

Received July 20, 2020, accepted July 31, 2020, date of publication August 11, 2020, date of current version August 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3015789

An Automated Process for the Repository-Based Analysis of Ontology Structural Metrics

JOSÉ ANTONIO BERNABÉ-DÍAZ¹, MANUEL FRANCO-NICOLÁS²,
JUANA MARÍA VIVO-MOLINA², MANUEL QUESADA-MARTÍNEZ³, ASTRID DUQUE-RAMOS⁴,
AND JESUALDO TOMÁS FERNÁNDEZ-BREIS¹, (Member, IEEE)

¹Departamento de Informática y Sistemas, Universidad de Murcia, CEIR Campus Mare Nostrum, IMIB-Arrixaca, 30100 Murcia, Spain

²Departamento de Estadística e Investigación Operativa, Universidad de Murcia, CEIR Campus Mare Nostrum, IMIB-Arrixaca, 30100 Murcia, Spain

³Center of Operations Research, Miguel Hernández University of Elche, 03202 Elche, Spain

⁴Grupo de Investigación INGECO, Facultad de Ingenierías, Universidad Autónoma Latinoamericana (UNLA), Medellín 050010, Colombia

Corresponding author: Jesualdo Tomás Fernández-Breis (jfernand@um.es)

This work was supported in part by the Spanish Ministry of Economy, Industry and Competitiveness, and in part by the European Regional Development Fund (ERDF) Programme under Grant TIN2017-85949-C2-1-R.

ABSTRACT Quantitative metrics are generally applied by scientists to measure and assess the properties of data and knowledge resources. In ontology engineering, a number of metrics have been developed to analyse different features of ontologies in the last few years. However, this community has not generated any standard framework for studying the properties of ontologies or generated sufficient knowledge about the usefulness and validity as the measurement instrument of these metrics for evaluating and comparing ontologies. Recently, 19 ontology structural metrics were studied using the OBO Foundry and AgroPortal ontology repositories. This study was based on how each metric partitioned the two datasets into five groups by applying the k -means algorithm. The results suggested that the use of five clusters for every metric might be suboptimal. In this paper, we propose an automated process for the study of ontology structural metrics by including the selection of an optimal number of clusters for each metric. This optimal number is automatically obtained by using statistical properties of the generated clusters. Moreover, the cosine similarity is used for estimating the similarity of two repositories from the perspective of the behaviour of the same set of metrics. The results on the two datasets allow for a more realistic perspective on the behaviour of the metrics. In this paper, we show and discuss the difference observed in the comparative behaviour of the metrics on the two repositories when using the optimal number with respect to a predetermined number of clusters for every metric. The proposed method is not specific for ontology metrics and therefore, can be applied to other types of metrics.

INDEX TERMS Knowledge-based systems, knowledge engineering, clustering methods, biomedical informatics, biomedical ontologies, quality metrics.

I. INTRODUCTION

Semantic web technologies and ontologies are widely used in the development of knowledge-based systems [1]–[7]. Their success lies in the combination of four main features present in almost all ontologies: standard identifiers for classes and relations that represent the phenomena within a domain; a vocabulary for a domain; metadata that describe the intended meaning of the classes and relations; and machine-readable axioms and definitions that enable computational access to some aspects of the meaning of classes and relations [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara¹.

The development of hundreds of biological and biomedical ontologies has resulted in the need for repository-based initiatives to facilitate the finding and sharing of biomedical knowledge. Examples of such repositories are AgroPortal [9], OBO Foundry [10], BioPortal [11], OntoBee [12], Ontology Lookup Service (OLS) [13] or AberOWL [14].

Consequently, in these repositories, researchers can find different ontologies covering similar biological subdomains, and they need support in making informed decisions about which to use, the differences between the ontologies, and so on. The ontology community has recognised the need for reference methods to measure the quality of ontologies [15], which could contribute to guiding users in their decisions,

but there is no community agreement thus far. In the last few decades, different types of approaches for evaluating the quality of ontologies have been proposed:

- **Qualitative approaches:** These approaches are based on the application of qualitative criteria. A diagnostic task based on ontology descriptions, using three categories of criteria (structural, functional and usability profiling) was proposed in [16]. In [17], an approach that applies four qualitative criteria (philosophical rigour, ontological commitment, content correctness, and fitness for a purpose) was proposed.
- **Quantitative approaches:** These approaches use quantitative metrics for measuring ontology properties [18]–[25]. They vary in the number and type of metrics used, which means that each approach may be interested in the quantitative assessment of the concrete facets of ontology quality. This has been the most active area in the last few years.
- **Principles and guidelines:** These approaches propose to follow a series of recommendations to ensure the quality of the ontology. Examples are the OBO Foundry principles [10] or the MIRO guideline [26].

In this work, we were interested in quantitative metrics, as they enable objective and reproducible evaluation processes. The relevance of quantitative metrics has been sufficiently demonstrated in the literature in the last few decades for different fields, including software engineering [27], [28] and recruitment [29]. Unfortunately, the metrics displayed by ontology repositories are limited, and each repository includes different metrics.

In our case, we believe that increasing the knowledge about the usage of metrics can help to increase the knowledge about the engineering of ontologies and to define effective methods for linking the values of the metrics to the structural properties of ontologies. Our belief is reinforced by recent works which have started to apply quantitative metrics to ontology design patterns [30]. This is why in the last few years, we have been studying and learning how quantitative metrics can be applied for analysing and understanding the structure and engineering of ontologies. Early works studied how to combine quantitative metrics to analyse the features of ontologies [19], [31] and how to use them to study the evolution of ontologies from the metrics perspective [32]. The main target of these works was the analysis of ontologies.

In particular, every ontology was assigned a rating between 1 and 5 for every metric. In [31], [32], such a rating was based on the application of the best practices. In [32], as all the versions of an ontology could be considered a repository, the rating was based on the distribution of the data in the repository. One feature of this approach is the use of five categories for every metric. After these works, an important question remained unanswered: which metrics provide a better classification of ontologies so that they can be more useful for discriminating between ontologies? In [33], a method based on public repository data for answering such questions was

presented, by applying 19 ontology structural metrics to the OBO Foundry repository and to AgroPortal. The behaviour of some metrics was the same in both the repositories, but it changed for the others. To the best of our knowledge, this past study was the first one that focused on the reliability of ontology metrics. In this previous study, the behaviour was measured by the stability and the goodness of the clustering, two statistical properties of the five clusters in which the ontology repository was partitioned for a given metric. However, the study had the limitation of using five clusters for every metric, which might not be optimal.

In the present study, we addressed the abovementioned limitation by defining a method for selecting the optimal number of clusters for each metric and analysed the behaviour of each metric by using the OBO Foundry and AgroPortal datasets used in [33]. The optimal number of clusters for a metric might vary for different datasets or repositories. Many validity indexes were considered for analysing the clustering result with a prefixed k number, and different rules were formulated to choose a k value into a plausible range of cluster numbers. However, each rule was based on a specific index, and no standard consensus on the best validity index and rule was found in the literature. Our main aim was to search the optimal number of clusters by combining two validity indexes: stability and goodness of the clustering. Hence, the first contribution of this work was an automated process for analysing ontology structural metrics for classifying the ontologies included in a repository. Note that a certain metric was studied in [33] using five clusters in each repository, but currently, the optimal number of clusters for the same metric might differ between repositories. Metrics with similar behaviour in different repositories could be the most appropriate for standardisation. This led to the second contribution, as the characterisation of the metrics described the content of the repository. Finally, we proposed a metric which measured the similarity between repositories on the basis of the behaviour of the metrics. The proposed methods were applied to ontology metrics and repositories in this study, which provided new insights in the field of ontology evaluation methods. However, these methods are generic, so they can be applied to study the behaviour of other quantitative metrics.

II. METHODS

A. ONTOLOGY METRICS AND THEIR PROCESSING

In this work, we studied a set of 19 ontology structural metrics, $M = \{M_1, \dots, M_{19}\}$ described in Table 7. These metrics accounted for different features of an ontology (e.g. cohesion, multiple inheritance, ratio of properties per class, and ratio of annotations per class). All the metrics included in this work were quantitative. Therefore, for every metric, we could define a function $f(x)$ with an ontology as its domain and the raw values of the metric as its range. Note that the metrics might have different ranges of raw values. Our method was designed for application to a repository, in this work, a set of

ontologies $\theta = \{\theta_1, \dots, \theta_n\}$. Hence, once the raw values for a given metric were computed for all of the ontologies in the repository, we applied a clustering algorithm to partition the ontologies into k non-empty and non-overlapping categories. The clustering could be seen as a function $n(f(x))$, which generated an ordered factor of k categories. Furthermore, $n(f(x))$ partitioned the range of $f(x)$ in k non-prefixed continuous intervals that contained all of the observed samples in the experimental data. This process was repeated for each metric. The results of the process for partitioning the range of $f(x)$ in k intervals were therefore metric and dataset dependent, as the clusters were dynamically generated for a given dataset. Each ontology would then be associated with a particular cluster for each metric. The assessment of all of the ontologies included in the same cluster for the feature measured by a certain metric would be the same. This could also serve to perform a qualitative evaluation of the ontologies, but this was out of the scope of the present study. Our focus was on the metrics and on the repositories.

Fig. 7 describes the process applied to a repository of ontologies for a certain value of k and for each metric. First, the raw values for all the ontologies and for all the metrics were obtained. Second, the clustering algorithm performed, for each metric, the scaling of the raw values of the ontologies into the k categories.

B. CLUSTERING METHODS

There are different clustering algorithms for the partitioning of objects into a predefined number k of non-empty and non-overlapping clusters [34], [35]. Here, we applied the k -means clustering method, which is one of the most well-known and widely used clustering methods [36]. Such an algorithm aims to find the partitioning by maximising both the compactness of the objects within a cluster and the separability between the clusters.

Nevertheless, our algorithm for selecting the optimal k could be applied to any partitioning method, such as the Partitioning Around Medoids (PAM) and Clustering LARge Applications (CLARA) algorithms [37]. The PAM method minimises the sum of the dissimilarities of the objects in the dataset to their nearest medoid (the most centrally located object of each cluster), which produces spherical clusters. The CLARA method takes a sample of objects from the dataset in which PAM is applied to find medoids and to compute the best clustering as a result.

Furthermore, accelerated techniques for these clustering methods have been recently developed to reduce the computational cost [38]. These improvements are mainly related to the BUILD initialisation and SWAP refinement algorithms, which in the PAM method consists of the following: (1) the complexity of the BUILD phase for choosing an initial clustering is reduced from $O(kn^2)$ to $O(kn)$, and (2) the workload of finding the closest new medoids of the SWAP refinement is reduced from $O(k(n-k)^2)$ to $O((n-k)^2)$. As CLARA uses PAM, CLARA is also improved by these accelerated

techniques. Both of them will also be applied to show the comparison of different clustering methods for selecting the optimal k for each metric.

C. ANALYSIS OF THE OPTIMAL NUMBER OF CATEGORIES

The methods described in the previous section can be applied for different values of k , which means that different groupings of the ontologies can be obtained, as the optimal number of clusters is usually unknown. The lack of a gold standard has led to the need for a validation process of an unsupervised classification method. We propose to validate the results of each clustering by measuring two robust features of the k categories generated: stability and goodness of the clustering. In a nutshell, the optimal k is the one whose clustering is the most stable and accurate. Nevertheless, the two criteria involved in determining such an optimum may not return the same k .

Next, the cluster validation criteria applied are described. We also propose an algorithm which automates the process for selecting the best k when the ones reported by both indexes are different; this algorithm can be applied to any unsupervised partitioning method.

1) CLUSTERING STABILITY

Stability refers to whether the structure of a meaningful valid cluster is significantly altered by small variations in the data. Therefore, it seems reasonable to assess the effect of potential changes on each category of the clustering (C_1, \dots, C_k) by a bootstrap resampling procedure [39]. For each metric M_i in M , we measured the stability $S_{M_i,k}(C_j)$ of each category C_j by using the Jaccard coefficient of similarity between sets [40]. In a repository, this index provides the proportion of concordant ontologies between C_j and the most similar cluster in a bootstrapped clustering of a metric M_i . Thereby, the stability measure $S_{M_i,k}(C_j)$ is derived as the mean of the values obtained from the b bootstrap samples. Such scores can be interpreted as degrees of statistical stability [41], thus classifying each category into the types displayed in Table 1.

TABLE 1. Stability classification.

Range	Category
[0, 0.60)	Unstable
[0.60, 0.75]	Doubtful
(0.75, 0.85]	Stable
(0.85, 1]	Highly stable

In addition, to deal with the comparison of the metrics M_i for $i = 1, \dots, m$ on a repository in terms of the clustering stability, the mean of the category stability scores for each metric $S_k(M_i)$ can be used as a global stability criterion of M_i , assuming the same relative importance of the categories.

Algorithm 1 Selecting the Optimal k

```

1: procedure OPTIMAL_K(metrics, stability, goodness)
2:    $k_{op} \leftarrow 1$ 
3:    $k_s \leftarrow \max(\text{stability})$ 
4:    $k_g \leftarrow \max(\text{goodness})$ 
5:   for each  $M_i \in \text{Metrics}$  do
6:     if  $k_s == k_g$  then
7:        $k_{op} \leftarrow k_g$ 
8:     else if  $\{S_{k_s}(M_i) > 0.75 \wedge S_{k_g}(M_i) > 0.75\}$ 
9:        $\vee \{S_{k_s}(M_i) \leq 0.75 \wedge S_{k_g}(M_i) \leq 0.75\}$  then
10:       $k_{op} \leftarrow k_g$ 
11:    else
12:      if  $\{S_{k_s}(M_i) > 0.75 \wedge sil_{k_s}(M_i) > 0.5 \wedge S_{k_g}(M_i) \leq 0.75\}$ 
13:      then
14:         $k_{op} \leftarrow k_s$ 
15:      else
16:        if  $\{S_{k_s}(M_i) > 0.75 \wedge sil_{k_s}(M_i) \leq 0.5$ 
17:           $\wedge S_{k_g}(M_i) \leq 0.75 \wedge sil_{k_g}(M_i) > 0.5\}$  then
18:           $k_{op} \leftarrow k_g$ 
19:        else
20:           $k_{op} \leftarrow k_s$ 

```

2) CLUSTERING GOODNESS

The goodness of the clustering is assessed by two internal cluster validation measurements, namely, cohesion and separation of the clusters. The first feature reflects how closely related the ontologies in a category are, whereas the other quantifies how well-separated a category is from the rest of the categories. Advantageously, both of them can be gathered into several validity indexes, such as the Silhouette width (sil) [42], Calinski-Harabasz (ch) [43], Dunn ($dunn$) [44], and Davies-Boudin (db) [45] measurements. Among them, we propose to use the Silhouette width (sil) because it can provide the classification quality for both ontologies and metrics, which gives it a competitive edge over the others. For each metric, the sil coefficient of a particular ontology represents the degree of confidence in the clustering and is given by $sil_{M_i,k}(\theta_l) = (b_l - a_l) / \max(a_l, b_l)$, for $l = 1, \dots, n$, where a_l is the average distance between the ontology θ_l and all the others belonging to the same category, and b_l is the average distance between the ontology θ_l and the ones of the closest neighbouring category. In particular, $sil_{M_i,k}(\theta_l)$ measures how well each ontology θ_l has been clustered, varying its value between -1 and 1 , which can be interpreted as in [42]. A value close to 1 (-1 , respectively) means that the ontology tends to be ‘well-classified’ (‘misclassified’, respectively). A value close to zero signifies that the ontology can be assigned to the nearest neighbouring category as well. For comparison purposes, the overall goodness of the clustering for a metric M_i is evaluated by the global Silhouette coefficient, which is defined by the average of the sil scores, $sil_k(M_i) = \sum_{l=1}^n sil_{M_i,k}(\theta_l) / n$, for $i = 1, \dots, m$. As suggested by Kaufman and Rousseeuw [37], the global Silhouette width score can be interpreted as the effectiveness

TABLE 2. Structure classification.

