **U** and **J** matrices, followed by the **PCA** matrix, and then the high-dimensional **M** and **B** matrices. Notably, though PCA increases informational content, it does not achieve the informational content of **U** and **J** matrices. The only exception to this trend is Chebyshev distance, where **PCA** data exhibits a higher entropy than **U** and **J** data.

Additionally, the reader may notice that entropy values for cosine and correlation distances in Fig. 2 are quite similar. This occurs because correlation and cosine distances are in fact closely related. If means are computed and subtracted from each cell (i.e., thus leaving no empty cells), and if cosine distances are computed from the new matrix, the resulting cosines are in fact the correlation values computed from Eq. (3). As an anonymous reviewer noted, this may be considered a way of removing sparseness. However, this achieves only a nominal removal of sparseness (i.e., there are no more zero frequency cells). It is easy to see that the new nonzero values add no information. The slight differences in entropy values that the reader may note in Fig. 2 are only due to details in the algorithm that does the computations, and for all intents and purposes entropy values should be considered to be the same for cosine and correlation matrices.

Regarding the relation between sparsity and dimensionality, Fig. 3 illustrates their relation in our data. To generate this graph, we started by graphing dimensions exhibiting the lowest sparsity (lowest percentage of zeros), and then continued including dimensions with higher sparsity. Note that the increase in sparsity as the number of dimensions grows, is larger for the **M** and **B** data, than for **U** and **J**. Also, note that sparsity is always proportionally much smaller for **U** and **J** data than for **M** and **B** data.

Finally, given that Principal Component Analysis reduces dimensionality so that all extracted dimensions explain some variance of the data, the **PCA** matrix does not have zeros and thus, its sparsity is always zero (i.e., there are no zero values in their cells). However, as we discuss next, this does not automatically guarantee its informational content.

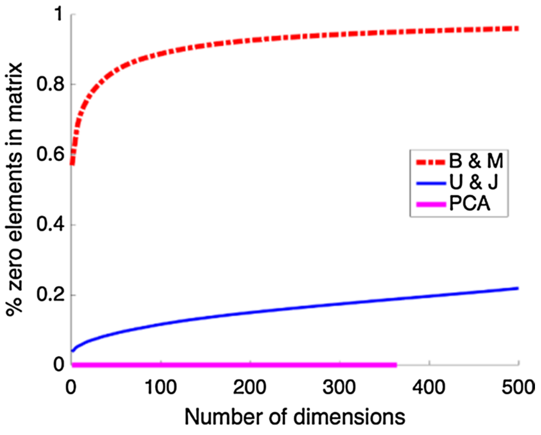


Fig. 3 Sparsity (% of zeros in corresponding distance matrix) computed for different number of dimensions (up to 500, and 364 for PCA)

To further elucidate the effect of dimensionality on our data's entropy values (i.e., informational content), Fig. 4 graphs that relationship for the four distances considered in this work across the five matrices, considering the first 500 dimensions (i.e., to have a fair comparison, and note also that **PCA** has only 364 dimensions). To produce each of the graphs, we used a similar strategy used to build Fig. 3. We iteratively computed entropy, progressively including dimensions starting with those exhibiting the lowest sparsity (i.e., lowest percentage of zeros) and moving toward those with higher sparsity. In the particular case of **PCA**, given that there are no zeros in cells, we kept the order defined by PCA's output (i.e., descending order based on the percentage of variance explained by each dimension). As shown in Fig. 4, as the number of dimensions increases, entropy tends to increase and then levels out or decreases. Graphs in Fig. 4 clearly show the trade-off between dimensionality and sparsity.

An analysis of Fig. 4 reveals interesting patterns. First, irrespective of the distance being computed (i.e., cosine, correlation, Euclidean), **U** and **J** matrices show higher entropy values, while **M** and **B** matrices show the lowest. The only exception is when computing distances from low dimensionality **PCA** vectors. However, distances computed on the **PCA** matrix lead to a large reduction in entropy as dimensions increase further, particularly for cosine and correlation distances. Second, for Chebyshev distance, entropy noticeably remains higher for **PCA** than for **U** and **J** matrices. Given that for **PCA** all dimensions have some informational content, and that Chebyshev distance uses only the dimension in which two concepts are maximally different, the informational content conveyed by dimensions in **PCA** remains relatively high for the first dimensions (i.e., most of the informational content is conveyed in the first dimensions).

