Advanced Engineering Informatics 56 (2023) 101970

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

Full length article

A quantitative analysis of parametric CAD model complexity and its relationship to perceived modeling complexity

Manuel Contero [a,*], David Pérez-López [a], Pedro Company [b], Jorge D. Camba [c]

[a] Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, Universitat Politècnica de València, Valencia, Spain
[b] Dpto. de Ingeniería Mecánica y Construcción, Universitat Jaume I, Castellón, Spain
[c] School of Engineering Technology, Purdue University, West Lafayette, IN, USA

ARTICLE INFO

Keywords:
CAD model complexity
Perceived complexity
CAD model complexity metrics
CAD model reusability

ABSTRACT

Digital product data quality and reusability has been proven a critical aspect of the Model-Based Enterprise to enable the efficient design and redesign of products. The extent to which a history-based parametric CAD model can be edited or reused depends on the geometric complexity of the part and the procedure employed to build it. As a prerequisite for defining metrics that can quantify the quality of the modeling process, it is necessary to have CAD datasets that are sorted and ranked according to the complexity of the modeling process. In this paper, we examine the concept of *perceived CAD modeling complexity,* defined as the degree to which a parametric CAD model is perceived as difficult to create, use, and/or modify by expert CAD designers. We present a novel method to integrate pair-wise comparisons of CAD modeling complexity made by experts into a single metric that can be used as ground truth. Next, we discuss a comprehensive study of quantitative metrics which are derived primarily from the geometric characteristics of the models and the graph structure that represents the parent/child relationships between features. Our results show that the perceived CAD modeling complexity metric derived from experts' assessment correlates particularly strongly with graph-based metrics. The Spearman coefficients for five of these metrics suggest that they can be effectively used to study the parameters that influence the reusability of models and as a basis to implement effective personalized learning strategies in online CAD training scenarios.

1. Introduction

This work is part of a broader initiative by the authors to develop new intelligent methods and tools for improving the quality and reusability of parametric CAD models, which are highly relevant topics today in the context of the Model-Based Enterprise (MBE). MBE is a paradigm where annotated 3D CAD models serve as primary elements to support the design, analysis, and manufacturing of industrial products [40]. In an MBE, the quality of the native CAD models (typically parametric feature-based solid models) is paramount, as they are the primary source from which the secondary models required for CAE and CAM purposes derive. In the context of this paper, only models that are featured-based and contain a construction history are considered. Neutral files derived from the native CAD files are out of the scope of the paper.

Despite the vital role of CAD quality in the various aspects of the product development process, particularly in an MBE context [9,14,28,17], there are no standard metrics to quantitatively evaluate parametric models and estimate their responses to changes [16].

CAD models are often reused for future redesigns, shared within collaborative design teams, and used as the basis for automated design and shape optimization tools. In this context, the ability of a parametric model to adjust to changes is critical. It corresponds with the "semantic level" in the CAD quality model proposed by Contero et al. [23].

The ability to alter the geometry of a parametric model depends on intrinsic factors (i.e., the geometric complexity of the part to be modeled) and extrinsic factors (i.e., the modeling methodology and practices employed to create it that are reflected on the complexity of the feature tree). A thorough examination of complexity metrics (many of which are borrowed from other areas such as graph theory and software engineering) and the definition of a generalized instrument that can produce a reliable figure of merit to quantify the quality of a parametric model remain open problems.

In this paper, we build on our previous works [15,16] to present a

* Corresponding author.
E-mail addresses: mcontero@upv.es (M. Contero), dapelo@i3b.upv.es (D. Pérez-López), pcompany@uji.es (P. Company), jdorribo@purdue.edu (J.D. Camba).

https://doi.org/10.1016/j.aei.2023.101970
Received 2 November 2022; Received in revised form 28 January 2023; Accepted 3 April 2023
Available online 25 April 2023

Fig. 1. Sample of 95 models used in our study.

study in which we examined, analyzed, and compared a series of complexity metrics for characterizing parametric CAD models. Specifically, we study the assessment of parametric part modeling complexity as perceived by designers, and compare it to various quantitative complexity metrics. The paper is structured as follows: first, we provide an overview of the relevant literature on CAD quality metrics and parametric model complexity. Next, we describe the methodology used to integrate pair-wise comparisons of CAD modeling complexity made by experts into a single metric that can be used as ground truth. To the best of our knowledge, this is the first time this idea has been proposed in the context of CAD complexity metrics. Next, we discuss a comprehensive study of quantitative metrics which are derived primarily from the geometric characteristics of the models and the graph structure that represents the parent/child relationships between features, followed by the correlation analysis between these metrics and the CAD modeling complexity metric derived from experts, which proves to be a valid