Range	Clustering Structure
$[-1, 0.25)$	Not substantial
$[0.25, 0.50]$	Weak, probably artificial
$(0.50, 0.70]$	Reasonable
$(0.70, 1]$	Strong

of the clustering structure found, in terms of the metrics (see Table 2).

3) ALGORITHM FOR SELECTING THE OPTIMAL k

The methods presented in the previous sections permit the application of the partitioning of a set of ontologies based on the values for a certain metric M_i and for a range of k -values. For each k within this range of k -values, the clustering is carried out, allowing one to quantify its validity indexes, such as the stability and the goodness of the clustering. We propose to use the stability and the goodness scores to select the value of k that provides the best clustering of the ontologies for a metric M_i . The clustering obtained for the optimal k should be used to analyse the set of ontologies from the perspective of the metric M_i .

The algorithm for determining the optimal k is presented in Algorithm 1 from $k_s = \arg_k \max S_k(M_i)$ and $k_g = \arg_k \max sil_k(M_i)$ for each metric M_i . Its purpose is to search the optimal k number of clusters by combining stability and goodness, which provides the most robust clustering with respect to such validity measures. This goal is pursued by the following reasoning:

- 1) If the highest stability and the highest goodness are obtained for the same value of k , $k_s = k_g$, then return this value as the optimal one.
- 2) If the highest stability and the highest goodness are obtained for different values, $k_s \neq k_g$, then the optimal k value should be calculated as follows:
 - a) If both k_s and k_g provide at least stable classifications or both provide non-stable classifications, the optimal k should be the value with the largest Silhouette width, i.e. k_g .
 - b) If k_s provides at least stable and reasonable classifications and k_g does not provide stable classifications, we should select k_s .
 - c) If k_s provides at least stable, but less than reasonable, classifications and k_g does not provide stable classifications, then if k_g provides at least a reasonable silhouette, then we should select k_g . Otherwise, k_s should be selected.

D. REPOSITORY PROFILE

The calculation of the optimal k value for each metric in a repository enables one to define the profile of a repository as a vector with m elements, one per metric. Each element in the profile is the optimal k value for the metric M_i . The repository profile may be considered as characteristic information on the repository from the perspective of a set of metrics, which can be used to compare repositories.

Given an ordered set M of m metrics that can be applied over a repository, this can have different q -dimensional profiles, where q is the number of metrics used for constructing the profile. In the case of a q -dimensional profile with a size smaller than m , there will be different possibilities for selecting the q metrics. The elements in the profile preserve the order defined in M . In this work, we used a q -dimensional profile corresponding to q metrics given in Table 7 with $q \leq 19$.

Moreover, the similarity between two repositories based on their q -dimensional profiles, $q \leq m$, can be interpreted as the homogeneity degree of the repositories with respect to these q metrics, which could be obtained by applying the cosine similarity given as follows:

$$\cos(\vec{r}_1, \vec{r}_2) = \frac{\vec{r}_1 \cdot \vec{r}_2}{\|\vec{r}_1\| \|\vec{r}_2\|} = \frac{\sum_{i=1}^q r_{1,i} r_{2,i}}{\sqrt{\sum_{i=1}^q r_{1,i}^2 \sum_{i=1}^q r_{2,i}^2}}$$

where \vec{r}_1 and \vec{r}_2 are the corresponding q -dimensional profiles of the repositories.

Note that this similarity would relate the behaviour of the metrics on the repositories, not the content. Moreover, the similarity between profiles would require both the vectors to be homogeneous; that is, they represent the same ordered set of metrics.

III. RESULTS

A. EXPERIMENTAL SETUP

We applied our process to two repositories of ontologies, AgroPortal and OBO Foundry. More concretely, we analyzed

78 AgroPortal ontologies ($\theta^{agr} = \{\theta_1^{agr}, \dots, \theta_{78}^{agr}\}$) and 119 OBO Foundry ones ($\theta^{obo} = \{\theta_1^{obo}, \dots, \theta_{119}^{obo}\}$), which were studied in [33]. Our results are based on the raw scores of ontology structural metrics retrieved from the OQuaRE REST service,¹ which can be reproduced through either the *evaluome* web service [46] or the R package version *evaluomeR* [47]. The *evaluome* tool allows the evaluation of the reliability of metrics with a user-friendly interface. It analyses the stability and the goodness of the clusters reported for each metric by using both the global validation criteria described in Sections II-C1 and II-C2. The package *evaluomeR* [47] is available in Bioconductor. In addition, the optimal number of clusters for each metric is assessed by the decision-making algorithm proposed in Section II-C3, which is implemented in *evaluomeR* release v1.3.4 hosted GitHub.² *evaluomeR* depends on the following packages: *cluster* [48], *corrplot* [49], *Rdpack* [50], *plotrix* [51], *fpc* [52] *SummarizedExperiment* [53], and *MultiAssayExperiment* [54]. It requires R version 3.6 or higher to run [55]. Other dependencies such as the Bioconductor or CRAN R packages are automatically downloaded via the Bioconductor install manager. Note that the accelerated techniques described in Section II-B for the clustering methods are implemented in the R *cluster* package 2.0.9 version onwards, and consequently, the experimental results by *evaluomeR* are obtained through these accelerated optimisations of the PAM and CLARA algorithms. The resulting tables (CSVs), figures, and *evaluomeR* scripts are available in the directory *usecases/kcomparison*³ in our GitHub repository.

B. CLUSTERING METHODS COMPARISON

A comparison of the three clustering methods described in Section II-B is shown in Fig. 1 for AgroPortal and Fig. 2 for OBO Foundry. In general, the optimal k value selected for each metric on both the repositories from k -means is lower or equal to the values reported for PAM and CLARA, except for NACOnto, Ponto, and WMCOnto2 in the AgroPortal repository (see Fig. 8). In most of the cases, k -means provides similar or better classifications according to Tables 1 and 2 from their global stability and global goodness scores. In the rest of the paper, we will apply the k -means algorithm to analyse and discuss the automated process proposed for the study of ontology structural metrics, because it is one of the most well-known methods. The users of our process can choose the most suitable clustering method for their study.

C. DESCRIPTIVE STATISTICS OF THE METRICS

We used violin plots [56] to describe the behaviour of the distributional shape of the raw scores on each repository (see Fig. 3 and Fig. 4). In these graphical representations,

¹<http://sele.inf.um.es/ontology-metrics/>

²<https://github.com/neobernad/evaluomeR>

³<https://github.com/neobernad/evaluomeR/tree/master/usecases/kcomparison>

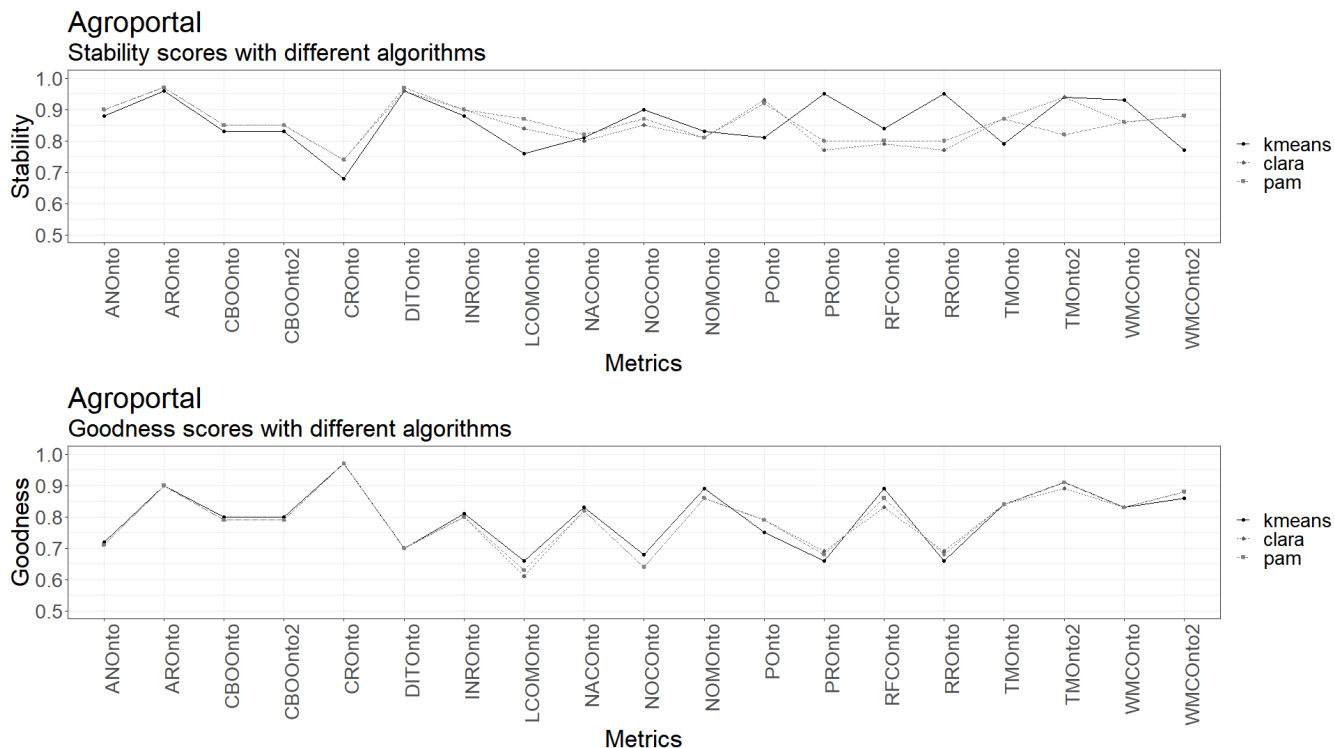


FIGURE 1. Visual comparison of different clustering methods in the θ^{agr} repository for the stability and goodness scores from the optimal k value selected for each metric.

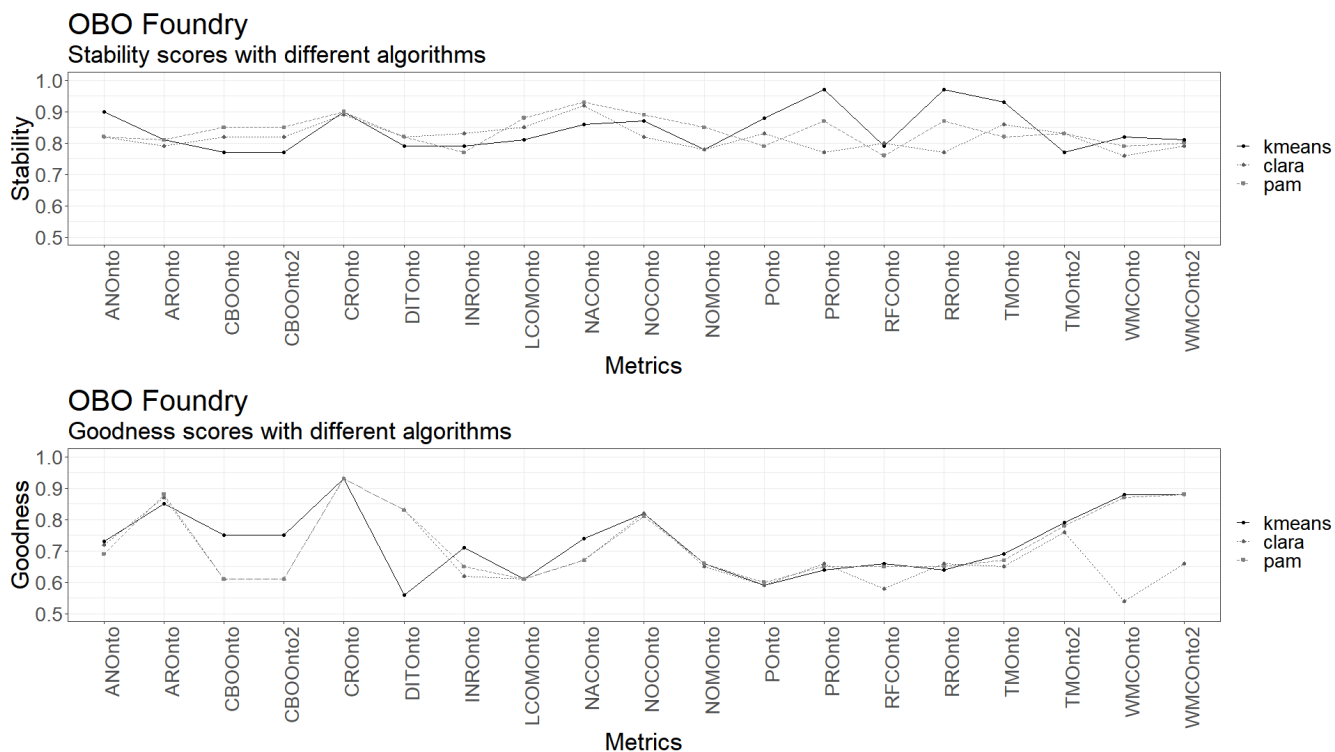


FIGURE 2. Visual comparison of different clustering methods in the θ^{obo} repository for the stability and goodness scores from the optimal k value selected for each metric.

the kernel density trace that overlaid the boxplot might assist in revealing the existence of groups of ontologies, which were scaled into interval [0, 1] for a visual comparison. At a quick

glance, these figures convey different behaviours of most of the ontology structural metrics in each repository, as well as between both the repositories.