In conclusion, all our comparisons show that distances computed from **M** matrix data (property frequency vectors) tend to be high-dimensional and sparse, something which decreases their informational content. Reducing dimensions through **PCA** improves their behavior, but not as much as might be presumed. In contrast, distances computed from **U** and **J** matrices ($N_c \times N_c$ matrices that disregard frequency information) produce noticeably better distributions (higher informational content, i.e., entropy). Summarizing, our analyses clearly show that sparseness increases with dimensionality (Fig. 3), such that beyond certain cut-off point entropy drops-off (Fig. 4), which makes matrices less useful from a statistical point of view, as visually illustrated by distributions and entropy values in Fig. 2.

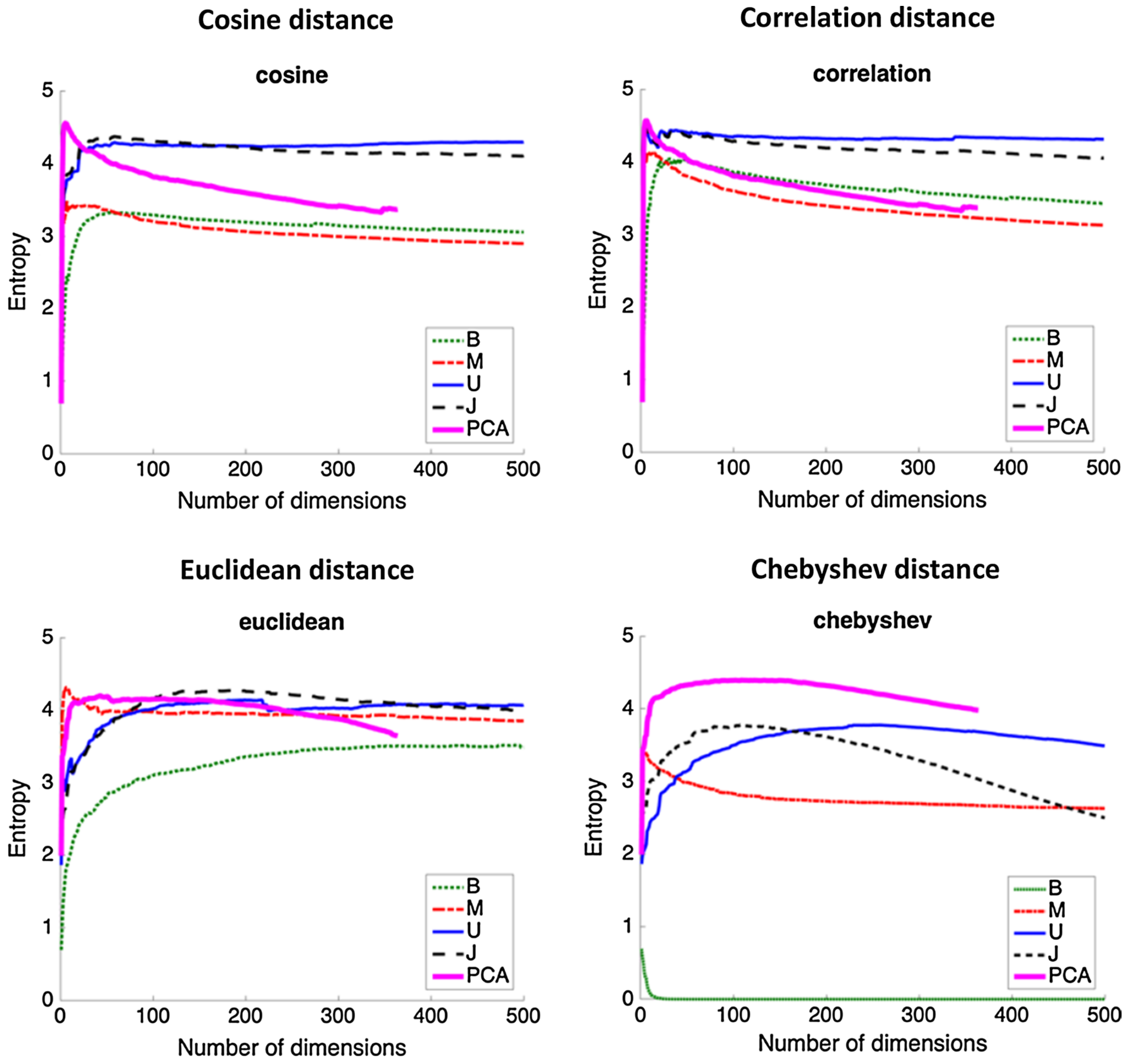


Fig. 4 Entropy versus dimensionality for four distances computed across **B**, **M**, **U**, **J** and **PCA** matrices

Clustering

Given that clustering algorithms (Johnson 1967; Shepard and Arabie 1979) separate data into intuitively similar groups based on some similarity or distance measure, they can be used to analyze similarity data obtained through the PLT. Across psychology, clustering is often used to uncover the similarity structure underlying a set of concepts (e.g., Maki and Buchanan 2008; Verbeemen et al. 2007). Thus, we resorted to clustering analysis to show a concrete example of how distance measures computed from high- and low-dimensional data affect the quality of the corresponding solutions.