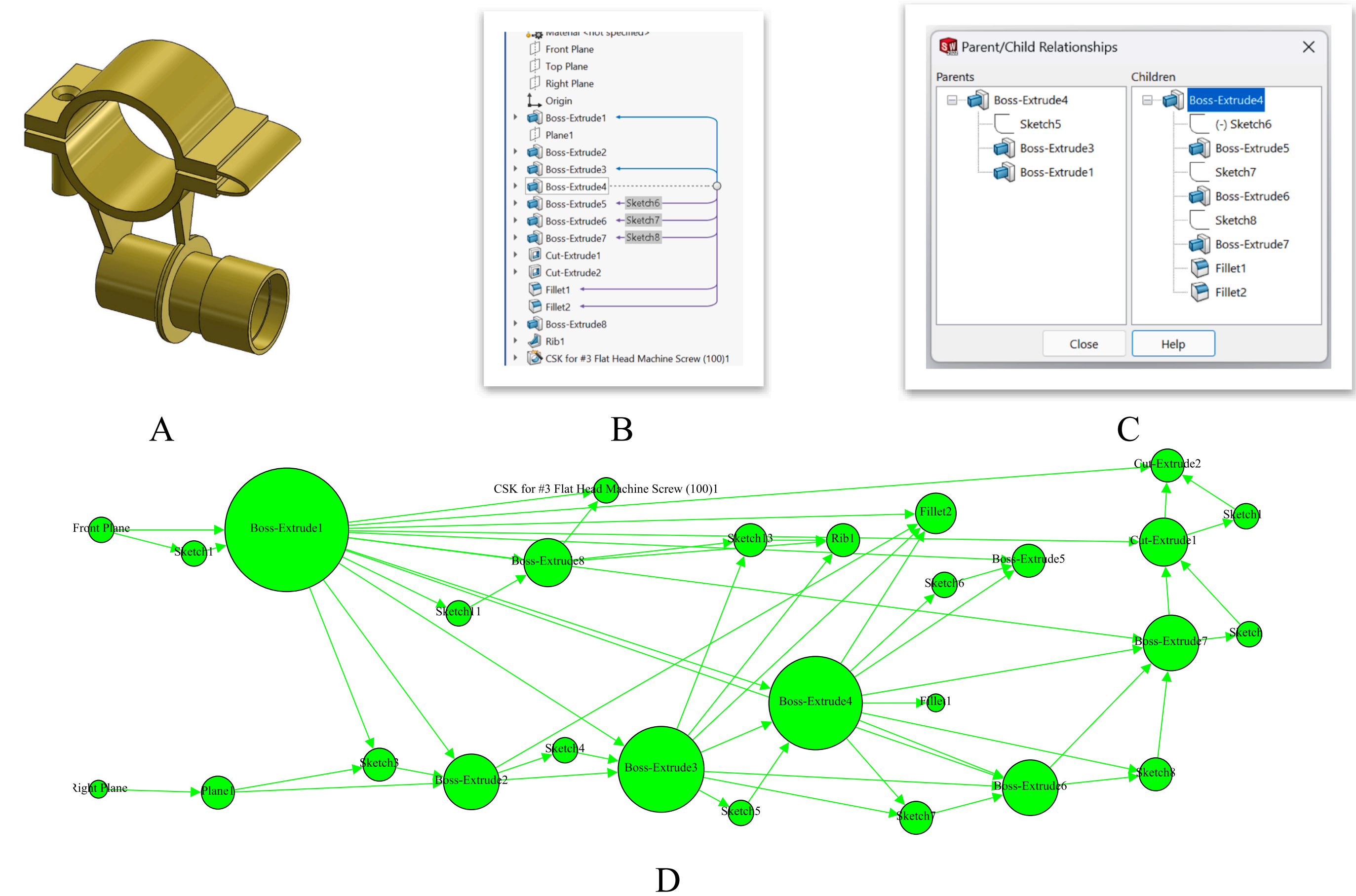

Fig. 2. 3D Part (A), feature tree (B), parent/child relationships for a feature (C), and full graph of the model (D).

ground truth to determine the validity of current (and hopefully future) complexity metrics. Finally, we discuss the results of our statistical analyses and the implications for the quantitative assessment of parametric models and identify directions for future work.

2. Background

3D model complexity has significant effects in the design and manufacturing fields [48,57], particularly in activities that rely heavily on the CAD model and the communication of design intent [10]. The nature of additive manufacturing processes, for example, has enabled a significant increase in terms of complexity of the models that can be manufactured [39]. However, it is difficult to agree on a precise definition as well as to develop objective reliable metrics to assess and compare CAD models [3]. One reason has to do with the various levels at which complexity can be understood and the various sources that can influence complexity, such as the geometric characteristics of the model (e.g., number of faces, edges, etc.), the internal data structure and organization, and the properties associated with the software functionality [20].

Rossignac [46] identified five dimensions of complexity related to CAD models: algebraic, topological, morphological, combinatorial, and representational. Algebraic complexity refers to the complexity of the polynomials that are used to represent the shape exactly in its implicit or parametric form. Topological complexity refers to the number of handles and components or the existence of non-manifold singularities, non-regularized components, holes or self-intersections. Morphological complexity is related to feature size and “smoothness” (e.g., components that have more, and smaller features would be deemed more complex). Combinatorial complexity refers to the vertex count in a polynomial mesh. Representational complexity is a measure of the file size and the ease-of-use of the data structure employed to store the model [46]. These types of complexity are informed by both the geometric representation of the component as well as the design intent, i.e., the manner in which modeling decisions (and thus the constraints imposed on the geometry) influence how the final model is built [32].

A significant body of work has been devoted to developing quantitative metrics to evaluate the geometric complexity of a model. Some of these metrics include the number of triangles, surfaces, and vertices in the representation of the model [56,25,46] shape measurements based on area and volume [33] and the similarity between its views [47]. However, geometric complexity is only one aspect of a broader definition, which includes the complexity of the CAD data, the organizational system, and the operational complexity, to name a few [20].

In the context of parametric CAD, there are many aspects that contribute to the complexity of a model, specifically in terms of its internal structure and organization. For example, the parametric relationships between features and sketches and the particular constraining strategies and decisions used to build the model can have a significant impact on robustness and reusability. From a user’s standpoint, these aspects are critical, as modeling complex parts requires strategic and effective modeling strategies to manage dependencies efficiently, which can be extremely challenging even for experienced users [7,31,5]. According to Camba et al. [15], CAD quality metrics can be classified based on the specific dimension of quality that they measure.

Some researchers have adopted complexity metrics from areas such as graph theory and software engineering and applied them to parametric CAD modeling [36,24]. Graph theory is particularly useful in parametric CAD to represent and visualize constraint relationships and

Table 1
Metrics analyzed in our study.

Metric type	Metric	Description
Complexity scale	S_MAP	Maximum a posteriori (MAP) complexity scale estimation based on Thurstone's case V model [55]
Computer file based	File size (MBytes)	Part file size measured on Mbytes
Geometry	No. of faces	Total number of individual faces in the model
	No. of vertices	Total number of individual vertices in the model
	No. of edges	Total number of individual edges in the model
	Volume/area ratio	Ratio between the object volume and its surface
	Volume ratio	Calculated as 1–Volume/bBoxVol, where bBoxVol is the volume of the bounding box of the model
	Cube ratio	Calculated as 1–aCube/Surface, where aCube is the area of a cube with the same volume as the model
	Sphere ratio	Calculated as 1–aSphere/Surface, where aSphere is the area of a sphere with the same volume as the model
	Fractal dimension	Calculated as presented in [12], that is, the slope of log(N(s)) vs log(1/s), where s is the size of voxel and N(s) is the number of voxels needed to fill the model with voxels of s size
Graph	Size	No. of nodes in the graph
	Cyclomatic complexity	No. of independent paths through the graph [24]
	Modularity	Measure of the strength of division of the graph into groups [38]
	Diameter	Maximum eccentricity (greatest distance) between any two nodes in the graph
	Mean degree	Degree of a vertex is the number of its adjacent edges
	Graph density	Ratio between number of edges and number of possible edges
	Average path	Average number of steps along the shortest paths for all possible pairs of nodes
	Dependencies	No. of edges in the graph
	Kolmogorov complexity	String length when the graph is encoded as a binary string [41]
	Li entropy	Graph entropy measure, describe the uncertainty of a system following the algorithm from [37]
	Solé-Valverde entropy	Graph entropy measure, describe the uncertainty of a system following the algorithm from [52]
	Shannon entropy	Graph entropy measure, describe the uncertainty of a system following the algorithm from [50].

parametric dependencies [30,18]. A model can be represented as a graph G = (V, E) where V is a set of "vertices" or "nodes" in the model tree (2D sketches or features) and the "edges" E represent the relationships between sketches and/or features). Similar studies have demonstrated the value strategies grounded on graph theory for determining differences [43] and similarities [51] between parametric models, and for estimating assembly time [42].

It has been argued that complexity is an inherently subjective concept; what is complex depends upon how one looks [19]. In this regard, the role of the observer in the acknowledgement of complexity can be characterized as "perceived complexity" [19,49]. Indeed, complexity can be understood as the result of a particular perception of a situation made by an observer. In the context of CAD, Johnson et al. [32] stated that an objective quantitative assessment of model complexity should also be well correlated to the subjective assessment of complexity by CAx tool users. In their study, the authors examined the relationship between perceived complexity and a series of objective metrics. They concluded that the volume ratio (calculated as 1–Volume/bboxVol, where bboxVol is the volume of the bounding box of the model) and the normalized volume ratio (measured by number of features) are significantly correlated with subjective assessments of model complexity. Subjective assessments were conducted through 5-point Likert scale questions in which a group of participants scored a series of models from "very simple" to "very complex.".

3. Methodology

In this section, we present the dataset of parametric 3D models and the complexity metrics used in the study. We also describe our strategy to obtain a ground truth based on the integration of the assessments of a group of CAD experts.