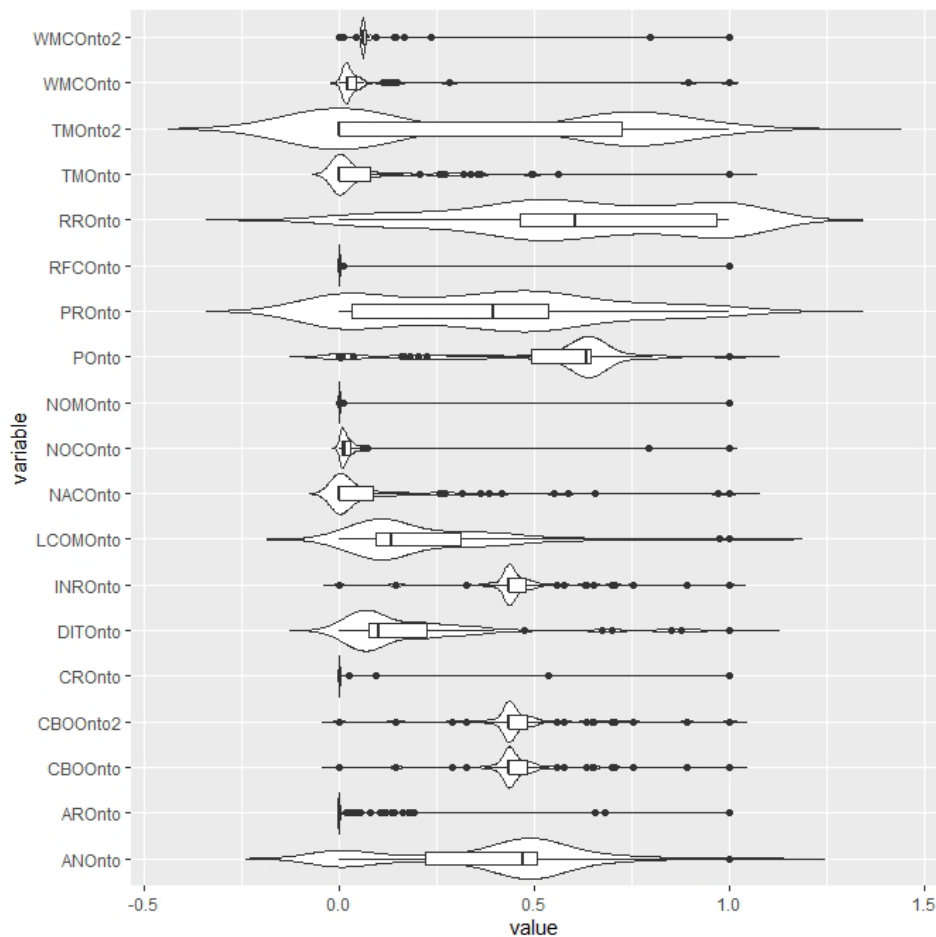


FIGURE 3. Violin plots of the scaled raw data of metrics for the AgroPortal repository.

D. SELECTION OF THE OPTIMAL k VALUES

In order to determine the optimal number of categories, we performed our evaluation on the generated clusterings with k varying between 3 and 15. This range was selected to retain all the ontology variability in both the repositories, θ^{agr} and θ^{obo} , thereby avoiding elementary binary classifications. The number of bootstrap replications was set at 500 in order to reach relative reliable and accurate outcomes (as reported in [33]).

Table 3 and Table 8 display the global stability scores for each metric on the AgroPortal and OBO Foundry repositories, respectively. The highest stability $S_k(M_i)$ was reached when $k = 3$ for 12 out of the 19 metrics on AgroPortal and for 11 on OBO Foundry. Furthermore, the highest stability score for each metric fell within the range of 0.71 to 0.96 on AgroPortal (of 0.79 to 0.97 on OBO Foundry), so there were no ‘Unstable’ clusterings of the metrics, and there was one only ‘Doubtful’ clustering, CROnto, on the AgroPortal repository. Moreover, there are 8 (9) ‘Stable’ clusterings for the metrics CBOOnto, CBOOnto2, LCOMOnto, NOMOnto, POnto, RFCOnto, TMOnto, and WMCOnto2 (AROnto, CBOOnto, CBOOnto2,

DITOnto, INROnto, NOMOnto, RFCOnto, WMCOnto, and WMCOnto2), and 10 (10) ‘Highly stable’ clusterings of the rest of the metrics. In detail, 5.26% (0%) metrics were classified as ‘Doubtful’, 42.11% (47.37%) were ‘Stable’, and 52.63% (52.63%) were ‘Highly stable’.

In addition, Table 4 and Table 9 show the global goodness scores for each metric on the AgroPortal and OBO Foundry repositories, respectively. The highest goodness $sil_k(M_i)$ was achieved when $k = 3$ for 12 out of 19 metrics on both the repositories. Moreover, the highest goodness score per metric ranged from 0.66 to 0.91 on AgroPortal (from 0.60 to 0.93 on OBO Foundry), so no metrics generated either unstructured clusterings or weak ones. Indeed, 73.68% (57.89%) of the metrics provided categories with ‘Strong structure’, and 26.32% (42.11%) reproduced categories with ‘Reasonable structure’ in AgroPortal (OBO Foundry).

Table 5 reports the optimal k values, and Fig. 5 shows the scores returned from the performed evaluation analysis on each of the 19 ontology structural metrics in the AgroPortal and OBO Foundry repositories, respectively. Note that there were 11 (57.89%) metrics in AgroPortal and 13 (68.42%) in OBO Foundry whose corresponding k s for the global

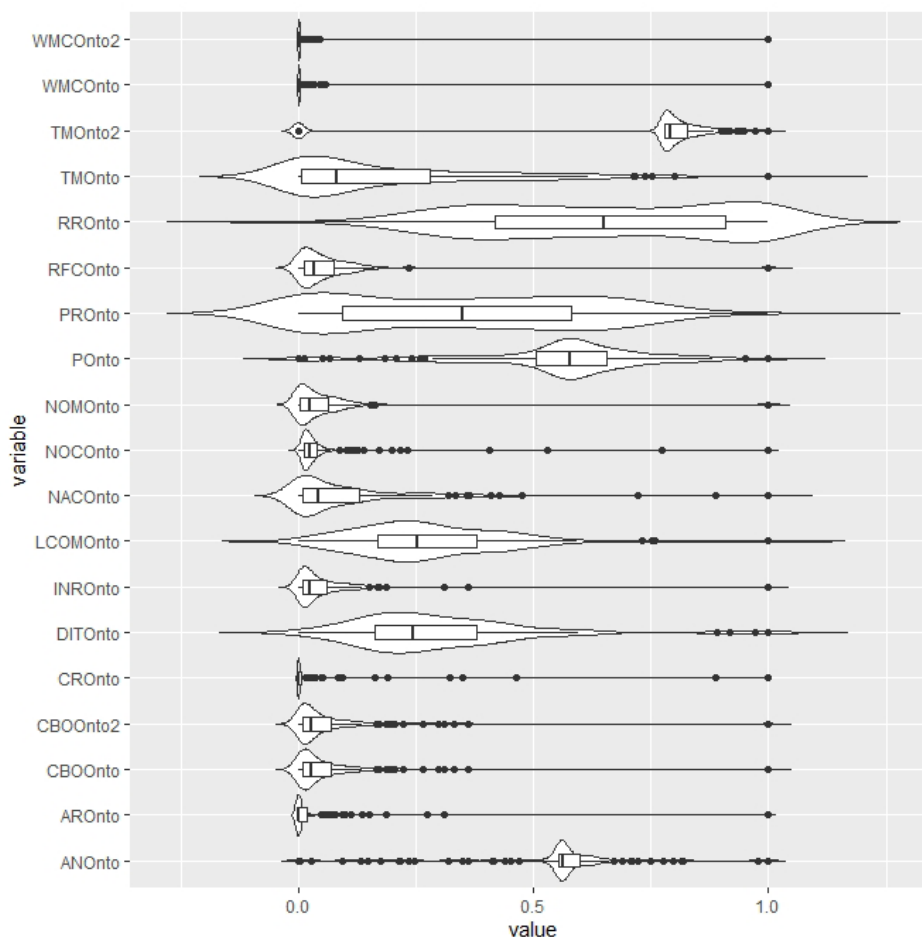


FIGURE 4. Violin plots of the scaled raw data of metrics for the OBO Foundry repository.

stability and goodness criteria were different. The optimal k displayed in the column ‘Optimal k ’ was the result of the application of the algorithm proposed in Section II-C3, thus providing an automatic decision-making process to carry out the statistical optimisation of the repository-based analysis of the 19 ontology structural metrics. This automatic process revealed that $k = 3$ was optimal for most of the metrics in both the repositories, that is for 15 metrics from θ^{agr} and 16 metrics from θ^{obo} out of the 19 in total (78.94% and 84.21%, respectively).

E. ANALYSIS OF THE OPTIMAL k VALUES ACROSS REPOSITORIES

The optimal k values for stability and goodness across the metrics is shown in Table 5 for the θ^{agr} and θ^{obo} repositories. The graphical representation of the stability and the goodness scores for the optimal k are shown in Fig. 5. Furthermore, the qualitative interpretation of their stability and goodness is shown in Table 6. Moreover, 18 metrics from θ^{agr} and 19 from θ^{obo} provided at least stable clusterings (>0.75). All the metrics exhibited at least a reasonable clustering

structure in both the θ^{agr} and θ^{obo} repositories by using their optimal k .

In addition, we could compare the behaviour of the metrics in both the θ^{agr} and the θ^{obo} repositories when we used their optimal k value. The main findings were as follows:

- The following metrics had the same optimal k ($k = 3$) in both the repositories: ANOnto, CBOOnto, CBOOnto2, CROnto, DITOnto, INROnto, NOCOnto, NOMOnto, PROnto, RFCOnto, RROnto, and WMCOnto.
- The clustering structures presented an alike distribution for the metrics with the same optimal k , except for CBOOnto, CBOOnto2, CROnto, INROnto, and RFCOnto. This difference in the Silhouette scores for CBOOnto and RFCOnto in both the repositories is clearly observed in Fig. 9 and 10 for CBOOnto, and Fig. 11 and 12 for the RFCOnto metric.
- The following metrics had similar global stability scores in both the repositories: ANOnto, CBOOnto, CBOOnto2, CROnto, DITOnto, INROnto, NOCOnto, PROnto, RFCOnto, RROnto, and WMCOnto.

TABLE 3. Stability scores for the θ^{agr} repository in the k range [3, 15]. The highest score for each metric is highlighted in bold lettering, and NA is reported when there are insufficient different values to build k clusters.

	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	k=12	k=13	k=14	k=15
ANOnto	0.88	0.79	0.72	0.76	0.75	0.76	0.69	0.62	0.72	0.67	0.75	0.74	0.74
AROnto	0.96	0.79	0.81	0.73	0.76	0.75	NA	NA	NA	NA	NA	NA	NA
CBOOnto	0.83	0.76	0.78	0.81	0.76	0.77	0.75	0.81	0.77	0.78	0.79	0.79	0.75
CBOOnto2	0.83	0.76	0.78	0.81	0.76	0.77	0.75	0.81	0.77	0.78	0.79	0.79	0.75
CROnto	0.68	0.62	0.6	0.7	0.7	0.71	0.62	0.71	0.67	0.7	0.68	0.67	0.68
DITOnto	0.96	0.64	0.7	0.67	0.68	0.62	0.6	0.64	0.61	0.63	0.66	0.73	NA
INROnto	0.88	0.76	0.79	0.83	0.71	0.74	0.8	0.77	0.76	0.78	0.78	0.76	0.77
LCOMOnto	0.76	0.67	0.82	0.74	0.78	0.71	0.75	0.72	0.66	0.65	0.62	0.6	0.72
NACOnto	0.86	0.71	0.79	0.83	0.8	0.78	0.79	0.81	0.77	0.74	0.71	0.72	0.7
NOCOnto	0.9	0.91	0.83	0.85	0.86	0.75	0.69	0.73	0.72	0.7	0.72	0.71	0.71
NOMOnto	0.83	0.78	0.72	0.7	0.79	0.78	0.8	0.81	0.8	0.78	0.77	0.73	0.72
POnto	0.8	0.81	0.81	0.79	0.72	0.72	0.76	0.77	0.76	0.75	0.75	0.75	0.74
PROnto	0.95	0.93	0.74	0.75	0.75	0.69	0.65	0.69	0.73	0.73	0.7	0.67	0.67
RFCOnto	0.84	0.78	0.72	0.78	0.72	0.73	0.73	0.77	0.71	0.74	0.74	0.74	0.69
RROnto	0.95	0.93	0.74	0.75	0.75	0.69	0.65	0.69	0.73	0.73	0.7	0.67	0.67
TMOnto	0.75	0.74	0.78	0.79	0.74	0.79	0.74	0.75	0.71	0.71	0.68	0.68	0.68
TMOnto2	0.94	0.96	0.8	0.78	0.83	0.81	NA	NA	NA	NA	NA	NA	NA
WMCOnto	0.93	0.91	0.88	0.82	0.79	0.81	0.79	0.69	0.69	0.69	0.72	0.69	0.7
WMCOnto2	0.63	0.77	0.83	0.84	0.76	0.78	0.74	0.77	0.74	0.7	0.69	0.69	0.67

TABLE 4. Goodness scores for the θ^{agr} repository in the k range [3, 15]. The highest score for each metric is highlighted in bold lettering.