Clustering algorithms can be roughly classified into hard or soft clustering. Hard clustering creates a partition of the data, where each data point can belong to only one cluster. Some common hard algorithms are hierarchical and K-means clustering (Johnson 1967; MacQueen 1967). In contrast, soft clustering allows data points to belong to more than one cluster. Some common soft algorithms are additive and fuzzy clustering (Dunn 1973; Shepard and Arabie 1979). However, given that results from soft clustering algorithms are notoriously hard to interpret (see Wilderjans et al. 2011), our analyses focus on Agglomerative Hierarchical Clustering (AHC).

Clustering solutions' quantitative analysis

As a way to judge the goodness of clustering when using the AHC algorithm, results were compared by using the Silhouette Coefficient (SC) and the Cophenetic Correlation coefficient (CC). As will become clear next, the SC is related to the goal of obtaining high within-group and low between-group similarity. The CC provides an estimate of how much variance in the similarity data is accounted for by distances in the clustering solution. A good similarity or distance measure should reflect on these two indices.

The SC of a concept varies between − 1 and 1, and measures the goodness of each concept with respect to the assigned cluster. If the distance of a concept with respect to the other concepts of the same cluster is smaller than the distance with respect to other clusters, then its SC value is greater than 0. Otherwise, SC is less than 0 and implies that the concept could as well belong to another cluster. Note that a cluster consisting of a single element has an SC value of 1, adding an important bias to the average SC. For that reason, we omitted single element clusters when computing SC.

The CC measures the correlation between the distance among concepts based on the dendrogram and the actual distance used in the hierarchical clustering process. Given a dendrogram and two concepts, the distance between them corresponds to the height of the dendrogram where these two concepts are joined. As with other correlation measures, the CC varies between − 1 and 1, where a high value implies a better representation of the original distances between the concepts, leading to a better clustering of the data.

Several methods have been proposed to compute distances between clusters (see Kuiper and Fisher 1975). In the current work, we use complete and average linkages. In complete linkage (CL), the distance between two clusters is determined by the largest distance among all concepts belonging to those clusters. This method generates spherical clusters, and is robust to noisy points (Aggarwal 2015). In average linkage (AL), the distance between two clusters is determined by the average distance between all pairs of concepts, where each pair is made up of one concept from each group (Aggarwal 2015).

Given that we used four distance measures (cosine, correlation, Euclidean, Chebyshev) computed from five different data matrices (**M**, **B**, **U**, **J** and **PCA**), and that we used two different methods to compute distances between clusters (CL and AL), in principle, we could compare clustering solutions from all those forty possible combinations. However, to reduce those combinations to a more manageable number, we decided to use the four distance measures (cosine, correlation, Euclidean, Chebyshev), with the data matrix that renders the overall highest entropy (**U**). We also included **PCA** because of its relevance for discussing the issue of dimensionality and the **M** matrix, because it corresponds to the raw CPN data. Those 3 (matrices) × 4 (distances) × 2 (CL or AL) = 24 combinations allowed us comparing the effect of high and low-dimensional data on clustering performance.

Accordingly, Fig. 5 shows the Silhouette Coefficient (SC) for our 24 combinations. As can be seen, low-dimensional data (**U** and **PCA** matrices) obtain better clustering solutions than high-dimensional data (**M** matrix). In particular, **U** data consistently fares better than the other two, except for Chebyshev distance and average linkage, where **PCA** data renders a higher SC. Comparing Fig. 5 with the corresponding entropies displayed in Fig. 2, we can see that both results are consistent.

Generally, a higher-entropy distance measure renders a higher-SC clustering solution. And consequently, a lower-dimensional distance measure renders a higher-SC solution than a higher-dimensional measure.