3.1. Dataset

We compiled a set of 569 parametric CAD models of mechanicals parts of varying complexity. We restricted our selection to native SolidWorks (.sldprt) files and used the analysis software developed in our previous works. Our models were obtained from six sources:

- GrabCAD, an online community where students, engineers, designers, and manufactures can share their models [26].
- Official Solidworks training files, a public repository of SolidWorks tutorials and files [54].
- SolidWorks Exercises – Learn by Practicing, a practice CAD book with exercises [53].
- 3D Content Central, an online community that provides access to 3D models from component suppliers and individuals in all major CAD formats [1].
- SPECapc, a performance evaluation software for vendors and users of computing systems running SolidWorks CAD/CAM software on Microsoft Windows 10 64-bit platforms. The software includes several 3D models for testing purposes [54].
- "CAD 3D con SolidWorks vol. I & II," a two-volume CAD book with SolidWorks exercises [21,22].

Because of the computational costs of the sorting algorithms and the difficulty of having a single expert sort all the models in the sample, we considered a theoretical maximum of 1,000 comparisons and, thus, selected a representative subset of 95 models randomly chosen from the total sample. The selection included only models built as a single solid (i.e., excluding assemblies and multi-body parts). The selected models used in our study are shown in Fig. 1.

In order to characterize parametric models according to complexity, we selected a group of metrics based on an extensive literature review that spanned a number of disciplines (e.g., software engineering, graph theory, engineering design, geometry, etc.) and a study on their applicability to parametric CAD. Our selection criteria included attributes related both to the part geometry and the metrics based on the representation of the parametric model tree as a graph structure.

Model trees do not fully and accurately reflect the intricacies of the interdependencies that are created between features during the parametric modeling process. The part shown in Fig. 2 is presented as an example to illustrate this problem. The model tree of the 3D part (Fig. 2-A), modeled in the commercial CAD system SolidWorks, is shown in Fig. 2-B as presented by the software (with dynamic reference visualization enabled). The blue arrows indicate parent features whereas the purple arrows represent child features. All the arrows originate from a circle, which indicates the currently selected feature (Boss-Extrude4, in the example). An alternative method to visualize these parent/child relationships is through the use of commands to display dependencies, such as the window shown in Fig. 2-C. However, these visualizations cannot convey the full extent of the complexity of the interdependencies

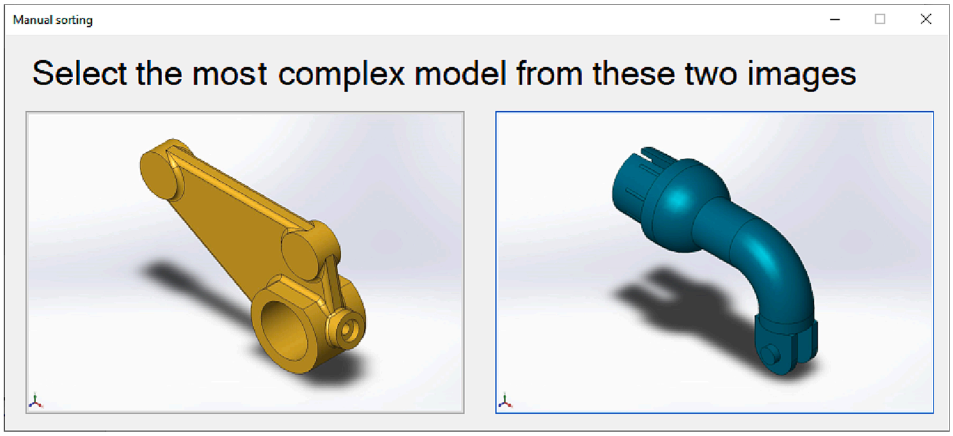

Fig. 3. Software tool for paired comparisons.

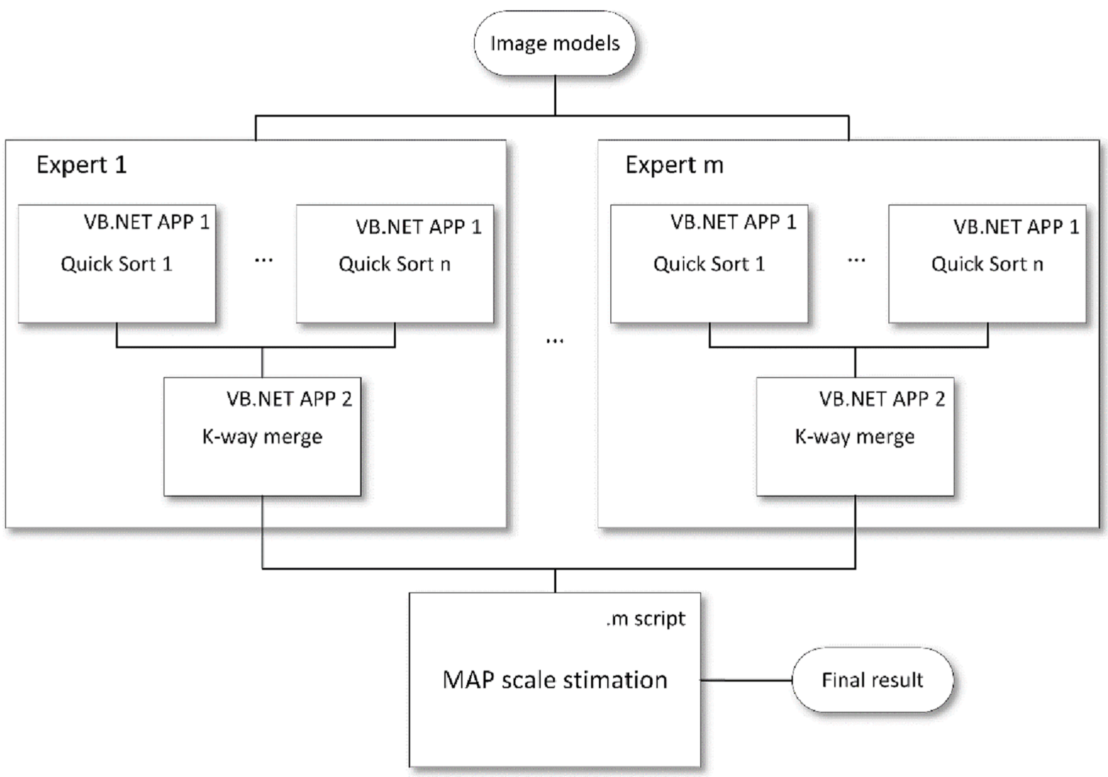

Fig. 4. Block diagram.

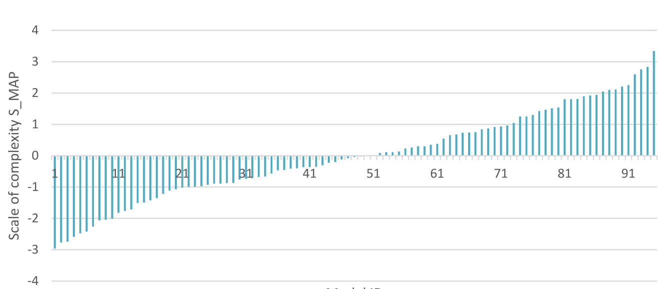

Fig. 5. Scale values vs. model ID representation.

between features, as illustrated in the complete graph shown Fig. 2-D, which paints a more realistic picture of the actual complexity within the internal structure of the parametric model by showing the dependencies that are not visible in the standard visualization mechanisms provided by current CAD tools. This graph was produced in Gephi, an application developed by Bastian et al. [6]. As part of our work, we developed a custom plugin for SolidWorks that can extract the parent/child relationships in the CAD model and export it to a GEXF (Graph Exchange XML Format) file, which can be imported in Gephi for visualization. In this example, the sizes of the nodes in the graph were made proportional to the number of edges that meet at those nodes to emphasize the most highly interconnected features in the model.