	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	k=12	k=13	k=14	k=15
ANOnto	0.72	0.64	0.68	0.68	0.62	0.64	0.62	0.61	0.66	0.66	0.66	0.6	0.66
AROnto	0.9	0.88	0.87	0.87	0.88	0.89	0.87	0.87	0.84	0.85	0.84	0.81	0.8
CBOOnto	0.8	0.62	0.65	0.69	0.67	0.68	0.51	0.66	0.67	0.67	0.64	0.59	0.6
CBOOnto2	0.8	0.62	0.65	0.69	0.67	0.68	0.51	0.66	0.67	0.67	0.64	0.59	0.6
CROnto	0.97	0.95	0.95	0.83	0.8	0.85	0.79	0.8	0.73	0.8	0.77	0.72	0.72
DITOnto	0.7	0.69	0.69	0.63	0.64	0.63	0.66	0.63	0.61	0.68	0.75	0.83	0.77
INROnto	0.81	0.79	0.66	0.68	0.44	0.66	0.67	0.64	0.65	0.67	0.63	0.57	0.55
LCOMOnto	0.66	0.62	0.59	0.53	0.61	0.54	0.57	0.58	0.58	0.59	0.58	0.57	0.6
NACOnto	0.79	0.78	0.78	0.8	0.81	0.82	0.81	0.83	0.82	0.8	0.79	0.79	0.78
NOCOnto	0.68	0.66	0.61	0.54	0.59	0.53	0.51	0.56	0.55	0.55	0.56	0.55	0.55
NOMOnto	0.89	0.84	0.77	0.56	0.67	0.61	0.67	0.65	0.65	0.66	0.66	0.66	0.62
POnto	0.72	0.75	0.72	0.64	0.57	0.59	0.63	0.66	0.62	0.62	0.61	0.57	0.56
PROnto	0.66	0.68	0.68	0.64	0.64	0.63	0.62	0.65	0.69	0.69	0.67	0.65	0.63
RFCOnto	0.89	0.84	0.76	0.61	0.55	0.6	0.59	0.61	0.58	0.59	0.6	0.59	0.59
RROnto	0.66	0.68	0.68	0.64	0.64	0.63	0.62	0.65	0.69	0.69	0.67	0.65	0.63
TMOnto	0.78	0.75	0.79	0.81	0.77	0.84	0.81	0.82	0.79	0.81	0.81	0.81	0.8
TMOnto2	0.91	0.89	0.9	0.87	0.89	0.89	0.89	0.86	0.9	0.86	0.87	0.88	0.85
WMCOnto	0.83	0.67	0.57	0.59	0.63	0.59	0.61	0.6	0.62	0.6	0.55	0.54	0.53
WMCOnto2	0.75	0.86	0.78	0.76	0.76	0.71	0.72	0.71	0.78	0.76	0.77	0.75	0.74

F. REPOSITORY PROFILES AND SIMILARITY

The repository profiles could be affected by the clustering algorithm used; Fig. 8 displays a visual comparison of the 19-dimensional profiles according to the three methods for each repository, although we focused on the k -means algorithm to compare repositories, as mentioned before.

The comparison between the AgroPortal and the OBO Foundry repositories was carried out through their 19-dimensional profiles, which were defined from the optimal k values in Table 5 by using the 19 metrics included in the study and ordered as shown in Table 7:

$$\vec{r}_{19}^{agr} = (3, 3, 3, 3, 3, 3, 3, 3, 10, 3, 3, 4, 3, 3, 3, 8, 3, 3, 4)$$

$$\vec{r}_{19}^{obo} = (3, 4, 3, 3, 3, 3, 3, 8, 3, 3, 3, 3, 3, 3, 3, 10, 3, 3)$$

TABLE 5. Optimal k values for the metrics in the θ^{agr} and θ^{obo} repositories. The column *Stability* presents the most stable k value for each metric, whereas *Goodness* exhibits the k value with the greatest silhouette for such a metric. The *Optimal k* column displays the optimal k obtained by the proposed selection algorithm.

	Agroportal			OBO Foundry		
	Stability	Goodness	Optimal k	Stability	Goodness	Optimal k
ANOnto	3	3	3	3	3	3
AROnto	3	3	3	4	9	4
CBOnto	3	3	3	7	3	3
CBOnto2	3	3	3	7	3	3
CROnto	8	3	3	3	3	3
DITOnto	3	14	3	3	15	3
INROnto	3	3	3	9	3	3
LCOMOnto	5	3	3	4	8	8
NACOnto	3	10	10	4	3	3
NOCOnto	4	3	3	3	3	3
NOMOnto	3	3	3	6	3	3
POnto	5	4	4	3	4	3
PROnto	3	11	3	3	12	3
RFCOnto	3	3	3	4	3	3
RROnto	3	11	3	3	12	3
TMOnto	6	8	8	3	3	3
TMOnto2	4	3	3	3	10	10
WMCOnto	3	3	3	3	3	3
WMCOnto2	6	4	4	3	3	3

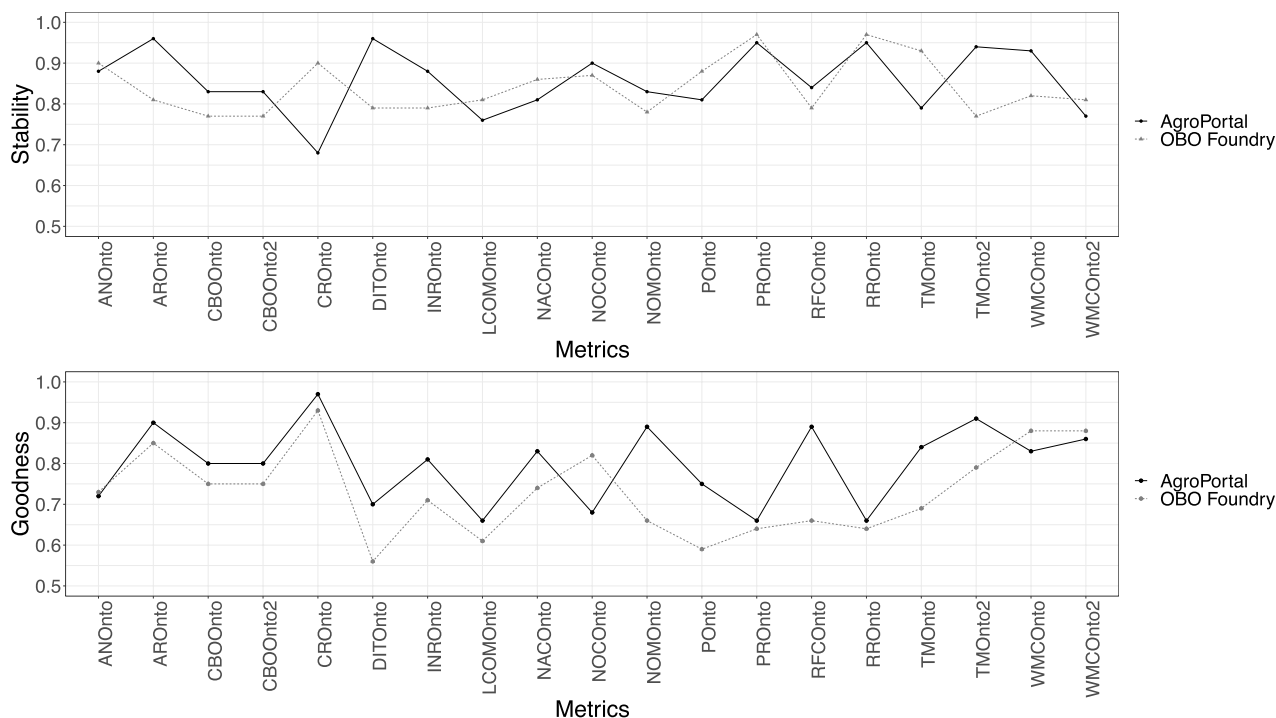


FIGURE 5. Visual comparison between θ^{agr} and θ^{obo} for the stability and goodness scores from the optimal k value selected for each metric.

The cosine similarity of these two 19-dimensional profiles was 0.77. Hence, the same optimal k was achieved in both the repositories for 12 metrics, i.e. the same value in \vec{r}_{19}^{agr} and \vec{r}_{19}^{obo} . Thus, the similarity of the corresponding 12-dimensional profiles was 1, which could be interpreted as

the two repositories being fully homogeneous with respect to the subset of 12 metrics.

Furthermore, note that in [33], perfect correlations were found for certain metrics. For instance, the CBOnto and CBOnto2 metrics presented a positive perfect correlation, and

TABLE 6. Classification of the stability and goodness scores across metrics for their optimal k value in the θ^{agr} and θ^{obo} repositories.

	Agropotal		OBO Foundry	
	Stability	Goodness	Stability	Goodness
ANOnto	Highly stable	Strong	Highly stable	Strong
AROnto	Highly stable	Strong	Stable	Strong
CBOOnto	Stable	Strong	Stable	Strong
CBOOnto2	Stable	Strong	Stable	Strong
CROnto	Doubtful	Strong	Highly stable	Strong
DITOnto	Highly stable	Reasonable	Stable	Reasonable
INROnto	Highly stable	Strong	Stable	Strong
LCOMOnto	Stable	Reasonable	Stable	Reasonable
NACOnto	Stable	Strong	Highly stable	Strong
NOCOnto	Highly stable	Reasonable	Highly stable	Strong
NOMOnto	Stable	Strong	Stable	Reasonable
POnto	Stable	Strong	Highly stable	Reasonable
PROnto	Highly stable	Reasonable	Highly stable	Reasonable
RFCOnto	Stable	Strong	Stable	Reasonable
RROnto	Highly stable	Reasonable	Highly stable	Reasonable
TMOnto	Stable	Strong	Highly stable	Reasonable
TMOnto2	Highly stable	Strong	Stable	Strong
WMCOnto	Highly stable	Strong	Stable	Strong
WMCOnto2	Stable	Strong	Stable	Strong

CBOnto and INROnto had an almost positive perfect correlation, which were detected in both repositories. Nevertheless, dependent metrics do not affect the characteristic information on a repository, as such characteristic information is provided by the independent metrics. For each metric, dependent or independent, from a subset of q metrics, the optimal k value is independently calculated by evaluomeR to determine the q -dimensional profile, and the component corresponding to a dependent metric is a function of the remaining metrics. It may provide redundant information without affecting the comparison, although it could lightly modify the similarity because of the increase in the dimension.

For example, let us take the following set of 6 independent metrics: ANOnto, AROnto, CBOOnto, CROnto, NOCOnto, and POnto. From Table 5, the 6-dimensional profiles for θ^{agr} and θ^{obo} based on these independent metrics were as follows:

$$\vec{r}_6^{agr} = (3, 3, 3, 3, 3, 4)$$

$$\vec{r}_6^{obo} = (3, 4, 3, 3, 3, 3)$$

Hence, the cosine similarity between both the repositories was 0.983. Next, we added two metrics highly correlated with CBOnto to consider an 8-dimensional profile. As stated before, these metrics were CBOnto2 and INROnto. Therefore, the 8-dimensional profiles with respect to the ordered set of metrics (ANOnto, AROnto, CBOOnto, CBOnto2, CROnto, INROnto, NOCOnto, POnto) were as follows:

$$\vec{r}_8^{agr} = (3, 3, 3, 3, 3, 3, 3, 4)$$

$$\vec{r}_8^{obo} = (3, 4, 3, 3, 3, 3, 3, 3)$$

which did not affect the components of the 6-dimensional profiles, although its similarity slightly increased to 0.987. In any case, these results reported a high homogeneity degree between both the repositories based on these 6 or 8 metrics.

IV. DISCUSSION

A. PROPOSED METHOD

The method presented in this work is a natural evolution of [33]. There, the goal was to develop a mechanism for analysing the behaviour of a set of metrics for a prefixed value of k . This method was extended to deal with certain ranges of values of k and to automatically suggest the optimal number of clusters for each metric. Note that this algorithm is just one possibility and that the users of the method are entitled to design their own algorithm in case they prefer to prioritise and interpret the stability and the goodness scores in a different manner. Therefore, the proposed method optimises the analysis, thereby constituting an excellent benchmark for the developers of ontology metrics, the developers of ontology evaluation methods, and the managers of ontology repositories. The method aims at helping to make informed decisions on the use and interpretation of metrics in a repository. Note that the method used in this study can also be applied to other types of quantitative metrics, such as the ones presented in [57]–[59]. The only requirement for the application of the method is to have a dataset for which at least one metric has been measured and that the range of values for the metric permits one to build a certain number of k clusters. Consequently, we believe that the method is of interest to data scientists in general.

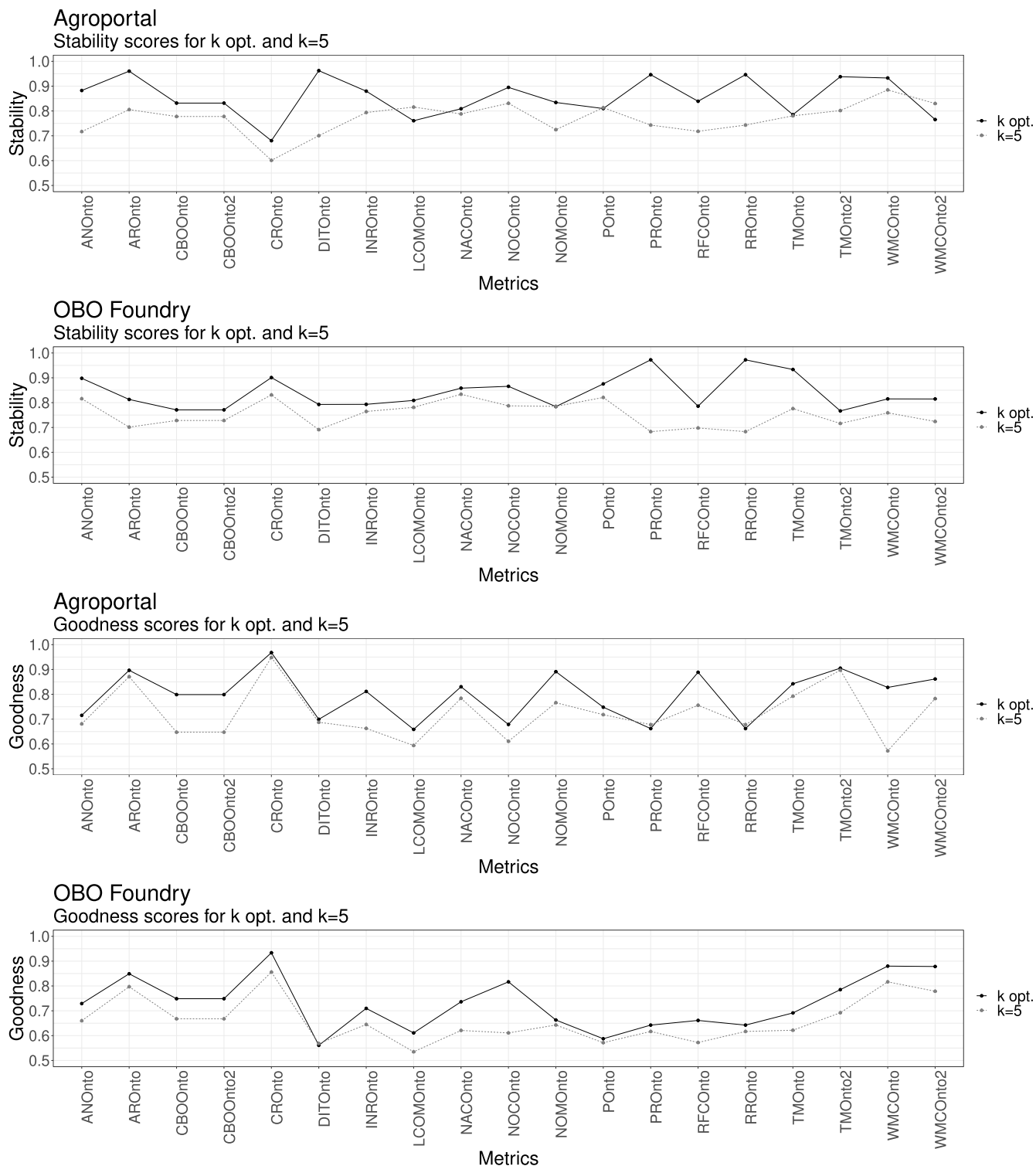


FIGURE 6. Visual comparison of the stability and goodness scores between $k = 5$ and the optimal k for each metric in each repository.