Table 1 presents the Cophenetic Coefficient (CC) for the same 24 combinations, and shows the same results as before. The **U** data matrix exhibits higher CC than the other two data matrices, except for **PCA** with Chebyshev distance and

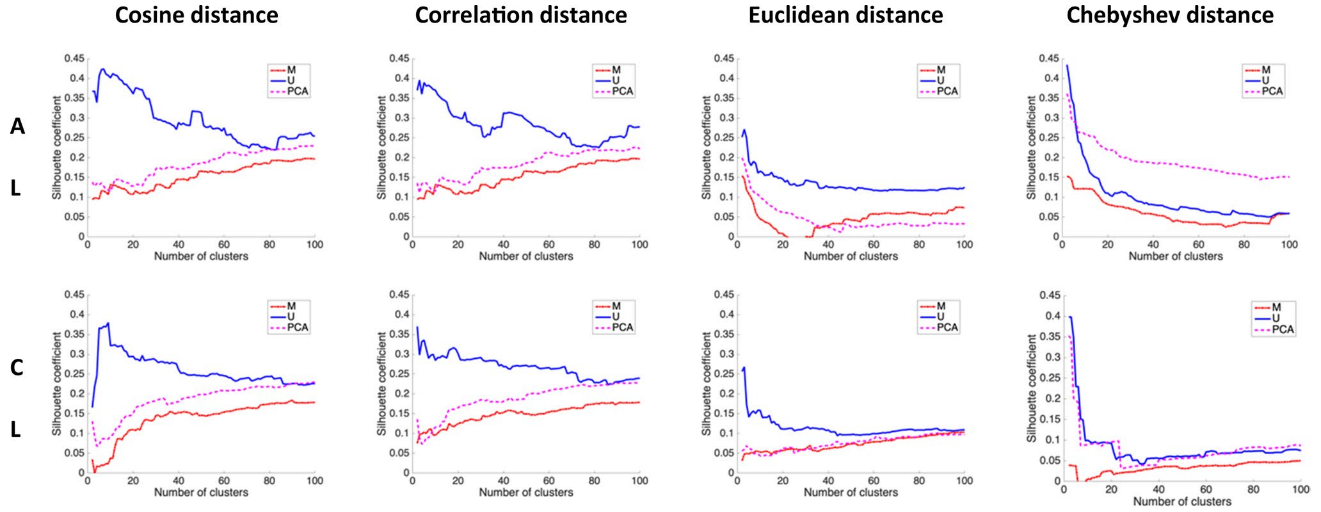


Fig. 5 Silhouette Coefficient (SC) for four different distance measures, three different data matrices (**M**, **U** and **PCA**), and average (AL) and complete linkage (CL)

Table 1 Cophenetic Coefficients (CC) for different similarity measures

	Complete linkage				Average linkage			
	Cosine	Correlation	Euclidean	Chebyshev	Cosine	Correlation	Euclidean	Chebyshev
M	0.62	0.62	0.42	0.47	0.84	0.84	0.61	0.64
U	0.78	0.87	0.77	0.79	0.88	0.91	0.80	0.86
PCA	0.68	0.68	0.37	0.68	0.81	0.80	0.70	0.89

Higher values imply more similarity variance explained. Frequency = **M** Matrix; #(C_i∩C_j) = **U** matrix; PCA = **PCA** matrix

AL. In that case, **U**'s CC (0.86) is smaller than **PCA**'s CC (0.89), which agrees with **PCA**'s higher entropy relative to **U** (respectively, 3.97 and 3.29).

In conclusion, the quantitative analysis of the clustering solutions show that distances calculated from low-dimensional data produce better clusters (in terms of SC and CC) than distances computed from high-dimensional data. Overall, distances obtained from the **U** data matrix fare better than the other alternatives.

Clustering solutions' qualitative analysis

To complement the quantitative analysis, here we present a brief examination of those concepts that belong to each cluster for some of the 24 combinations included in the "Clustering solutions' quantitative analysis" section. To keep this section within a reasonable size, rather than providing and exhaustive list, Table 2 presents the number of concepts per cluster and an intuitive name for each.

To compute the clustering solutions presented in Table 2, we used AHC with complete linkage, as a way of avoiding clusters with only one concept. Additionally, we requested 18 clusters. The SC in Fig. 5 shows that we obtain a maximum SC when using approximately between 5 and 18 clusters. We decided to use the upper limit (18 clusters) to avoid obtaining large clusters with many concepts, something which was likely to make their interpretation difficult. The clustering solutions displayed in Table 2 were selected based on the CC shown in Table 1. We included the clustering solutions corresponding to all the distances calculated from the **U** matrix that had the highest CC for complete linkage. To compare those solutions to other ones, we also used the **M** and **PCA** data with Cosine and Correlation distances, because they exhibited a higher CC than Euclidean and Chebyshev distances.