The metrics are described in Table 1 and are mainly based on a selection of those compiled by Camba et al. [16]. The metrics *volume* and *surface* were excluded since non-normalized metrics are not relevant in our study. *Dimensionality* was also excluded as we assume that there

Table 2
Descriptive statistics.

Metric type	Metric	M	SD
Complexity scale	S_MAP	0.00	1.44
Computer file based	File size (MBytes)	0.61	0.58
Geometry	No. of faces	160.63	160.11
	No. of vertices	228.19	258.29
	No. of edges	389.57	413.98
	Volume/area ratio	0.003	0.002
	Volume ratio	0.78	0.12
	Cube ratio	0.63	0.14
	Sphere ratio	0.70	0.11
	Fractal dimension	2.79	0.09
Graph	Size	41.37	38.37
	Cyclomatic complexity	70.47	80.92
	Modularity	0.37	0.11
	Diameter	4.18	1.19
	Mean degree	4.23	0.88
	Graph density	0.08	0.04
	Average path	2.23	0.40
	Dependencies	97.39	108.58
	Kolmogorov complexity	1260.85	1732.31
	Li entropy	296.28	402.16
	Solé-Valverde entropy	2.43	0.70
	Shannon entropy	0.90	0.08

Table 3
Spearman ρ correlations.

Metric type	Metric	Spearman ρ	*p-value*
Computer file based	File size (MBytes)	0.758	<0.001
Geometry	No. of faces	0.762	<0.001
	No. of vertices	0.713	<0.001
	No. of edges	0.755	<0.001
	Volume/area ratio	−0.059	0.57
	Volume ratio	0.414	<0.001
	Cube ratio	0.428	<0.001
	Sphere ratio	0.428	<0.001
	Fractal dimension	-0.497	<0.001
Graph	Size	0.792	<0.001
	Cyclomatic complexity	0.804	<0.001
	Modularity	0.680	<0.001
	Diameter	0.565	<0.001
	Mean degree	0.553	<0.001
	Graph density	−0.731	<0.001
	Average path	0.713	<0.001
	Dependencies	0.797	<0.001
	Kolmogorov complexity	0.801	<0.001
	Li entropy	0.799	<0.001
	Solé-Valverde entropy	0.277	<0.001
	Shannon entropy	0.398	<0.001

cannot be features without parents, except for the initial reference planes.

It is important to note that although some metrics may be mutually related, they are all used to assess model quality in an objective and quantitative manner. For example, the number of edges, vertices, and faces in a model are highly correlated, but they are all widely accepted metrics to evaluate geometric complexity. A similar situation occurs with graph-based metrics (number of paths, connectivity, etc.). In this paper, we do not intend to review the accuracy of these metrics or determine the extent to which they correlate to one another, but to provide a comprehensive view of complexity with well-established metrics so it can be compared to perceived complexity.

3.2. Ground truth for perceived CAD model complexity

We define perceived CAD model complexity as the degree to which a parametric CAD model is perceived as difficult to create, use, and/or modify. We obtain a ground truth by aggregating the evaluations of a group of experts. The procedure uses a pairwise comparison technique [44] in which each expert is shown side-by-side renderings of the most significant view of the models and asked to identify which of the two models is more complex. To facilitate the assessment and ensure all users receive the exact same information without manipulating the geometry, all models are presented in isometric view as a static image.

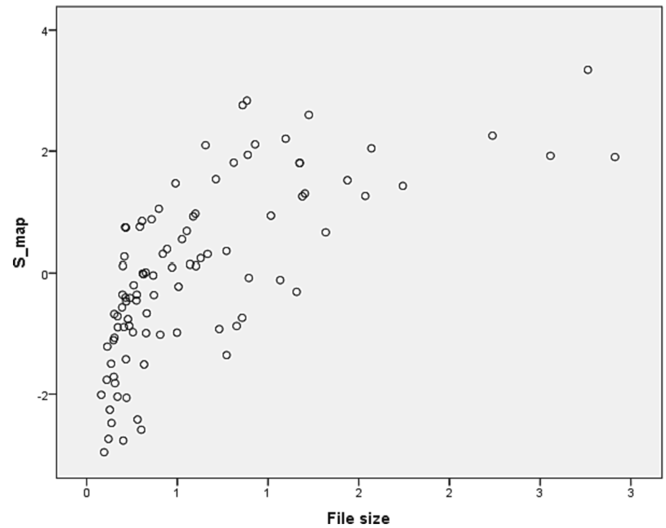

Fig. 6. Scatter plot for computer file-based metric: S_MAP vs. File size, **ρ** = 0.758.

To minimize the number of comparisons to be made by experts, we implemented a sorting tool based on the classical Quick Sort algorithm [29] using Microsoft Visual Studio.NET. The tool displays the images of the parts in pairs, as shown in Fig. 3 and asks the user to identify the most complex part. Ties are not allowed. Quick Sort was selected because is an efficient sorting algorithm; it takes on average $O(n \log(n))$ comparisons to sort n items. In the worst case, $O(n^2)$ comparisons are required, though this situation is rare.

The sorting tool was designed to break the sorting task into smaller subtasks, avoid fatigue, and obtain more precise evaluations by the human expert. To combine these smaller ordered subsets back into a unique ordered sequence, we applied a merging technique based on the classical k-way merge algorithm, which takes in k sorted lists and merges them into a single sorted list [35]. More specifically, the algorithm works as a tournament tree where, in each game, two of the input elements contend and the winner is promoted to the next level. The list is sorted in ascending order, so the winner of a game is the smallest of the two elements.

Regarding the computational cost, the algorithm takes on average $O(n \log(k))$, where n is the total number of items to sort, and k is the number of sorted lists which are taken as input. As a result of this process, our software tool generates a *csv* file which contains the list of models ordered from lowest to highest complexity according to the user's estimation.

To combine the evaluations of multiple users, we used a square count matrix C of size N, where N is the total number of elements to categorize, scale, or order. Each matrix element, $c_{i,j}$, represents the number of times that option i was preferred over option j. We assumed that each paired comparison is independent, i.e., the order of the comparisons is not relevant. We used the maximum a posteriori (MAP) scale estimation based on Thurstone's case V model presented by Tsukida & Gupta [55]. This technique allowed us to combine all the evaluations in a single sorted list, including a scale to quantify the items. In order to apply this technique, we used MATLAB and code functions provided by Tsukida & Gupta [55]. In addition, we developed a custom script to combine the results from the group of experts. The script processed a set of *csv* files,