B. METRICS

In [33], 19 ontology structural metrics were evaluated assuming a partitioning in five clusters of the values of each ontology for each metric in each repository. The results obtained in this work showed two main differences (see Table 5):

- (1) The optimal k value was different from five for all the metrics included in the study.
- (2) The optimal k value was the same in both the repositories for 12 out of 19 metrics. This is an indicator of the existence of differences between the two repositories of ontologies studied from the perspective of

TABLE 7. Definition of the 19 metrics evaluated: Column 1 shows the acronym of the metric and its reference; Column 2 describes the ontology facet measured by the metric; and Column 3 describes how the metric is calculated.

Metric	Facet measured	Description
ANOnto [23]	Annotations	Mean number of annotations properties per classes
AROnto [23]	Attribute richness	Number of restrictions of the ontology per classes
CBOnto [61], [62]	Coupling	Number of direct ancestors of classes divided by the number of classes minus subclasses of thing
CBOnto2 [62]	Coupling	Mean number of direct ancestors per classes
CROnto [23]	Individuals	Mean number of individuals per classes
DITOnto [61], [62]	Depth of the hierarchy	Length of the longest path from thing to a leaf classes
INROnto [23]	Descendants	Mean number of subclasses per classes
NACOnto [63]	Ancestors of leaf classes	Mean number of superclasses per leaf classes
NOCOnto [61], [62]	Descendants	Number of the direct subclasses divided by the number of classes minus the number of leaf classes
NOMOnto [63]	Properties	Mean number of object and data property usages per class
POnto [62]	Ancestors	Mean number of direct ancestors per class
LCOMOnto [64]	Cohesion	Mean length of all paths from leaf classes to thing
PROnto [23]	Property richness	Number of subclass of relationships divided by the number of subclass of relationships and properties
RFCOnto [61], [62]	Properties usage	Number of usages of object and data properties and superclasses divided by the number of classes
RROnto [23]	Properties usage	Number of usages of object and data properties and super classes divided by the number of classes
TMOnto [62]	Multiple inheritance	Mean number of classes with more than one ancestor
TMOnto2 [62]	Multiple inheritance	Mean number of direct ancestors of classes with more than 1 direct ancestor
WMCOnto [61], [62]	Complexity	Mean length of the paths from thing to a leaf classes
WMCOnto2 [62]	Complexity	Mean number of paths from thing to a leaf classes

TABLE 8. Stability scores for ρ^{obo} repository in k range [3, 15]. The highest score for each metric is highlighted in bold lettering.

	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	k=12	k=13	k=14	k=15
ANOnto	0.9	0.73	0.82	0.72	0.64	0.7	0.72	0.72	0.73	0.73	0.71	0.71	0.75
AROnto	0.73	0.81	0.7	0.74	0.72	0.67	0.74	0.76	0.73	0.7	0.74	0.68	0.66
CBOOnto	0.77	0.78	0.73	0.78	0.82	0.81	0.78	0.75	0.74	0.7	0.7	0.65	0.67
CBOOnto2	0.77	0.78	0.73	0.78	0.82	0.81	0.78	0.75	0.74	0.7	0.7	0.65	0.67
CROnto	0.9	0.85	0.83	0.78	0.86	0.81	0.83	0.8	0.77	0.75	0.76	0.75	0.75
DITOnto	0.79	0.66	0.69	0.6	0.62	0.66	0.56	0.59	0.63	0.63	0.6	0.61	0.65
INROnto	0.79	0.72	0.76	0.77	0.74	0.76	0.79	0.75	0.7	0.71	0.71	0.67	0.65
LCOMOnto	0.81	0.86	0.78	0.83	0.81	0.81	0.75	0.72	0.71	0.68	0.7	0.69	0.68
NACOnto	0.86	0.93	0.83	0.79	0.76	0.8	0.77	0.77	0.76	0.74	0.72	0.68	0.65
NOCOnto	0.87	0.79	0.79	0.76	0.79	0.77	0.77	0.8	0.73	0.7	0.74	0.74	0.73
NOMOnto	0.78	0.75	0.79	0.8	0.76	0.76	0.76	0.72	0.7	0.69	0.69	0.69	0.7
POnto	0.88	0.71	0.82	0.78	0.68	0.68	0.71	0.7	0.73	0.72	0.72	0.71	0.72
PROnto	0.97	0.72	0.68	0.62	0.61	0.72	0.69	0.7	0.71	0.7	0.67	0.64	0.65
RFCOnto	0.79	0.81	0.7	0.71	0.74	0.76	0.72	0.69	0.71	0.72	0.71	0.71	0.68
RROnto	0.97	0.72	0.68	0.62	0.61	0.72	0.69	0.7	0.71	0.7	0.67	0.64	0.65
TMOnto	0.93	0.76	0.78	0.77	0.64	0.6	0.53	0.59	0.55	0.64	0.71	0.71	0.66
TMOnto2	0.94	0.91	0.72	0.74	0.74	0.79	0.79	0.77	0.74	0.75	0.73	0.73	0.72
WMCOnto	0.82	0.81	0.76	0.79	0.74	0.78	0.77	0.73	0.7	0.67	0.69	0.69	0.74
WMCOnto2	0.81	0.81	0.72	0.78	0.77	0.78	0.77	0.74	0.7	0.73	0.73	0.74	0.72

the ontology features measured by these metrics. Our interpretation for one metric having the same optimal number of clusters in both repositories is that it has the same behaviour in both the repositories; hence, it classifies the data in the

same number of groups. However, this does not necessarily mean that the distribution of values for the metric in both the repositories is the same. For example, the optimal number of clusters for ANOnto or CBOnto was 3, but the shapes of their

TABLE 9. Goodness scores for θ^{obo} repository in k range [3, 15]. The highest score for each metric is highlighted in bold lettering.

	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	k=11	k=12	k=13	k=14	k=15
ANOnto	0.73	0.71	0.66	0.67	0.67	0.68	0.59	0.6	0.62	0.56	0.53	0.54	0.58
AROnto	0.82	0.85	0.8	0.82	0.85	0.85	0.87	0.87	0.86	0.85	0.87	0.85	0.83
CBOnto	0.75	0.67	0.67	0.65	0.62	0.6	0.55	0.53	0.54	0.48	0.53	0.49	0.51
CBOnto2	0.75	0.67	0.67	0.65	0.62	0.6	0.55	0.53	0.54	0.48	0.53	0.49	0.51
CROnto	0.93	0.9	0.86	0.84	0.85	0.76	0.76	0.69	0.68	0.7	0.71	0.71	0.7
DITOnto	0.56	0.51	0.57	0.56	0.56	0.57	0.54	0.59	0.63	0.63	0.62	0.63	0.68
INROnto	0.71	0.71	0.64	0.59	0.61	0.6	0.57	0.56	0.49	0.5	0.56	0.52	0.54
LCOMOnto	0.54	0.59	0.53	0.56	0.58	0.61	0.58	0.56	0.56	0.56	0.55	0.55	0.56
NACOnto	0.74	0.67	0.62	0.63	0.61	0.6	0.61	0.6	0.61	0.61	0.6	0.61	0.59
NOCOnto	0.82	0.82	0.61	0.63	0.59	0.53	0.55	0.55	0.54	0.54	0.55	0.55	0.54
NOMOnto	0.66	0.6	0.64	0.64	0.6	0.59	0.62	0.61	0.59	0.59	0.6	0.62	0.63
POnto	0.59	0.6	0.57	0.59	0.54	0.55	0.57	0.52	0.54	0.55	0.54	0.54	0.56
PROnto	0.64	0.6	0.62	0.58	0.55	0.62	0.62	0.64	0.63	0.65	0.6	0.57	0.56
RFCOnto	0.66	0.62	0.57	0.55	0.59	0.57	0.53	0.52	0.55	0.56	0.56	0.58	0.57
RROnto	0.64	0.6	0.62	0.58	0.55	0.62	0.62	0.64	0.63	0.65	0.6	0.57	0.56
TMOnto	0.69	0.63	0.62	0.6	0.64	0.6	0.62	0.6	0.61	0.63	0.62	0.63	0.61
TMOnto2	0.74	0.72	0.69	0.72	0.72	0.72	0.78	0.79	0.75	0.77	0.75	0.76	0.75
WMCOnto	0.88	0.83	0.82	0.66	0.62	0.53	0.52	0.48	0.48	0.49	0.51	0.52	0.55
WMCOnto2	0.88	0.85	0.78	0.73	0.72	0.64	0.6	0.61	0.62	0.61	0.57	0.55	0.55

violin plots in AgroPortal and OBO Foundry were different (see Fig. 3 and Fig. 4)

In [33], six metrics (NACOnto, NOCOnto, POnto, TMOnto, TMOnto2, and WMCOnto) provided stable or highly stable clusters. Now, with the use of the optimal value of k for each metric (see Fig. 5 and Table 6), only CROnto (0.68) in AgroPortal was not stable or highly stable. Consequently, 18 out of 19 metrics were stable or highly stable in both the repositories when the optimal k values suggested by our method were used. The use of the optimal k value revealed that all the metrics had at least a reasonable structure, and such clusters were equal or more strongly structured than the ones built using a prefixed $k = 5$ in both the repositories. Moreover, 14 metrics had a strong structure in AgroPortal and 11 in OBO Foundry. In most cases, the optimal k value was the one providing the largest goodness value. When the optimal k value was provided by the largest stability value, the metrics were highly stable and had at least a reasonable structure in AgroPortal (DITOnto, PROnto, and RROnto) and were at least stable and had a reasonable structure in the OBO Foundry (AROnto, DITOnto, POnto, PROnto, and RROnto). It can be seen that the optimal k value was provided by the stability for DITOnto, PROnto, and RROnto in both the repositories. These three metrics were related to the existence of properties and relations, including taxonomic ones in the ontology. The following 10 metrics had a strong structure in both the repositories: ANOnto, AROnto, CBOnto, CBOnto2, CROnto, INROnto, NACOnto, TMOnto2, WMCOnto, and WMCOnto2. These metrics were also stable or highly stable in both the repositories; therefore,

we can conclude that they exhibit the same behaviour in both the repositories. Most of these metrics are related to important features of ontologies, such as the taxonomy (CBOnto, CBOnto2, INROnto, NACOnto, TMOnto2, WMCOnto, and WMCOnto2), annotations (ANOnto), attributes (AROnto), and individuals (CROnto). These results show the benefits of using the optimal k value and reveal the difference between the repositories.

Fig. 6 shows the changes in stability and goodness when the optimal k value was used instead of $k = 5$. Both aspects improved for most of the metrics when the optimal k value was used, which was the expected result. There were cases such as LCOMOnto in AgroPortal in which the stability score was higher for $k = 5$, although the optimal k was 3 because of the results obtained for the goodness, as the latter was given priority when the same degree of stability was achieved for different values of k .

We could also compare the results of the content of the clusters obtained between both the optimal k and $k = 5$ for the 19 metrics. Figs. 13 and 14 provide a visual comparison of the content of the clusters generated in each repository. We could infer different types of situations. Two examples are described next for the AgroPortal case (see Fig. 13). For WMCOnto, the red and black clusters were common for both values of k , and the blue one (optimal k) included the remaining three clusters from $k = 5$. For ANOnto, we observed how the content of the clusters when $k = 5$ shown with red data points was distributed in two clusters of the optimal k , and the same happened for the cluster of green points. Consequently, there was no common pattern relating

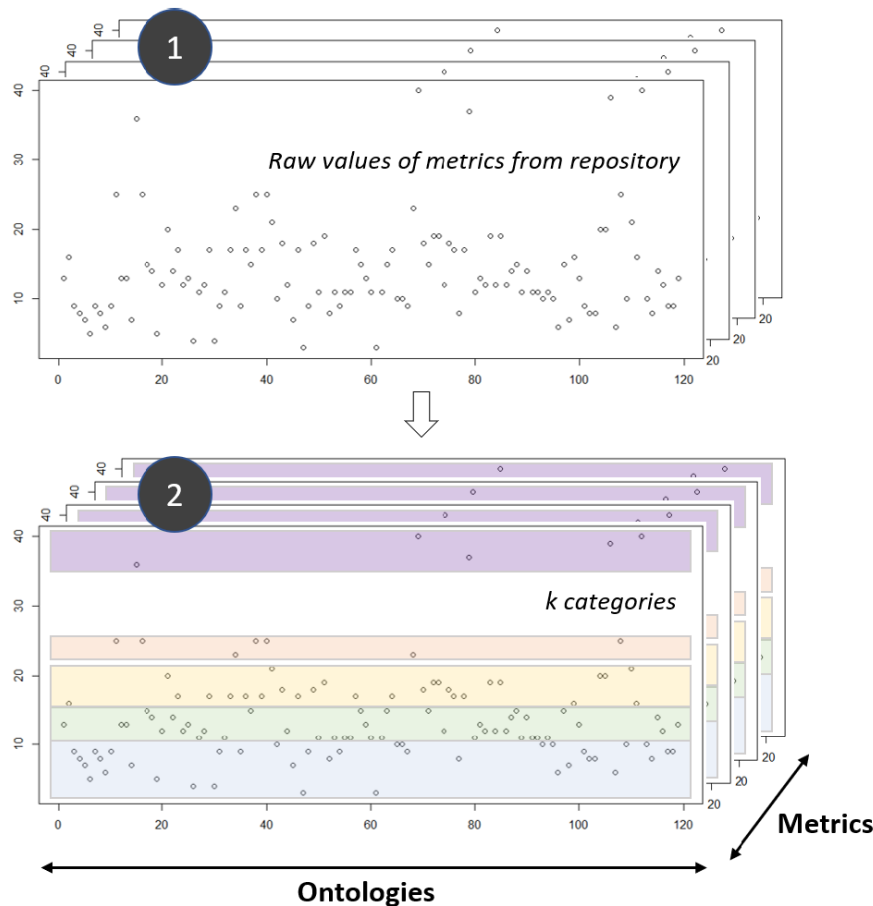


FIGURE 7. Applying our method to an ontology repository: (1) computation of the values of the metrics for the ontologies; (2) application of a clustering algorithm.

the content of the clusters obtained with the optimal k and the ones obtained with $k = 5$; hence, each clustering generated its own structure.