Though qualitative interpretations of clustering solutions are highly dependent on theoretical assumptions (e.g., if clusters are expected to reveal a linguistic superordinate organization or perceptual-based groupings), we believe that discussion of our solutions' qualitative structure is useful. It would be surprising that different solutions produce the same qualitative information, regardless of their quantitative goodness of clustering. With this caveat in mind, several things are noteworthy regarding Table 2. First, there seems to be a greater tendency for distances computed from the **M** matrix (frequency vectors) to produce unnatural clusters, relative to clusters generated from distances obtained from the **U** matrix. In typical clusters, it is possible to naturally capture them with some higher level abstraction (e.g., superordinates like mammals, furniture, vehicles, fruits, food). This may not be possible with other less natural clusters. Consider the difference between the "Liquids" cluster for correlation distance computed from the **U** matrix and the "Drinkables" cluster for cosine distance computed from the **U** matrix. The "Liquids" cluster includes drinkable liquids such as cider and coffee, but also non-drinkable liquids such as glue, nail-polish and perfume. There is no superordinate category that naturally captures all these subordinates. In contrast, the "Drinkables" cluster includes only liquids that can also be drunk, such as beer, coffee, gin, lemonade and milk. Intuitively, this is a much more natural cluster.

Other examples may help clarifying what we mean by cluster naturalness. Consider the "Flying things" cluster for cosine distance with **M** matrix, which includes otherwise dissimilar things such as "aeroplane", "hummingbird" and "moth", and the "Dangerous things" cluster for correlation distance with **M** matrix, which includes otherwise dissimilar things such as "cigar", "poison", "revolver" and "rattlesnake". This issue of cluster unnaturalness seems to be reduced with PCA, given that difficult to interpret clusters are less when using the **PCA** matrix. Some unnatural clusters are also found for the **U** matrix, though less frequently.

As Table 2 illustrates, distances computed from the **M** matrix seem to produce clusters that are harder to intuitively understand (the ND clusters in Table 2), even to a greater extent than the unnatural clusters mentioned above, for which some grouping principle still transpires. Inevitably, intuitiveness is a subjective judgment. However, we believe anyone would be challenged if trying to find the underlying principle that clusters together the following concepts: "dice", "harmonica", "sledge", "swing" and "umbrella" which are part of cluster 5 for correlation distance computed from the **PCA** matrix.

Finally, it is worth noting that as our quantitative analyses showed, Chebyshev distance produced poor results for our CPN data. Chebyshev distance computed from the **U** matrix

Table 2 Clustering solutions for different distance measures computed from different data matrices

Cluster	Correl. U	Cosine U	Euclidean U	Chebysh. U	Correl. PCA	Cosine PCA	Correl. M	Cosine M
1	Smoking related—3	Heart—1	Body parts—2	Coin—1	Means of Transport.—8	Means of Transport.—8	*Scents—3	*Scents—3
2	Prescription drugs—5	Certificate—1	Reptiles—6	Wetsuit—1	Marine Transport.—9	Marine Transport.—9	*Big things—5	*Red things—7
3	*Hard things—5	Buildings—2	Fish—8	Horse—1	*Dangerous things—11	*Long things—13	*Red things—7	*Glass-made things—8
4	ND—6	Prescription drugs—5	Means of Transport.—10	Television—1	*Round things—14	ND—14	Body parts—8	Body parts—8
5	ND—8	*Scents—5	Drinkables—19	Tent—1	ND—19	ND—18	*Long things—12	Drinkables—11
6	Printed—10	*Flat things—7	Animals—24	Eye—1	Musical instrument—20	Musical instrument—20	Printed—12	*Round things—11
7	*Things to rest on—11	Body parts—7	Flora—25	Large reptiles—2	*Green things—28	ND—22	*Dangerous things—14	Printed—12
8	*Liquids—13	Body parts—9	ND—28	Marine mammals—3	Flora—29	*Green things—28	ND—28	*Long things—12
9	Body parts—14	Printed—10	Food—31	ND—6	Means of Transport.—29	Means of Transport.—29	ND—32	*Dangerous things—14
10	Outdoor activities—17	Drinkables—11	Animals—34	Body parts—7	ND—31	ND—31	Marine creatures—33	Musical instrument—21
11	Musical instrument—24	Fish—24	Musical instrument—36	Printed—7	Marine creatures—32	Marine creatures—32	ND—38	Marine creatures—33
12	Flora—25	Musical instrument—26	Insects & crustaceans—37	Fruits—7	Clothing—36	ND—33	Food—46	Food—35
13	Marine creatures & insects—30	Flora—26	Birds—37	Animals—8	Birds—37	Birds—37	Fruits & flowers—48	Clothing—48
14	Clothing—45	ND—33	Means of Transport.—39	Animals—9	ND—46	Clothing & flowers—51	Clothing—48	Fruits & flowers—48
15	Means of Transport.—50	Clothing—46	Clothing—39	Cars—9	ND—67	Animals & insects—68	*Flying things—49	Flying things—49
16	Food—86	Food—84	Fruits—49	Animals—10	Animals—68	Food—72	ND—57	ND—57
17	Animals & insects—122	Animals & insects—128	ND—58	ND—15	Food—72	*Metal things—74	Animals & insects—66	ND—115
18	Other—164	Other—213	Other—156	Other—549	Other—82	Other—79	Other—132	Other—146