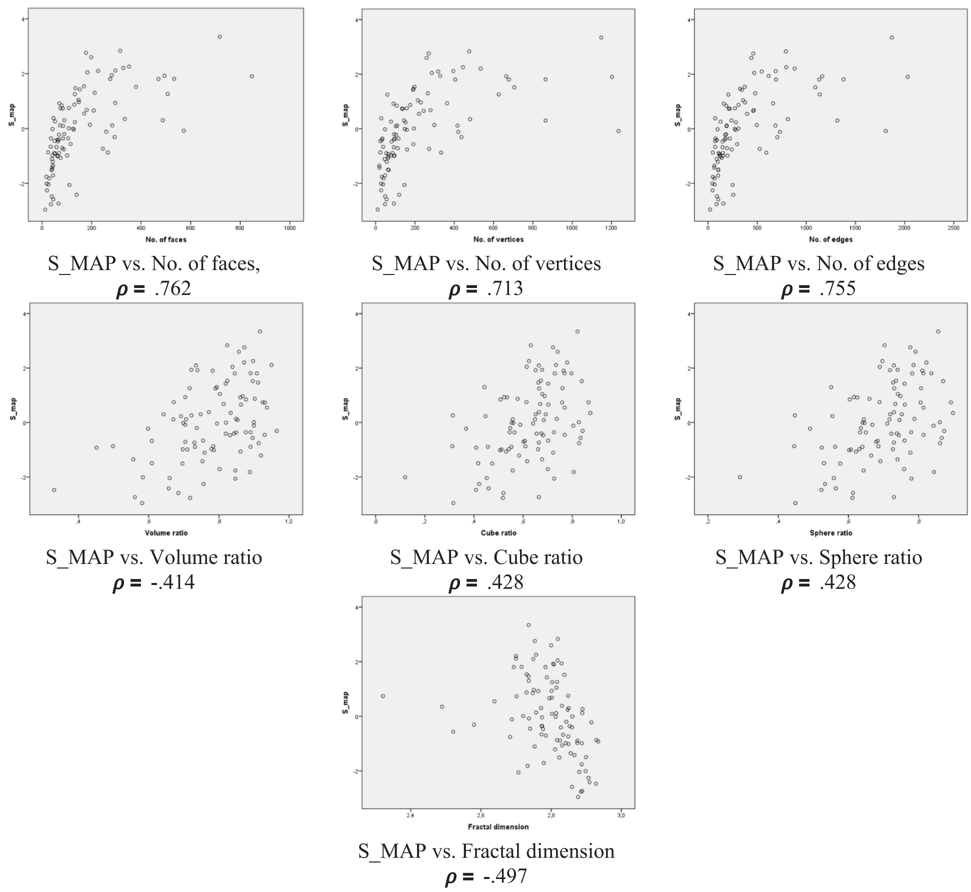

Fig. 7. Scatter plots for geometry-based metrics.

sorted the estimations from the experts, stored the files in a folder, and then built a matrix as input for the described method.

3.3. Implementation

Two software tools were developed in Microsoft Visual Basic.NET (VB.NET) to implement the sorting system. A custom script was also created using the online version of MATLAB R2021b. A block diagram of the system is shown in Fig. 4.

The first tool allows the experts to sort each subsample of the models using the QuickSort algorithm. An XML configuration file is used to indicate the specific subset of images that will be loaded into the system as a vector. This vector is then used as the input parameter for a traditional QuickSort recursive function, where the sorting decisions are determined by the user, instead of mathematically. When the sorting function ends, the results are displayed on screen as a table and exported as a Comma Separated Value (CSV) file with two columns, model id and order, as well as the total time employed by the user to complete the task.

The second tool allows the expert user to sort the models from the lists created in the previous step by applying the k-way merge algorithm. The tournament tree is built as a recursive function, where mathematical comparisons are again replaced by expert judgements. The function ends when there are no elements left to order. Final results are shown as a table and exported to a new CSV file with the same characteristics as the previous one.

Finally, the MATLAB script contains the code functions developed by Tsukida & Gupta [55] as well as other custom operations. These functions take a square matrix C as input, as described above. The custom code loads all the CSV files from the sorting tasks performed by each expert, which are then combined through several iterations and comparisons to create the matrix.

4. Results

Five CAD experts were asked to rank the models using the developed sorting tool. The tool and the instructions on how to complete the activity were shared with the experts via email. The experiment was structured in five tasks and participants were instructed to complete the assessments in order and in a timely manner. Tasks 1 to 4 correspond to the sorting tasks of 24, 24, 24, and 23 models, respectively, whereas task 5 was used to merge the results from the previous tasks. As a result, each expert produced a *csv* file with a list of 95 models sorted by their perceived complexity, the time the participant spent competing the sorting task, and the total number of comparisons. Experts spent, on average, 43.29 min (SD = 18.72) sorting the 95 models and made 566.6 comparisons (SD = 13.48). The maximum a posteriori scale of complexity (S_MAP) obtained by applying the procedure described in the previous point is shown in Fig. 5.

Next, we studied possible correlations between the complexity scale S_MAP and the complexity metrics discussed by Camba et al. [16]. Descriptive statistics are shown in Table 2.

Because of the non-normality of the data, a Spearman correlation test was applied to all data to measure the strength of association between the calculated scales and the complexity metrics as well as the direction of the relationship. The results are shown in Table 3.

To verify whether there is a monotonic relationship between the two variables between which we calculated Spearman's ρ, we present the scatter plots in Figs. 6-8 for each metric type with a p-value < 0.05.

The plots in Fig. 7 show that the strongest relationships correspond to metrics related to the number of geometric elements in the 3D model. However, the dispersion increases considerably when number of