C. REPOSITORIES

Our method uses the data existing in a given repository to study the behaviour of a set of metrics. It is not surprising that the optimal number of clusters for a metric might differ between repositories, as their content is expected to be different. The general interpretation of the results obtained in both the repositories is as follows: some metrics could be appropriate for certain repositories and not for the others. Note that the repositories are in continuous evolution, so our results would hold for the particular versions of the ontologies used in the study. Changes in the repositories might imply changes in our results.

In our study, both the repositories were analysed with k varying between 3 and 15 (see Tables 3-9). The results showed that $k > 8$ could not be applied for three metrics (AROnto, DITOnto, and TMOnto2) in AgroPortal, whereas such an application was possible in the OBO Foundry repository. This was attributed to the number of different values

for these metrics in AgroPortal. This could also be due to the lower number of ontologies in the AgroPortal repository. Generally speaking, not being able to obtain the stability or goodness for certain values of k should not be interpreted negatively for the metric, the repository, or the method. This made it explicit that such a number of categories does not make sense for the particular metric in the particular repository. We believe that this is another example of how the method captures the signal differentiating the repositories. The differences for these three metrics could be intuitively affirmed by inspecting the violin plots for the metrics. A clear example would be TMOnto2, as the violin plot in AgroPortal suggests three large groups, whereas the data were more disperse in OBO Foundry. If we compare the optimal k values for these three metrics in both the repositories, we can see that the optimal k for AROnto is 3/9 (AgroPortal/OBO Foundry), for DITOnto 3/3, and for TMOnto2 3/10. The optimal for OBO Foundry is larger, showing that the content of the repositories is different from the perspective of the metric.

In Section IV-B, 12 metrics were identified as having the same optimal k value in both the repositories; therefore, the behaviour was different for 7 metrics:

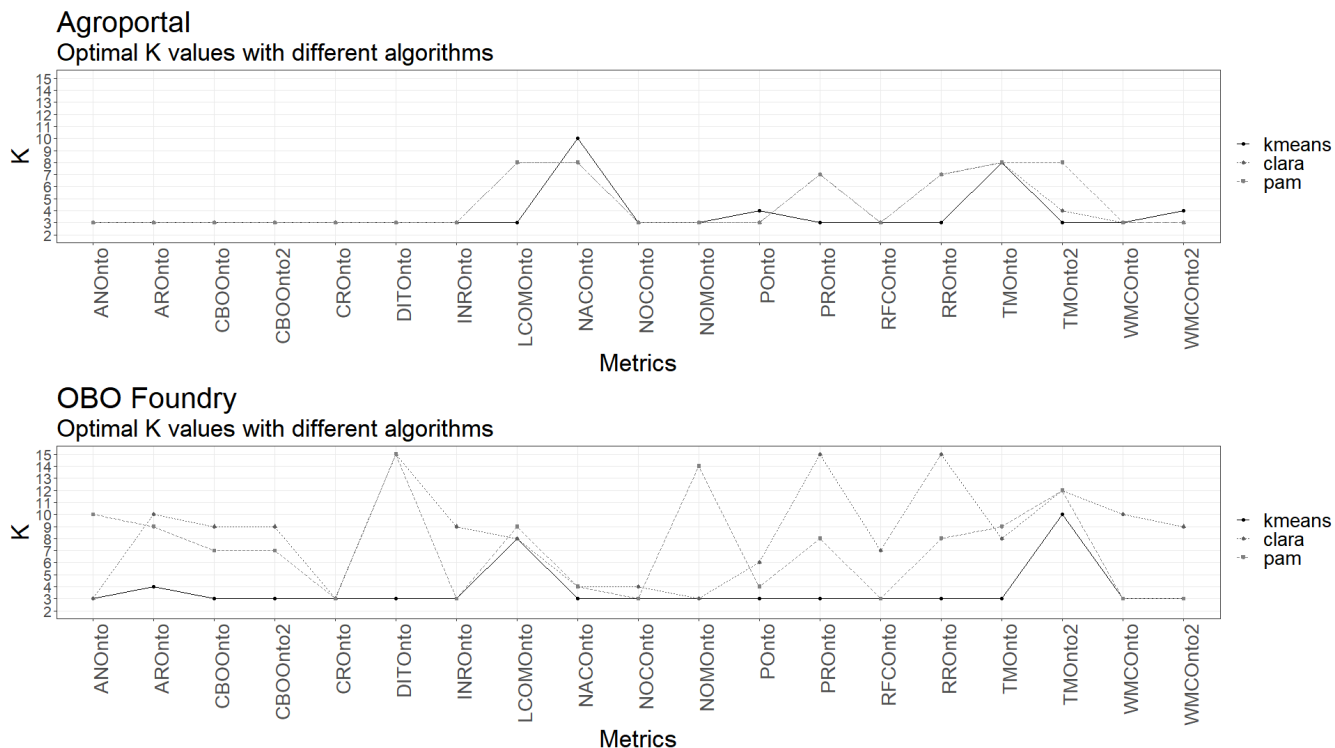


FIGURE 8. Visual comparison of the optimal k values for θ^{agr} and θ^{obo} by using the k -means, PAM and CLARA algorithms.

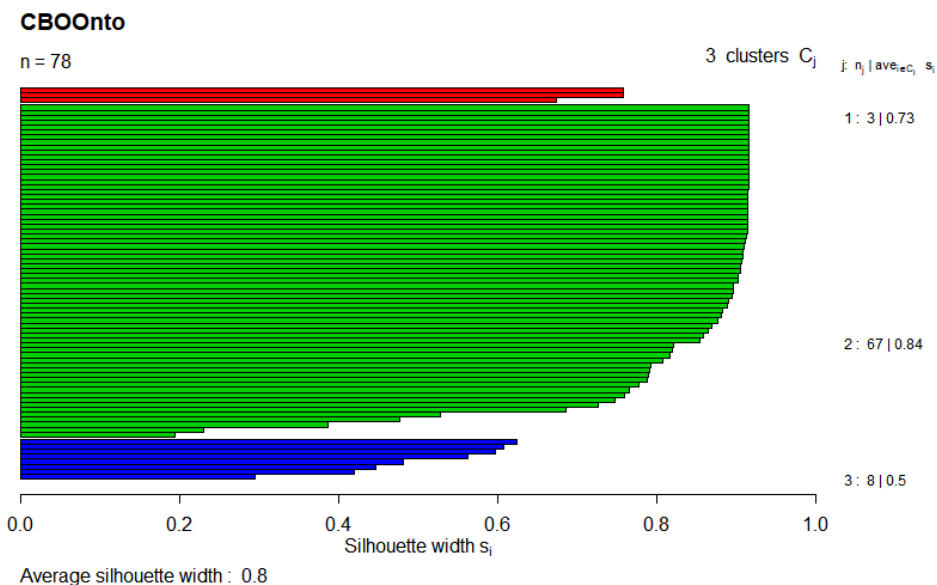


FIGURE 9. Silhouette plot of the clustering distribution in θ^{agr} for CBOnto metric in its optimal k .

- The optimal k value was higher in AgroPortal for the metrics NACOnto, POnto, TMOnto, and WMCOnto2.
- The optimal k value was higher in OBO Foundry for the metrics AROnto, LCOMOnto, and TMOnto2.

This difference implied that the optimal partitioning of the repositories using these metrics required a different number of groups. Moreover, 5 out of these 7 metrics measured facets related to the taxonomy. In three of these cases, the

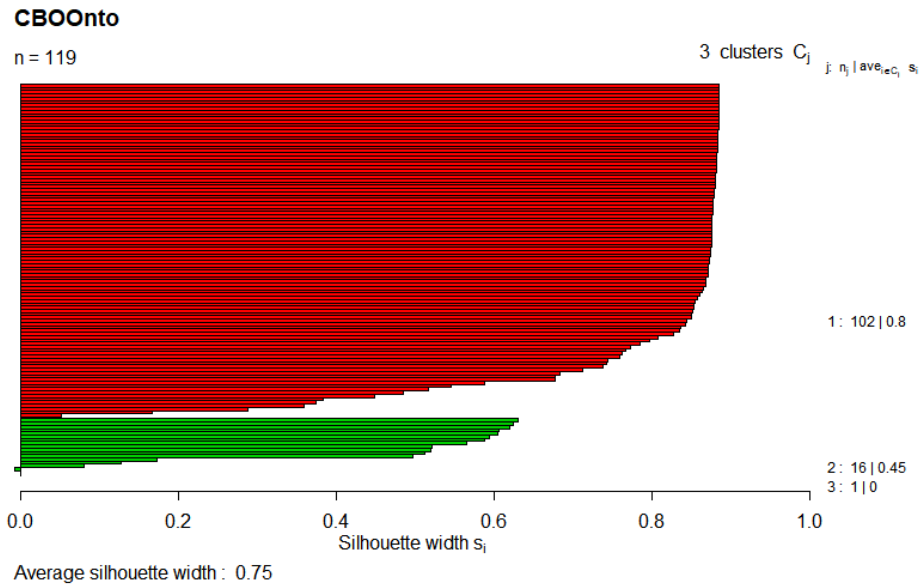


FIGURE 10. Silhouette plot of the clustering distribution in θ^{obo} for CBOnto metric in its optimal k .

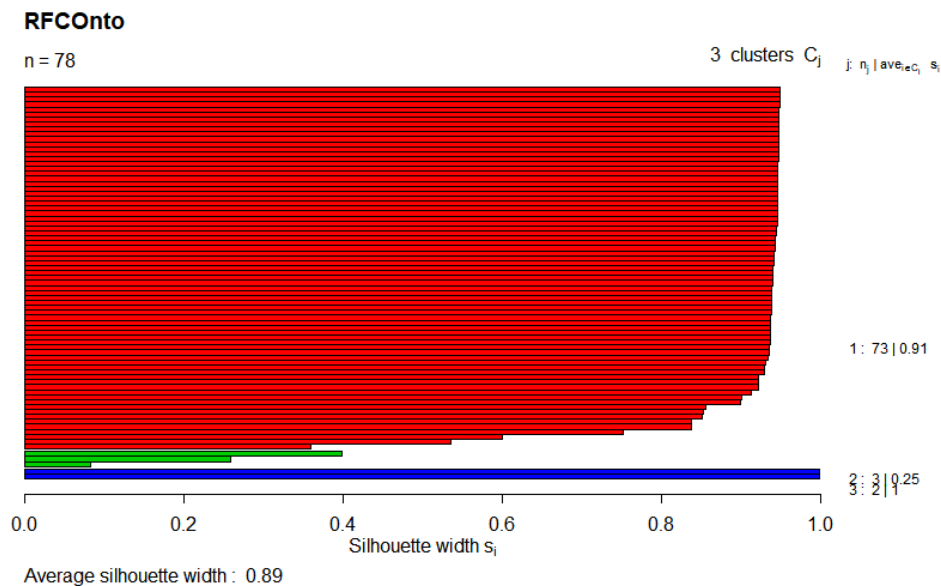


FIGURE 11. Silhouette plot of the clustering distribution in θ^{agr} for RFCOnto metric in its optimal k .

difference was in one cluster, but the situation was different for NACOnto (10-3), LCOMOnto (3-8), TMOnto (8-3), and TMOnto2 (3-10). For example, both TMOnto and TMOnto2 were related to ancestors, whereas TMOnto referred to classes with multiple ancestors, TMOnto2 referred to the direct ancestors of the classes targeted by TMOnto. Thus, we can conclude that this result confirms the adaptive nature of our method to the actual distribution of the data points and that each metric captures a specific feature.

The optimal number of clusters for a given metric may differ between repositories, which makes it a data-dependent feature. We could draw an analogy between this repository-based feature and parameters such as K and Λ in BLAST [60], which are calculated on the basis of the content of the sequence database and which are used for computing the scores and E-Value. In our case, the optimal number of clusters was used to build the repository profile, the q -dimensional profiles. We think that the profiles may be an

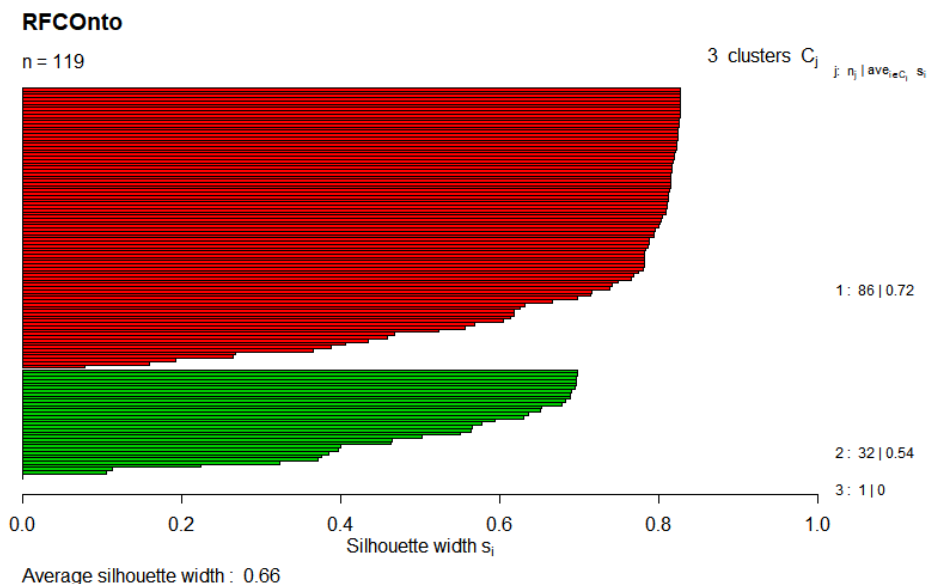


FIGURE 12. Silhouette plot of the clustering distribution in θ^{obo} for RFCOnto metric in its optimal k .

interesting and flexible tool to estimate whether two sets of metrics have a similar behaviour across repositories, because they can also be applied to subsets of the set of metrics. For example, the similarity of the behaviour of a set of metrics related to a particular feature (e.g. classes or annotations) could be studied using our q -dimensional profile approach.