Each cell shows: an intuitive descriptor of the cluster and the number of concepts that belong to the cluster. * = unnatural clusters. ND = no sensible descriptor was found for the cluster. The "other" cluster corresponds to a large cluster with no sensible descriptor. Concepts belonging to each cluster can be found at https://osf.io/ezy3h/

produced more single concept clusters (see Table 2), and also a substantially larger "other" cluster than the competing solutions, showing that clustering based on Chebyshev distance is noticeably worse than its counterparts. Incidentally, given that entropy for the Chebyshev distances computed from the **B** matrix is zero, we also submitted those distances to clustering, requesting 18 clusters (the same procedure used in creating Table 2). As expected, the clustering solution simply listed one cluster of 621 concepts (the first 621 concepts in alphabetical order, given that this was the order of the concepts in the distance matrix), and 17 other clusters of one concept each. Those 17 concepts corresponded to the last 17 concepts in alphabetical order. Hence, that solution makes no substantive sense and explains no variance in the data (in fact the statistical package was not able to calculate the CC).

In summary, our quantitative and qualitative analyses for our clustering solutions suggest that computing distances from the **U** matrix tends to avoid unnatural clusters, leading to better clusters that are easier to interpret. In contrast, distances computed from the **M** matrix produce less tightly clustered solutions that tend to be less intuitively organized and are harder to interpret. In general, this section shows that entropy values could be used to predict different distance measures' performance in clustering analyses. Presumably, entropy and the dimensionality/sparsity trade-off would also predict similarity data's usefulness for other kinds of statistical analyses.

Discussion

In the current work, we examined the relative merits of different distances computed from data matrices obtained from CPNs. Due to its widespread use, we included a cosine measure, and contrasted it with other available measures (i.e., correlation, Euclidean and Chebyshev distances). Because, as we have argued, a measure's dimensionality poses problems for statistical analyses, to compute our distance measures we used traditional high-dimensional property frequency vectors from CPN studies. To illustrate their performance, we compared the traditional data with other lower-dimensionality data derived from the same frequency matrix. Lower dimensionality was achieved by computing two set-theoretic matrices that register shared properties rather than property frequencies, and also by applying PCA to the property frequency vectors data.

We performed two main analyses. First, we characterized distances' probability distributions, and showed that distances computed from high-dimensionality data exhibit highly skewed distributions that are low in entropy relative to the same measures computed from lower-dimensionality data. Though our results are directly pertinent to CPN data in the form of property frequency vectors, they are also relevant for researchers who use high-dimensional data in general. Perhaps the most important result we show is that researchers need to consider their data's informational content to guide their data processing. Researchers need to consider that the apparent advantages of high dimensionality come at the cost of losing informational content contained in variability and also of increasing sparseness. Also, our analyses suggest that dimensionality reduction techniques (i.e., in this particular case, PCA) tended to ameliorate these problems, but not as much as one would expect.

A second analysis we performed was submitting our different distance measures to the AHC algorithm. Clustering was chosen as a way of illustrating the problems that dimensionality will create for statistical analyses. Our quantitative comparison of clustering efficiency showed that distances calculated from low-dimensional data produce better clusters than distances computed from high-dimensional data. Overall, distances obtained from the **U** data matrix (our set-theoretic matrix that disregards property frequencies) fare better than the other alternatives. Our qualitative evaluation of clustering solutions suggests that computing distances from the low-dimensionality **U** matrix tends to avoid unnatural clusters and leads to clusters that are easier to interpret. In contrast, distances computed from the high-dimensionality **M** matrix produce clustering solutions that are less natural (e.g., *dangerous things* clusters "rattlesnake" and "gun"; *green things* clusters "cucumber" and "frog") and harder to interpret.