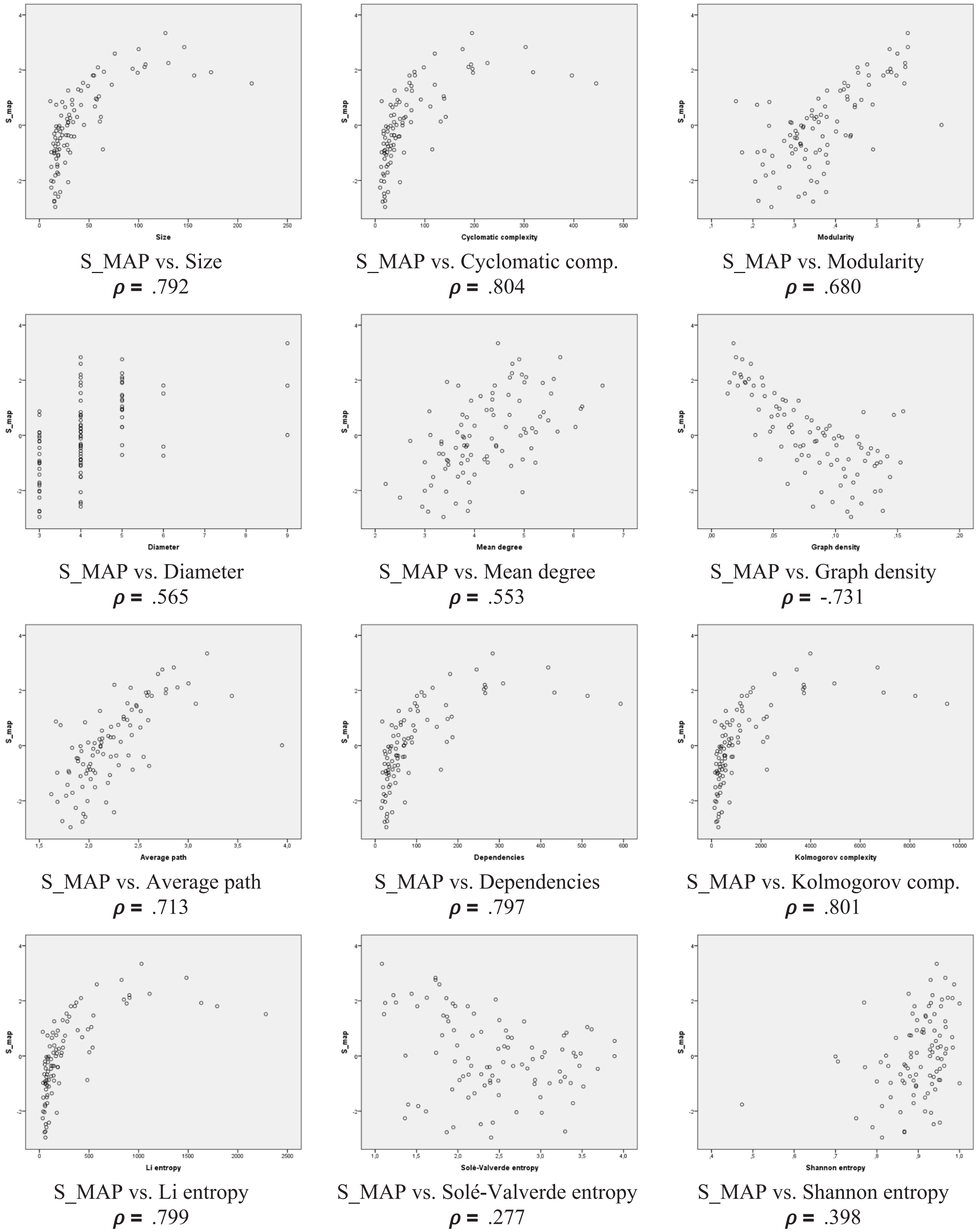

Fig. 8. Scatter plots for graph-based metrics map.

elements and S_MAP value increase. Ratio-based metrics show a great dispersion which is reflected in a lower ρ value.

5. Discussion

To interpret these results, it is important to remind that the experts were comparing the perceived complexity of the models defined as the

Table 4
Metrics sorted by absolute value of Spearman ρ.

Metric type	Metric	Spearman |ρ|
Graph	Cyclomatic complexity	0.804
Graph	Kolmogorov complexity	0.801
Graph	Li entropy	0.799
Graph	Dependencies	0.797
Graph	Size	0.792
Geometry	No. of faces	0.762
Computer file based	File size (MBytes)	0.758
Geometry	No. of edges	0.755
Graph	Graph density	0.731
Geometry	No. of vertices	0.713
Graph	Average path	0.713
Graph	Modularity	0.680
Graph	Diameter	0.565
Graph	Mean degree	0.553
Geometry	Fractal dimension	0.497
Geometry	Cube ratio	0.428
Geometry	Sphere ratio	0.428
Geometry	Volume ratio	0.414
Graph	Shannon entropy	0.398
Graph	Solé-Valverde entropy	0.277

perceived difficulty to create, use, and/or modify the CAD models. Unlike other studies on this subject, which focus solely on the geometric characteristics [4] or the complexity associated with the manufacturing processes of the parts [8,13], our work considers the complexity associated with the modeling process. It seems consistent that the highest Spearman ρ values, as presented on Table 3, are obtained from metrics associated with the parent/child relationship graph that represents the modeling process. Cyclomatic complexity was particularly strong, with $\rho = 0.804$, $p < .01$. Nevertheless, if we consider $|\rho| > 0.75$ as an indicator of a strong correlation between variables, as discussed by Akoglu [2] and Ratner [45], then we find a total of five metrics based on graphs, two based on geometry, and one based on a computer file metric that satisfy this condition, as shown in Table 4.

Our results reveal the suitability of graph-based metrics to adequately reflect the perception of CAD complexity by experienced CAD experts, who can anticipate the difficulty of creating or modifying a 3D CAD model by simply viewing an isometric depiction. To the best of our knowledge, our study is the first to use a large dataset of CAD models to analyze complexity metrics, and also to examine how graph-based complexity metrics compare to perceived complexity, as shown in Table 5. Only Johnson et al. [32] concluded that volume ratio was the

Table 5
Comparative analysis of previous studies that use complexity metrics in the context of CAD modeling.

Metric type	Metric	Valentan et al. [56]	Joshi & Ravi [33]	Johnson et al. [32]	Greco et al. [27]	Davis [24]	Bodein et. al. [11]	Camba et al. [18]	Johansson et al. [31]	Camba et al. [15]	Camba et al, [16]	Aramburu et al., [5]	This paper
	File size	x								x	x		x
Geometry	No. of faces			x	x		x			x	x	x	x
	No. of vertices										x	x	x
	No. of edges										x	x	x
	Vol./area ratio	x									x		x
	Volume ratio		x	x	x					x	x		x
	Cube ratio			x	x					x	x		x
	Sphere ratio		x	x	x					x	x		x
	Fractal dim.												x
Graph	Size					x		x	x		x	x	x
	Dimensionality					x			x		x		x
	Cyclomatic comp.					x					x	x	x
	Modularity										x		x
	Diameter										x		x
	Mean degree							x			x	x	x
	Graph density										x		x
	Average path										x		x
	Dependencies							x			x	x	x
	Kolmogorov comp.										x	x	x
	Li entropy								x				x
	Solé-Valverde ent.												x
	Shannon entropy												x
	Clustering										x		
	Eigenvector cent.										x		
	PageRank										x		

most highly correlated metric with experts' assessments, but their experiment used a small number of parts (10 models) and only students participated, not CAD experts.

Our results can also be explained by the fact that some metrics based on geometry are largely affected by patterning and symmetry operations, which may produce a large number of geometric elements (edges, faces, vertices) but have a very small impact on the difficulty to model a part. These situations are easily detected by CAD experts and probably contribute to an inferior performance of metrics that are exclusively based on geometric attributes.

We also observe similarities in the behavior of the parameters with higher ρ coefficients: a linear growth zone that contains most of the models and a plateau zone that represents around 10% of the models in our dataset (which are precisely the ones of greater complexity). These results may be explained by the difficulty of visually assessing parts with a very high degree of complexity. In future work, we plan to study this behavior further by checking whether more time should be allotted to decision making when evaluating these types of parts or by providing additional graphical information to better display the geometry.

We note that there is a need for studies that clearly identify the overlaps and gaps between existing metrics. Defining "similarity" or "complexity rate" in parametric CAD models is difficult, partly because it is a multidimensional construct. For example, two parametric models may be geometrically identical but very different internally because of the way they were built, which means they will react to changes in a very different manner. Likewise, a model that one may consider "complex" because of its geometric characteristics may be extremely simple in terms of its internal structure.

It is important to emphasize that this study focused on single-part modeling complexity, particularly parametric modeling complexity, where geometry is built as a sequence of interrelated featured controlled by parameters. The work could naturally be extended to study the complexity of assemblies, but other factors would come into play such as number of parts, length of the dependency chains, etc., which do not apply to part modeling. We speculate that the perceived complexity of an assembly could possibly depend, to a certain extent, on the complexity of the individual parts that make up the assembly. In any case, the treatment of assemblies is certainly multifaceted, and would require a separate rigorous study.

6. Conclusions and future work

In this paper, we described a novel methodology for producing a CAD complexity ground truth by integrating the evaluation provided by a group of experts who employed pairwise comparisons to evaluate the perceived complexity of parts. In addition, the preparation of a dataset of native parametric CAD models (SolidWorks, in our case) is also notable as it could be used in other studies in this field.

An additional contribution is the extensive statistical analysis which involves both graph-based and geometry-based complexity metrics of CAD models. Our results confirmed that the CAD complexity index strongly correlated with several metrics, particularly graph based. It is worth noting that two of the metrics with the strongest correlation with experts' perception, graph size and file size, are simple to compute in any commercial CAD system. Graph size measures the number of nodes in the graph, which represent the number of features used to model a part.