D. REPOSITORY PROFILES

The scope of our automated process was also enlarged by collecting the optimal k number of clusters for each metric on a repository into a vector referred as to the repository profile. This profile might be considered to be feature information of the repository with respect to an ordered set M of m metrics, allowing users to compare the classification performance of the set M on different ontology portals in terms of the similarity between their m -dimensional profiles by using the cosine function. As a result, the profiles could be ranked according to their cosine similarity measures. Furthermore, it is worthy to remark the flexibility of the repository profile which may be used for any q -dimensional ordered subset of metrics chosen by the user, $q \leq m$.

E. LIMITATIONS AND FURTHER WORK

Our method for obtaining the optimal number of clusters for the metric does not analyse the existence of exceptional cases, such as clusters with only one member. For example, such outlier cases can be identified by exploring the violin plots shown in Section III-C, but they must be manually identified. For instance, the NOCOnto metric depicts outlier cases in Fig. 3 and 4, and TMOnto2 can describe an exceptional case.

In the violin plot of θ^{obo} , the scaled raw data of TMOnto2 have two well-separated density functions on two small clusters, whereas in θ^{agr} , the distribution is visually different (these two density functions are overlapped).

The existence of such cases should be informed to the algorithm for selecting the optimal number of clusters. As further work, we will develop methods for identifying such cases automatically.

In addition, we have studied the metrics on two repositories. It would be interesting to include more repositories in further studies, which could contribute to finding metrics with (in)consistent behaviour across repositories. Moreover, the analysis of the similarity between the repositories by using the q -dimensional profiles could generate new insights.

The present study shares some limitations with [33], as we did not include metrics such as consistency or formal correctness, which are usually implemented as Boolean functions. In this study, we considered the minimum $k = 3$, as we believe that our method is less interesting for Boolean-like metrics and thus keeps the variability of the metrics, avoiding elementary binary classifications. We also did not include non-structural metrics, as we wanted to perform the study under the same conditions as those described in [33]. Finally, we proposed the repository profiles, which could be used to determine the homogeneity degree of repositories with respect to a set of metrics. Further experiments will be carried out to associate homogeneity degrees with ranges of similarity scores.

V. CONCLUSION

In this paper, we have presented a method for an automatic decision-making process based on the optimal analysis of

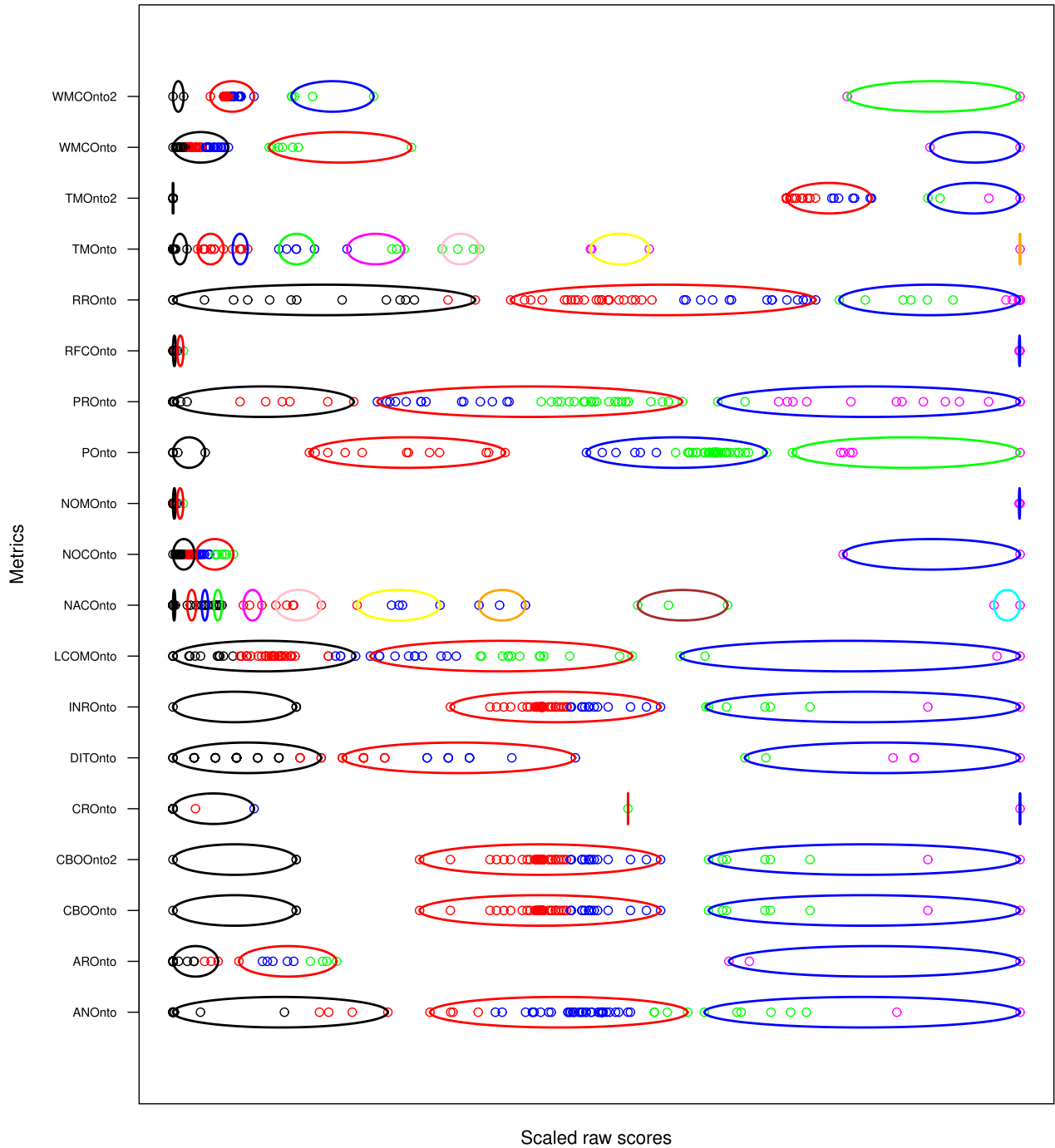


FIGURE 13. Comparison of the optimal k and $k = 5$ for θ^{agr} . Clusters for the optimal k are ellipses and the ones for $k = 5$ are represented by the colors of the data points.

repositories from the perspective of the features measured by using a set of metrics. The method is based on measuring and analysing the statistical properties of the clusterings obtained by applying the metrics to an ontology repository. The results showed that the use of the optimal k value for such an analysis allowed for a finer distinction of the profile of the ontology repositories. We believe that this type of study may well help users to generate new insights into the properties of an

ontology repository and its content. The method presented was applied to ontologies but is generic and, therefore, can be applied to other types of metrics and repositories.

**APPENDIX
ADDITIONAL TABLES AND FIGURES**

A. TABLES

See Tables 7–9.

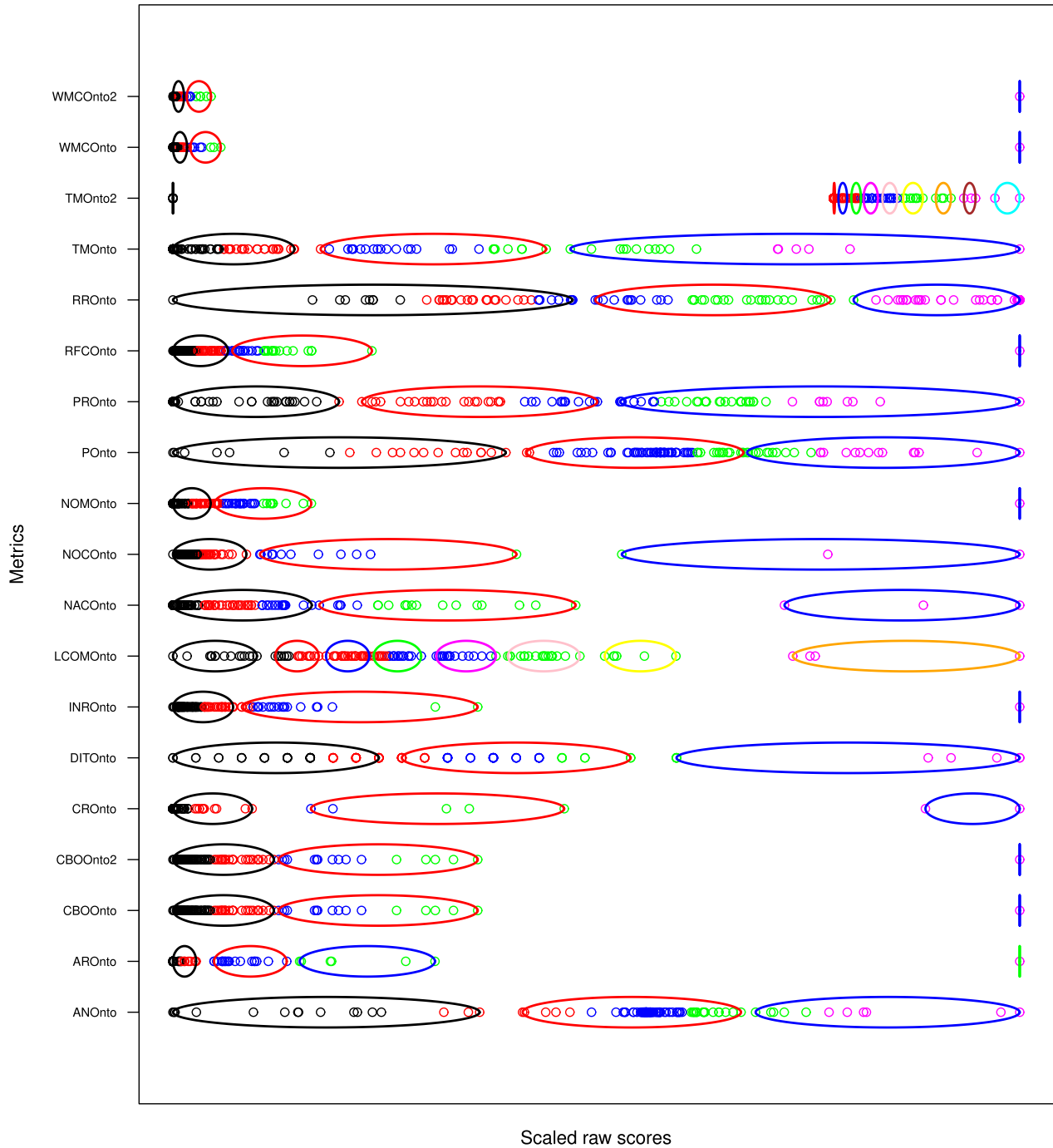


FIGURE 14. Comparison of the optimal k and $k = 5$ for θ^{obo} . Clusters for the optimal k are ellipses and the ones for $k = 5$ are represented by the colors of the data points.

B. FIGURES

See Figs. 7–14.

REFERENCES

[1] M. D. C. Legaz-García, C. Martínez-Costa, M. Menárguez-Tortosa, and J. T. Fernández-Breis, “A semantic Web based framework for the interoperability and exploitation of clinical models and EHR data,” *Knowl.-Based Syst.*, vol. 105, pp. 175–189, Aug. 2016.

[2] J. T. Fernández-Breis, D. Castellanos-Nieves, J. Hernández-Franco, C. Soler-Segovia, M. D. C. Robles-Redondo, R. González-Martínez, and M. P. Prendes-Espinosa, “A semantic platform for the management of the educative curriculum,” *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6011–6019, Apr. 2012.

[3] Y. Qin, W. Lu, Q. Qi, X. Liu, M. Huang, P. J. Scott, and X. Jiang, “Towards an ontology-supported case-based reasoning approach for computer-aided tolerance specification,” *Knowl.-Based Syst.*, vol. 141, pp. 129–147, Feb. 2018.

[4] D. L. Rubin, N. H. Shah, and N. F. Noy, “Biomedical ontologies: A functional perspective,” *Briefings Bioinf.*, vol. 9, no. 1, pp. 75–90, Oct. 2007.