On closing, the current work provides an example of how dimensionality may impact results in cognitive studies. A more ambitious endeavor would be to empirically test the statistical effect of dimensionality on the ability of different measures to predict cognitive performance. We hypothesize that a variable's entropy and dimensionality are important factors affecting its predictive power. Further work awaits to test this. Importantly, the current work suggests that producers and consumers of high-dimensional data should concern themselves with the issue of entropy. There may be situations in which this is not a problem, but for those wanting to use their data for predicting psychological dependent variables, a low entropy predictor is bound to be problematic. In the particular case of CPN studies, were similarity is frequently computed from property frequency vectors (and, furthermore, where the cosine is used as the de facto similarity measure), we suggest that researchers should routinely report entropy values for similarity computations. Also, given that entropy values are relative to the number of bins used, we suggest that researchers should report that number, or even better, adopt a convention regarding the number of bins. Also, perhaps the current results could convince some researchers to use different procedures when computing distances and similarities from CPN data. Finally, note that many of these conclusions may apply not only to CPN research, but more generally to research areas that use high-dimensional data (e.g., Latent Semantic Analysis, LSA; Landauer and Dumais 1997), which is why we think that this work may be useful beyond our main CPN focus.

Acknowledgements We want to thank two anonymous reviewers for their useful comments to a previous version of this manuscript. We also want to thank Eyal Sagi for his valuable input regarding the ideas discussed here. This research was carried out with funds provided by grant 1200139 from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) of the Chilean government to the first and second authors.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Aggarwal CC (2015) Data mining: the textbook. Springer, Cham. https://doi.org/10.1007/978-3-319-14142-8

Bellman R (1961) Adaptive control processes: a guided tour. Princeton University Press, Princeton. NJ

Beyer K, Goldstein J, Ramakrishnan R, Shaft U (1999) When is "nearest neighbor" meaningful? In: Beeri C, Buneman P (eds) Database theory—ICDT'99. ICDT 1999. Lecture Notes in Computer Science, vol 1540. Springer, Berlin. https://doi.org/10.1007/3-540-49257-7_15

Bruffaerts R, De Deyne S, Meersmans K, Liuzzi A, Storms G, Vandenberghe R (2019) Redefining the resolution of semantic knowledge in the brain: advances made by the introduction of models of semantics in neuroimaging. Neurosci Biobehav Rev 103:3–13. https://doi.org/10.1016/j.neubiorev.2019.05.015

Brusco MJ (2004) Clustering binary data in the presence of masking variables. Psychol Methods 9(4):510–523. https://doi.org/10.1037/1082-989X.9.4.510

Cree GS, McRae K (2003) Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). J Exp Psychol: Gen 132(2):163–201. https://doi.org/10.1037/0096-3445.132.2.163

De Deyne S, Verheyen S, Ameel E, Vanpaemel W, Dry MJ, Voorspoels W, Storms G (2008) Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. Behav Res Methods 40(4):1030–1048. https://doi.org/10.3758/brm.40.4.1030

Devereux BJ, Tyler LK, Geertzen J, Randall B (2014) The Centre for Speech, Language and the Brain (CSLB) concept property norms. Behav Res Methods 46(4):1119–1127. https://doi.org/10.3758/s13428-013-0420-4

Dry MJ, Storms G (2009) Similar but not the same: a comparison of the utility of directly rated and feature-based similarity measures for generating spatial models of conceptual data. Behav Res Methods 41(3):889–900. https://doi.org/10.3758/brm.41.3.889

Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J Cybern 3(3):32–57. https://doi.org/10.1080/01969727308546046

Hampton JA (1979) Polymorphous concepts in semantic memory. J Verbal Learn Verbal Behav 18(4):441–461. https://doi.org/10.1016/s0022-5371(79)90246-9

Harary F, Norman RA, Cartwright D (1965) Structural models: an introduction to the theory of directed graphs. Wiley, New York, NY

Hutchison KA, Balota DA, Cortese MJ, Watson JM (2008) Predicting semantic priming at the item level. Q J Exp Psychol 61(7):1036–1066. https://doi.org/10.1080/17470210701438111

Jaccard P (1901) Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. Bull de la Soc Vaud des Sci Nat 37:241–272

Johnson SC (1967) Hierarchical clustering schemes. Psychometrika 32(3):241–254. https://doi.org/10.1007/BF02289588

Kleinbaum DG, Kupper LL, Muller KE (1988) Applied regression analysis and other multivariate methods. PWS-Kent Publishing Co, Boston

Kremer G, Baroni M (2011) A set of semantic norms for German and Italian. Behav Res Methods 43(1):97–109. https://doi.org/10.3758/s13428-010-0028-x

Kuiper FK, Fisher L (1975) A Monte Carlo comparison of six clustering procedures. Biometrics 31(3):777–783. https://doi.org/10.2307/2529565

Landauer TK, Dumais ST (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol Rev 104(2):211–240. https://doi.org/10.1037/0033-295X.104.2.211