Our experiment demonstrates the viability of the proposed approach to measure perceived complexity. Future studies with more experts and an even larger dataset of parts would be relevant to confirm the results reported in this paper. We also acknowledge that although all the experts in our study have extensive experience in parametric modeling, it is possible that similar mindsets and workstyles could have affected the results. In this regard, the methodology and software tool are ready to be used in a more extensive experimental work.

An objective and quantitative measure of complexity is of great interest in studies that aim to assess the capability of a CAD model for reuse/modification. The results of this work could be applied to improve previous analyses that only used the mean degree metric [15] to determine the influence of complexity on CAD model editability. In our experience analyzing how formal modeling methodologies affect CAD model reusability, we have observed that role of methodology becomes increasingly relevant once the level of complexity of a part exceeds a certain threshold of complexity. The present work can be used to inform future strategies to determine more precisely where that threshold is. After a quick evaluation, engineers could then decide which modeling methodology is more appropriate for a specific design scenario.

The reusability of a CAD model depends on two factors that are highly coupled: the complexity of the part and the suitability of the modeling process followed to build the geometry. Since the S_MAP variable has a strong relationship both with geometric and graph-based complexity parameters, it would be interesting to analyze as future work how the variable correlates with reusability metrics such as the ones used in previous studies [15], or to incorporate other complexity parameters that have been proposed in the context of additive manufacturing [34] to determine whether they have a stronger relation with the S_MAP variable (experts' opinion) than with the parameters investigated in this work.

With regards to academia and workforce development, our work has direct application to CAD instruction where proper gauging of the level of difficulty of the exercises is critical, particularly in the assessment and evaluation of modeling skills. We envision our approach as part of an intelligent tutoring system that can offer a personalized learning experience where the system provides modeling exercises of increasing complexity based on the student's performance level.

As future work, we plan to replicate our study with students to determine whether their evaluations align with those provided by CAD experts. Additionally, we are interested in using complexity assessment tasks to evaluate student expertise. We speculate that high performing students will provide assessments with higher levels of agreement with CAD experts' assessment than low performing students.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by a grant provided by Vice-rectorate for Research from Universitat Politècnica de València (PAID-11-21) and the Spanish Ministry of Universities in the framework of the State Programme for the Promotion of Talent and its Employability in R&D&I, State Mobility Sub-programme of the State Plan for Scientific and Technical Research and Innovation 2017-2020 (Grant PRX21-00387).

References

[1] D CentralContent – Free D CAD Models, 2D Drawings, and Supplier Catalogs, 2020. <https://www.dcontentcentral.com>.
[2] H. Akoglu, User's guide to correlation coefficients, Turkish J. Emerg. Med. 18 (3) (2018) 91–93.
[3] K. Amadori, M. Tarkian, J. Ölvander, P. Krus, Flexible and robust CAD models for design automation, Adv. Eng. Inf. 26 (2) (2012) 180–195.
[4] M.F. Arslan, A. Haridis, P.L. Rosin, S. Tari, C. Brassey, J.D. Gardiner, M. Genctav, SHREC'21: quantifying shape complexity, Comput. Graph. 102 (2022) 144–153.
[5] A. Aranburu, J. Cotillas, D. Justel, M. Contero, J.D. Camba, How does the modeling strategy influence design optimization and the automatic generation of parametric geometry variations? Comput. Aid. Des. 151 (2022), 103364.

[6] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 3(1), 2009, pp. 361–362.

[7] D. Baxter, J. Gao, K. Case, J. Harding, B. Young, S. Cochrane, S. Dani, An engineering design knowledge reuse methodology using process modeling, Res. Eng. Des. 18 (1) (2007) 37–48.

[8] S. Ben Amor, A. Tahan, B. Louhichi, The impact of geometric complexity on printing time and cost for additive manufacturing (AM) process, in: International Conference on Advanced Materials Mechanics & Manufacturing, Springer, Cham, 2021, pp. 203–210.

[9] A. Benaouali, S. Kachel, An automated CAD/CAE integration system for the parametric design of aircraft wing structures, J. Theor. Appl. Mech. 55 (2) (2017) 447–459.

[10] Y. Bodein, B. Rose, E. Caillaud, A roadmap for parametric CAD efficiency in the automotive industry, Comput. Aid. Des. 45 (10) (2013) 1198–1214, https://doi.org/10.1016/j.cad.2013.05.006.

[11] Y. Bodein, B. Rose, E. Caillaud, Explicit reference modeling methodology in parametric CAD system, Comput. Ind. 65 (1) (2014) 136–147, https://doi.org/10.1016/j.compind.2013.08.004.

[12] P. Bourke, Box counting fractal dimension of volumetric data, 2014. <http://paulbourke.net/fractals/cubecount/>.

[13] H.D. Budinoff, M. Shafae, Connecting part geometry and cost for metal powder bed fusion, Int. J. Adv. Manuf. Technol. (2022), https://doi.org/10.1007/s00170-022-09688-x.

[14] X.T. Cai, S. Wang, X. Lu, W.D. Li, Y.W. Liang, Parametric and adaptive encryption of feature–based computer–aided design models for cloud–based collaboration, Integr. Comput.-Aid. Eng. 24 (2) (2017) 129–142.

[15] J.D. Camba, M. Contero, P. Company, D. Pérez–López, J. Otey, Identifying high–value CAD models: an exploratory study on dimensional variability as complexity indicator, in: ASME 2018 13th International Manufacturing Science and Engineering Conference, MSEC 2018, 2018. https://doi.org/10.1115/MSEC2018-6391.

[16] J.D. Camba, M. Contero, D. Pérez–López, P. Company, A database framework for the characterization and classification of parametric models based on complexity metrics to support data analytics, in: ASME 2019 14th International Manufacturing Science and Engineering Conference, MSEC 2019, 2019. https://doi.org/10.1115/MSEC2019-2812.

[17] J.D. Camba, M. Contero, P. Company, N. Hartman, The cost of change in parametric modeling: a roadmap, Comput.-Aid. Des. Appl. 18 (3) (2021) 634–643, https://doi.org/10.14733/cadaps.2021.634-643.

[18] J.D. Camba, M. Contero, Improved representation of dependencies in feature–based parametric cad models using acyclic digraphs, in: 10th International Conference on Computer Graphics Theory and Applications, GRAPP 2015, Berlin, Germany, 2015, pp. 16–25.

[19] L. Casti, Complexification: Explaining a paradoxical world through the science of surprise, Harper Collins Publishers, New York, 1995.

[20] S. Chase, P. Murty, Evaluating the complexity of CAD models as a measure for student assessment. In eternity, infinity and virtuality in architecture: Proceedings of ACADIA 2000, vol. 1, 2000, pp 173–182.

[21] P. Company, M.C. González–Lluch, CAD 3D con SolidWorks Tomo I: Diseño básico, second ed. Publicacions de la Universitat Jaume I, 2021. https://doi.org/10.6035/Sapientia176.

[22] P. Company, J. Dorribo Camba, CAD 3D con SolidWorks® Tomo II: Diseño avanzado, Publicacion. Univ. Jaume I (2018), https://doi.org/10.6035/sapientia145.

[23] M. Contero, P. Company, C. Vila, N. Aleixos, Product data quality and collaborative engineering, IEEE Comput. Graph. Appl. 22 (3) (2002) 32–42, https://doi.org/10.1109/MCG.2002.999786.

[24] D. Davis, Quantitatively analysing parametric models, Int. J. Archit. Comput. (2014), https://doi.org/10.1260/1478-0771.12.3.307.