- [5] S. El-Sappagh, J. M. Alonso, F. Ali, A. Ali, J.-H. Jang, and K.-S. Kwak, "An ontology-based interpretable fuzzy decision support system for diabetes diagnosis," *IEEE Access*, vol. 6, pp. 37371–37394, 2018.
- [6] Y. Ma, P. Zhang, and J. Ma, "An ontology driven knowledge block summarization approach for Chinese judgment document classification," *IEEE Access*, vol. 6, pp. 71327–71338, 2018.
- [7] J. Liu, X. Zhang, Y. Li, J. Wang, and H.-J. Kim, "Deep learning-based reasoning with multi-ontology for IoT applications," *IEEE Access*, vol. 7, pp. 124688–124701, 2019.
- [8] R. Hoehndorf, P. N. Schofield, and G. V. Gkoutos, "The role of ontologies in biological and biomedical research: A functional perspective," *Briefings Bioinf.*, vol. 16, no. 6, pp. 1069–1080, 2015.
- [9] C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. D. Yeumo, V. Emonet, J. Graybeal, M.-A. Laporte, M. A. Musen, V. Pesce, and P. Larmande, "AgroPortal: A vocabulary and ontology repository for agronomy," *Comput. Electron. Agricult.*, vol. 144, pp. 126–143, Jan. 2018.
- [10] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, and N. Leontis, "The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration," *Nature Biotechnol.*, vol. 25, no. 11, pp. 1251–1255, 2007.
- [11] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, "BioPortal: Enhanced functionality via new Web services from the national center for biomedical ontology to access and use ontologies in software applications," *Nucleic Acids Res.*, vol. 39, no. 1, pp. W541–W545, Jul. 2011.
- [12] Z. Xiang, C. Mungall, A. Ruttenberg, and Y. He, "Ontobee: A linked data server and browser for ontology terms," in *Proc. ICBO*, 2011, pp. 1–3.
- [13] R. Cote, F. Reisinger, L. Martens, H. Barsnes, J. A. Vizcaino, and H. Hermjakob, "The ontology lookup service: Bigger and better," *Nucleic Acids Res.*, vol. 38, pp. W155–W160, Jul. 2010.
- [14] R. Hoehndorf, L. Slater, P. N. Schofield, and G. V. Gkoutos, "Aber-OWL: A framework for ontology-based data access in biology," *BMC Bioinf.*, vol. 16, no. 1, p. 26, Dec. 2015.
- [15] F. Neuhaus, A. Vizedom, K. Baclawski, M. Bennett, M. Dean, M. Denny, M. Grüninger, A. Hashemi, T. Longstreth, L. Obrst, S. Ray, R. Sriram, T. Schneider, M. Vegetti, M. West, and P. Yim, "Towards ontology evaluation across the life cycle," *Appl. Ontol.*, vol. 8, no. 3, pp. 179–194, 2013.
- [16] A. Gangemi, C. Catenacci, M. Ciarmita, and J. Lehmann, "Modelling ontology evaluation and validation," in *The Semantic Web: Research and Applications*. Berlin, Germany: Springer, 2006, pp. 140–154.
- [17] J. E. Rogers, "Quality assurance of medical ontologies," *Methods Inf. Med.*, vol. 45, no. 3, pp. 267–274, 2006.
- [18] W. Ceusters, "Applying evolutionary terminology auditing to the gene ontology," *J. Biomed. Inform.*, vol. 42, no. 3, pp. 518–529, Jun. 2009.
- [19] A. Duque-Ramos, J. T. Fernández-Breis, R. Stevens, and N. Aussenac-Gilles, "OQuaRE: A square-based approach for evaluating the quality of ontologies," *J. Res. Pract. Inf. Technol.*, vol. 43, no. 2, pp. 159–176, 2011.
- [20] S. Kumar, N. Baliyan, and S. Sukalika, "Ontology cohesion and coupling metrics," *Int. J. Semantic Web Inf. Syst.*, vol. 13, no. 4, pp. 1–26, Oct. 2017.
- [21] M. McDaniel, V. C. Storey, and V. Sugumaran, "Assessing the quality of domain ontologies: Metrics and an automated ranking system," *Data Knowl. Eng.*, vol. 115, pp. 32–47, May 2018.
- [22] M. Poveda-Villalón, A. Gómez-Pérez, and M. C. Suárez-Figueroa, "OOPS! (Ontology pitfall scanner!): An on-line tool for ontology evaluation," *Int. J. Semantic Web Inf. Syst.*, vol. 10, no. 2, pp. 7–34, Apr. 2014.
- [23] S. Tartir and I. B. Arpinar, "Ontology evaluation and ranking using OntoQA," in *Proc. Int. Conf. Semantic Comput. (ICSC)*. Washington, DC, USA: IEEE Computer Society, Sep. 2007, pp. 185–192, doi: 10.1109/ICSC.2007.65.
- [24] H. Yao, A. M. Orme, and L. Etkorn, "Cohesion metrics for ontology design and application," *J. Comput. Sci.*, vol. 1, no. 1, pp. 107–113, Jan. 2005.
- [25] H. Zhu, D. Liu, I. Bayley, A. Aldea, Y. Yang, and Y. Chen, "Quality model and metrics of ontology for semantic descriptions of Web services," *Tsinghua Sci. Technol.*, vol. 22, no. 3, pp. 254–272, Jun. 2017.
- [26] N. Matentzoglou, J. Malone, C. Mungall, and R. Stevens, "MIRO: Guidelines for minimum information for the reporting of an ontology," *J. Biomed. Semantics*, vol. 9, no. 1, p. 6, Dec. 2018.
- [27] T. Hall and N. Fenton, "Implementing effective software metrics programs," *IEEE Softw.*, vol. 14, no. 2, pp. 55–65, Mar./Apr. 1997.
- [28] R. K. Lind and K. Vairavan, "An experimental investigation of software metrics and their relationship to software development effort," *IEEE Trans. Softw. Eng.*, vol. 15, no. 5, pp. 649–653, May 1989.
- [29] A. Abbott, D. Cyranoski, N. Jones, B. Maher, Q. Schiermeier, and R. Van Noorden, "Metrics: Do metrics matter?" *Nature*, vol. 465, no. 7300, pp. 860–862, Jun. 2010.
- [30] S. M. Alghamdi, B. A. Sundberg, J. P. Sundberg, P. N. Schofield, and R. Hoehndorf, "Quantitative evaluation of ontology design patterns for combining pathology and anatomy ontologies," *Sci. Rep.*, vol. 9, no. 1, p. 4025, Dec. 2019.
- [31] A. Duque-Ramos, J. T. Fernández-Breis, M. Iniesta, M. Dumontier, M. Egaña Aranguren, S. Schulz, N. Aussenac-Gilles, and R. Stevens, "Evaluation of the OQuaRE framework for ontology quality," *Expert Syst. Appl.*, vol. 40, no. 7, pp. 2696–2703, Jun. 2013.
- [32] A. Duque-Ramos, M. Quesada-Martínez, M. Iniesta-Moreno, J. T. Fernández-Breis, and R. Stevens, "Supporting the analysis of ontology evolution processes through the combination of static and dynamic scaling functions in OQuaRE," *J. Biomed. Semantics*, vol. 7, no. 1, p. 63, Dec. 2016.
- [33] M. Franco, J. M. Vivo, M. Quesada-Martínez, A. Duque-Ramos, and J. T. Fernández-Breis, "Evaluation of ontology structural metrics based on public repository data," *Briefings Bioinf.*, vol. 21, no. 2, pp. 473–485, Mar. 2020.
- [34] M. Franco and J.-M. Vivo, "Cluster analysis of microarray data," in *Microarray Bioinformatics*, V. Bolón-Canedo and A. Alonso-Betanzos, Eds. New York, NY, USA: Springer, 2019, pp. 153–183, doi: 10.1007/978-1-4939-9442-7_7.
- [35] M. Alhawarat and M. Hegazi, "Revisiting K-means and topic modeling, a comparison study to cluster Arabic documents," *IEEE Access*, vol. 6, pp. 42740–42749, 2018.
- [36] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020.
- [37] L. Kaufman and P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 1990.
- [38] E. Schubert and P. J. Rousseeuw, "Faster K-medoids clustering: Improving the PAM, CLARA, and CLARANS algorithms," in *Similarity Search and Applications*, G. Amato, C. Gennaro, V. Oria, and M. Radovanović, Eds. Cham, Switzerland: Springer, 2019, pp. 171–187.
- [39] C. Hennig, "Cluster-wise assessment of cluster stability," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 258–271, Sep. 2007.
- [40] C. Jaccard, "Distribution de la flore alpine dans le bassin de dranses et dans quelques régions voisines," *Bull. Societe Vaudoise Sci. Naturelles*, vol. 37, no. 140, pp. 241–272, 1901.
- [41] C. Hennig. (2019). *Fpc: Flexible Procedures for Clustering, R Package Version 2.2-3*. [Online]. Available: <https://CRAN.R-project.org/package=fpc>
- [42] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.
- [43] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Commun. Statist.-Simul. Comput.*, vol. 3, no. 1, pp. 1–27, 1974.
- [44] J. C. Dunn, "Well-separated clusters and optimal fuzzy partitions," *J. Cybern.*, vol. 4, no. 1, pp. 95–104, Jan. 1974.
- [45] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979.
- [46] *Evaluome-Evaluation of Metrics*. Accessed: Dec. 23, 2019. [Online]. Available: <http://semantics.inf.um.es/evaluome/>
- [47] J. A. Bernabé-Díaz, M. Franco, J.-M. Vivo, M. Quesada-Martínez, A. Duque-Ramos, and J. T. Fernández-Breis, "EvaluomeR: Evaluation of bioinformatics metrics," R package version 1.0.63, Bioconductor, 2019, doi: 10.18129/B9.bioc.evaluomeR.
- [48] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, M. Studer, P. Roudier, J. Gonzalez, and K. Kozłowski. (2018). *R Package 'Cluster': Finding Groups in Data: Cluster Analysis Extended Rousseeuw et al. (2.0.7-1)*. [Online]. Available: <https://cran.r-project.org/package=cluster>
- [49] T. Wei and V. Simko. (2017). *R package 'Corrplot': Visualization of a Correlation Matrix, (V. 0.84)*. [Online]. Available: <https://github.com/taiyun/corrplot>
- [50] G. N. Boshnakov and D. Murdoch, "Rdpack: Update manipulate RD documentation objects," R package V. 0.10-1, The R Foundation, 2018.
- [51] J. Lemon, "Plotrix: A package in the red light district of R," *R-News*, vol. 6, no. 4, pp. 8–12, 2006.
- [52] C. Hennig. (2018). *R Package, 'FPC': Flexible Procedures for Clustering*. [Online]. Available: <https://cran.r-project.org/package=fpc>

[53] M. Morgan, V. Obenchain, J. Hester, and H. Pagès, “Summarizedexperiment: Summarizedexperiment container,” package v. 1.12.0, R package v. 1.12.0, Bioconductor, 2018.

[54] M. Ramos, L. Schiffer, A. Re, R. Azhar, A. Basunia, C. R. Cabrera, T. Chan, P. Chapman, S. Davis, D. Gomez-Cabrero, A. C. Culhane, B. Haiibe-Kains, K. Hansen, H. Kodali, M. S. Louis, A. S. Mer, M. Reister, M. Morgan, V. Carey, and L. Waldron, “Software for the integration of multi-omics experiments in bioconductor,” *Cancer Res.*, vol. 77, no. 21, pp. e39–e42, 2017.

[55] R Foundation for Statistical Computing, Vienna, Austria, 2019. *R: A Language and Environment For Statistical Computing*. [Online]. Available: <https://www.R-project.org/>

[56] J. L. Hintze and R. D. Nelson, “Violin plots: A box plot-density trace synergism,” *Amer. Statistician*, vol. 52, no. 2, pp. 181–184, May 1998.

[57] J. D. Davis and E. O. Voit, “Metrics for regulated biochemical pathway systems,” *Bioinformatics*, vol. 35, no. 12, pp. 2118–2124, Jun. 2019, doi: 10.1093/bioinformatics/bty942.

[58] M. J. del Moral, F. Chiclana, J. M. Tapia, and E. Herrera-Viedma, “A comparative study on consensus measures in group decision making,” *Int. J. Intell. Syst.*, vol. 33, no. 8, pp. 1624–1638, Aug. 2018.

[59] S. Imbeaud, E. Graudens, V. Boulanger, X. Barlet, P. Zaborski, E. Eveno, O. Mueller, A. Schroeder, and C. Auffray, “Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces,” *Nucleic Acids Res.*, vol. 33, no. 6, p. e56, Mar. 2005.

[60] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[61] S. R. Chidamber and C. F. Kemerer, “A metrics suite for object oriented design,” *IEEE Trans. Softw. Eng.*, vol. 20, no. 6, pp. 476–493, Jun. 1994.

[62] A. Duque-Ramos, J. T. Fernández-Breis, R. Stevens, and N. Aussenac-Gilles, “Oquare: A square-based approach for evaluating the quality of ontologies,” *J. Res. Pract. Inf. Technol.*, vol. 43, no. 2, p. 159, 2011.

[63] W. Li, “Another metric suite for object-oriented programming,” *J. Syst. Softw.*, vol. 44, no. 2, pp. 155–162, Dec. 1998.

[64] M. D. Wilkinson, S.-A. Sansone, E. Schultes, P. Doorn, L. O. Bonino da Silva Santos, and M. Dumontier, “A design framework and exemplar metrics for fairness,” *Sci. Data*, vol. 5, no. 1, pp. 1–4, Dec. 2018.



JUANA MARÍA VIVO-MOLINA received the B.Sc. and M.Sc. degrees in mathematics from the University of Murcia, Spain, in 1993, and the Ph.D. degree from the University of Granada, Spain, in 2005. Since 2019, she has been a member of the IMIB-Arrixaca Bio-Health Research Institute. She is currently a Professor with the Statistics and Operations Research Department, University of Murcia. Her research interests include biostatistics, reliability systems and applications, and statistical methods in bioinformatics.



MANUEL QUESADA-MARTÍNEZ received the Computer Engineering degree and the Ph.D. degree in computer science from the University of Murcia, in 2009 and 2015, respectively. He is currently a Ph.D. Assistant Professor with the University Miguel Hernandez of Elche. He is also a member of the research group “Decision Support Web Systems”. His research interests focus on web semantics technologies, such as the axiomatic enrichment of biomedical resources and quality assurance of ontologies, and human-computer interaction, such as the integration of these technologies in e-health systems and applications.



ASTRID DUQUE-RAMOS received the Computer Engineering degree from the University of Antioquia, Medellín, Colombia, in 2005, and the Ph.D. degree in computer science from the University of Murcia, Spain, in 2016. She is currently a Professor with the School of Engineering, Latin American Autonomous University, Medellín. She is also a member of the research group “INGECO de UNAULA.” Her research interest includes web semantics technologies in three different domains: biomedical domain, specifically quality assurance of ontologies, ontologies for knowledge transfer, and ontologies for publishing domain.



JOSÉ ANTONIO BERNABÉ-DÍAZ received the B.S. degree in computer engineering and the M.S. degree in bioinformatics from the University of Murcia, Spain, in 2016 and 2017, respectively, where he is currently pursuing the Ph.D. degree in computer engineering. From 2017 to 2020, he was a Research Assistant with the Department of Computing and Systems, University of Murcia. His research interests focus on web semantics technologies, such as transformation to heterogeneous

to homogeneous data or the evaluation of ontologies through metrics and bioinformatics, where currently a research to characterize several varieties of *Cucumis melo* is carried out.



MANUEL FRANCO-NICOLÁS received the B.Sc. and M.Sc. degrees in mathematics and the Ph.D. degree in Mathematics from the University of Murcia, Spain, in 1991 and 1995, respectively. Since 1998, he has been a Professor with the Statistics and Operations Research Department, University of Murcia. Since 2019, he has been a member of the IMIB-Arrixaca Bio-Health Research Institute. His research interests include biostatistics, reliability systems and applications, and statistical methods in bioinformatics.



JESUALDO TOMÁS FERNÁNDEZ-BREIS (Member, IEEE) received the Computer Engineering and Ph.D. degrees in computer science from the University of Murcia, Spain, in 1999 and 2003, respectively. He is currently a Full Professor with the Faculty of Computer Science, University of Murcia. He is also a member of the IMIB-Arrixaca Bio-Health Research Institute. He has been leading research projects related to semantic web technologies since 2004. His current research interests include the application of semantic technologies for the development of learning health systems and the development of quality assurance methods for ontologies and terminologies.

...