Lenci A, Baroni M, Cazzolli G, Marotta G (2013) BLIND: a set of semantic feature norms from the congenitally blind. Behav Res Methods 45(4):1218–1233. https://doi.org/10.3758/s13428-013-0323-4

MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Cam LML, Neyman J (eds) Proceedings of 5th Berkeley symposium on mathematical statistics and probability. University of California Press, Berkeley, pp 281–297

Maki WS, Buchanan E (2008) Latent structure in measures of associative, semantic, and thematic knowledge. Psychon Bull Rev 15(3):598–603. https://doi.org/10.3758/PBR.15.3.598

Mandera P, Keuleers E, Brysbaert M (2015) How useful are corpus-based methods for extrapolating psycholinguistic variables? Q J Exp Psychol 68(8):1623–1642. https://doi.org/10.1080/17470218.2014.988735

McRae K, Cree GS, Westmacott R, Sa VRD (1999) Further evidence for feature correlations in semantic memory. Can J Exp rimental Psychol/Revue canadienne de psychologie expérimentale 53(4):360–373. https://doi.org/10.1037/h0087323

McRae K, Cree GS, Seidenberg MS, Mcnorgan C (2005) Semantic feature production norms for a large set of living and nonliving things. Behav Res Methods 37(4):547–559. https://doi.org/10.3758/bf03192726

Montefinese M, Ambrosini E, Fairfield B, Mammarella N (2013) Semantic memory: a feature-based analysis and new norms for Italian. Behav Res Methods 45(2):440–461. https://doi.org/10.3758/s13428-012-0263-4

Montefinese M, Zannino GD, Ambrosini E (2015) Semantic similarity between old and new items produces false alarms in recognition memory. Psychol Res 79(5):785–794. https://doi.org/10.1007/s00426-014-0615-z

Recchia G, Jones MN (2009) More data trumps smarter algorithms: comparing pointwise mutual information with latent semantic analysis. Behav Res Methods 41(3):647–656. https://doi.org/10.3758/BRM.41.3.647

Rosch E, Mervis CB (1975) Family resemblances: studies in the internal structure of categories. Cogn Psychol 7:573–605

Rosch E, Mervis CB, Gray WD, Johnson DM, Boyes-Braem P (1976) Basic objects in natural categories. Cogn Psychol 8(3):382–439. https://doi.org/10.1016/0010-0285(76)90013-x

Sahlgren M (2006) The word-space model: using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces. Ph.D. Dissertation, Department of Linguistics, Stockholm University

Shepard RN, Arabie P (1979) Additive clustering: representation of similarities as combinations of discrete overlapping properties. Psychol Rev 86(2):87. https://doi.org/10.1037/0033-295X.86.2.87

Simmons S, Estes Z (2006) Using latent semantic analysis to estimate similarity. In: Proceedings of the 28th annual conference of the cognitive science society, Austin, TX, pp 2169–2173

Steinbach M, Ertöz L, Kumar V (2004) The challenges of clustering high dimensional data. In: Wille LT (ed) New directions in statistical physics: econophysics, bioinformatics, and pattern recognition. Springer, Berlin, pp 273–309

Tversky A (1977) Features of similarity. Psychol Rev 84(4):327–352. https://doi.org/10.1037/0033-295X.84.4.327

Tversky B, Hemenway K (1984) Objects, parts, and categories. J Exp Psychol: Gen 113(2):169–197. https://doi.org/10.1037/0096-3445.113.2.169

Verbeemen T, Vanpaemel W, Pattyn S, Storms G, Verguts T (2007) Beyond exemplars and prototypes as memory representations of natural concepts: a clustering approach. J Mem Lang 56(4):537–554. https://doi.org/10.1016/j.jml.2006.09.006

Vigliocco G, Vinson DP, Lewis W, Garrett MF (2004) Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. Cogn Psychol 48(4):422–488. https://doi.org/10.1016/j.cogpsych.2003.09.001

Vivas J, Vivas L, Comesaña A, Coni AG, Vorano A (2017) Spanish semantic feature production norms for 400 concrete concepts. Behav Res Methods 49(3):1095–1106. https://doi.org/10.3758/s13428-016-0777-2

Wilderjans TF, Ceulemans E, Van Mechelen I, Depril D (2011) ADPROCLUS: a graphical user interface for fitting additive profile clustering models to object by variable data matrices. Behav Res Methods 43(1):56–65. https://doi.org/10.3758/s13428-010-0033-0

Wu LL, Barsalou LW (2009) Perceptual simulation in conceptual combination: evidence from property generation. Acta Physiol (Oxf) 132:173–189. https://doi.org/10.1016/j.actpsy.2009.02.002

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