[25] B. Denkena, J.T. Schürmeyer, R. Kaddour, V. Bö, Assessing mould costs analyzing manufacturing processes of cavities, Int. J. Adv. Manuf. Technol. 56 (2011) 943–949.

[26] GrabCAD: Design Community, CAD Library, 3D Printing Software, 2018. <https://grabcad.com/>.

[27] A. Greco, G. Piccicacchi, P. Manco, R. De Franchi, M. Ambrico, F. Caputo, S. Gerbino, Investigation on geometrical complexity techniques for assessing AM feasibility, in: Macromolecular Symposia, vol. 396(1), 2021, pp. 2000309.

[28] S. Harries, C. Abt, M. Brenner, Upfront CAD—Parametric modeling techniques for shape optimization, in: Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences, Springer, Cham, 2019, pp. 191–211.

[29] C.A.R. Hoare, Algorithm 64: quicksort, Commun. ACM 4 (7) (1961) 321, https://doi.org/10.1145/366622.366644.

[30] H.C. Huang, S.M. Lo, G.S. Zhi, R.K. Yuen, Graph theory–based approach for automatic recognition of CAD data, Eng. Appl. Artif. Intel. 21 (7) (2008) 1073–1079.

[31] J. Johansson, Analysing engineering knowledge in CAD–models and spread sheets using graph theory and filtering, in: 24th ISPE International Conference on Transdisciplinary Engineering. Singapore July 10–14, IOS Press, 2017, pp. 629–638.

[32] M.D. Johnson, L.M. Valverde, W.D. Thomison, An investigation and evaluation of computer–aided design model complexity metrics, Comput.-Aided Des. Applic. 15 (1) (2018) 61–75.

[33] D. Joshi, B. Ravi, Quantifying the shape complexity of cast parts, Comput.-Aided Des. Applic. 7 (5) (2010) 685–700.

[34] J. Jayapal, S. Kumaraguru, S. Varadarajan, A view similarity-based shape complexity metric to guide part selection for additive manufacturing, Rapid Prototyp. J. (2022), https://doi.org/10.1108/RPJ-04-2022-0122.

[35] D.E. Knuth, Art of computer programming, volume 2: Seminumerical algorithms, Addison–Wesley Professional, 2014.

[36] S. Kwon, B.C. Kim, D. Mun, S. Han, Graph–based simplification of feature–based three–dimensional computer–aided design models for preserving connectivity, J. Comput. Inf. Sci. Eng. 15 (3) (2015), 031010.

[37] J. Li, B.H. Wang, W.X. Wang, T. Zhou, Network entropy based on topology configuration and its computation to random networks, Chin. Phys. Lett. 25 (11) (2008) 4177.

[38] W. Li, D. Schuurmans, Modular community detection in networks, IJCAI Int. Joint Conf. Artif. Intell. (2011), https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-231.

[39] J. Liu, Y. Ma, A survey of manufacturing oriented topology optimization methods, Adv. Eng. Softw. 100 (2016) 161–175, https://doi.org/10.1016/j.advengsoft.2016.07.017.

[40] J. Lubell, K. Chen, J. Horst, S. Frechette, P. Huang, Model based enterprise / technical data package summit report (NIST Technical Note 1753), NIST Technical Note 1753 (2012), https://doi.org/10.6028/NIST.TN.1753.

[41] A. Mowshowitz, M. Dehmer, Entropy and the complexity of graphs revisited, Entropy (2012), https://doi.org/10.3390/e14030559.

[42] J.E. Owensby, J.D. Summers, Assembly time estimation: assembly mate based structural complexity metric predictive modeling, J. Comput. Inf. Sci. Eng. 14 (1) (2014), 011004.

[43] S. Patalano, F. Vitolo, A. Lanzotti, A graph based approach to CAD modeling: a digital pattern application to the sizing and modeling of manual transverse gearboxes, in: Proc. of International Conference on Graphic Engineering, Madrid, Spain, 2013.

[44] J.F. Petiot, B. Yannou, Measuring consumer perceptions for a better comprehension, specification and assessment of product semantics, Int. J. Ind. Ergon. 33 (6) (2004) 507–525, https://doi.org/10.1016/j.ergon.2003.12.004.

[45] B. Ratner, The correlation coefficient: Its values range between +1/−1, or do they? J. Target. Meas. Anal. Mark. 17 (2009) 139–142, https://doi.org/10.1057/jt.2009.5.

[46] J. Rossignac, Shape complexity, Vis. Comput. 21 (12) (2005) 985–996.

[47] W. Saleem, A. Belyaev, D. Wang, H.P. Seidel, On visual complexity of 3D shapes, Comput. Graph. 35 (3) (2011) 580–585.

[48] V. Salehi, C. McMahon, Methodological integration of parametric associative CAD systems in Product Lifecycle Management (PLM) environment, in: Proceedings ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2009, pp. 505–514.

[49] S.L. Schlindwein, R. Ison, Human knowing and perceived complexity: implications for systems practice, Emerg.: Complex. Org. 6 (3) (2004) 27–32.

[50] C.E. Shannon, A mathematical theory of communication, ACM SIGMOBILE Mob. Comput. Commun. Rev. 5 (1) (2001) 3–55.

[51] M. Shi, S. Zhang, A method of 3D CAD model retrieval based on feature adjacent graph, in: Proceedings of The International Conference on Graphic and Image Processing (ICGIP 2012), 2012, 87681, Paper ID 87681L.

[52] R.V. Solé, S. Valverde, Information theory of complex networks: on evolution and architectural constraints, in: Complex Networks, Springer, Berlin, Heidelberg, 2004, pp. 189–207.

[53] SolidWorks Exercises, 2018. <https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fwww.solidworks.com%2Fsupport%2Ftraining%2Ftraining-files&data=05%7C01%7Ch.gopu%40elsevier.com%7C75dc28f1dc584348880c08db430bce9e%7C9274ee3f94254109a27f9fb15c10675d%7C0%7C0%7C638177490217937890%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=H22B0XpVLfk1%2BsplbC1JO0odBa6b9Dg1IRyDaE%2F1fk0%3D&reserved=0>.

[54] SolidWorks Training Files, 2018. <https://nam11.safelinks.protection.outlook.com/?url=https%3A%2F%2Fgwpg.spec.org%2Fbenchmarks%2Fbenchmark%2Fspecapc-solidworks-2020&data=05%7C01%7Ch.gopu%40elsevier.com%7C75dc28f1dc584348880c08db430bce9e%7C9274ee3f94254109a27f9fb15c10675d%7C0%7C0%7C638177490217937890%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=huX6OgqzGm22XOIRalIXLcRLwoHweEYfuxCQGDPPbm0%3D&reserved=0>.

[55] SPECapc for Solidworks 2020 release, 2020. <https://www.spec.org/gwpg/apc.static/sw2020info.html>.

[56] K. Tsukida, M.R. Gupta, How to analyze paired comparison data, University of Washington Technical Report (2011). UWEETR-2011-0004.

[57] B. Valentan, T. Brajlih, I. Drstvenšek, J. Balič, Development of a part–complexity evaluation model for application in additive fabrication technologies, Strojniski Vestnik/J. Mech. Eng. 57 (10) (2011) 709–718.

[58] Y.H. You, X.Y. Kou, S.T. Tan, Adaptive tetrahedral mesh generation of 3D heterogeneous objects, Comput.-Aid. Des. Appl. 12 (5) (2015) 580–588.