



GRAPH-BASED SIMILARITY ANALYSIS OF BOM DATA TO IDENTIFY UNNECESSARY INNER PRODUCT VARIANCE

Schmidt, Michael (1); Gehring, Benedikt (1); Gerber, Jan-Sebastian (1); Stocker, Johannes Michael (1); Kreimeyer, Matthias (2); Lienkamp, Markus (1)

1: Technical University of Munich, Germany; 2: MAN Truck & Bus AG, Germany

Abstract

This paper contributes to the fields of variant management and product family design. The focus lies on analysing historically grown product portfolios in order to reduce unnecessary inner variety. Such inner variety adds no value to the customer, yet it induces complexity costs within the whole company. Increasing transparency in documented product variants is key when applying standardisation or modularisation methods as part of variant management. Studies of literature and industrial practice at a major German truck manufacturer show that analysing product structure information from BOM data yields the potential to point out promising candidates in companies' portfolios for effective standardization or modularisation. For modelling and analysing highly variant and complex product structures, we employ graph-based modelling of BOM data in combination with a state-of-the-art tree matching algorithm for similarity calculations. Actual product data of a truck manufacturer serves as a case study. Thereby, we propose a generally applicable approach that enables intuitive handling of large amounts of product family data and that effectively supports variety reduction efforts.

Keywords: Product families, Product structuring, Complexity, Data analysis, Data visualization

Contact:

Michael Schmidt
Technical University of Munich
Chair of Automotive Technology
Germany
m.schmidt@ftm.mw.tum.de

Please cite this paper as:

Surnames, Initials: *Title of paper*. In: Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 1: Resource-Sensitive Design | Design Research Applications and Case Studies, Vancouver, Canada, 21.-25.08.2017.

1 INTRODUCTION

1.1 Motivation

Commercial vehicles have to fulfil business customers' needs in all varieties, ranging from light distribution to severe duty mining services. Each application demands very specific configuration options of the vehicle type and of almost every component in the vehicle. In addition to this application-related variety, globally operating manufacturers of commercial vehicles are prone to a legislation-induced increase of product variety. At the same time, life cycles of commercial vehicles typically last 20-30 years, resulting in many incremental changes to product generations. These external factors for variety lead to historically grown portfolios of technical solutions. Internal factors for variety extend this problem. A major internal factor is the lack of a clearly focused variant management, which includes linking market demands to available technical solutions (Braun et al., 2013). Another major internal factor for variety is the lack of transparency in documentation of technical solutions across the portfolio. Especially the inner product variety, which is invisible to the customer, leads to unnecessary variants cluttering engineering processes and inducing complexity costs in the whole company. Thus, further support in terms of mastering the documented variety of technical solutions is required - in particular for manufacturers of highly variant complex products such as commercial vehicles.

1.2 Challenges and Goals

This paper aims at supporting standardization and modularization efforts for companies with large existing product families of complex products. An important boundary condition of the industrial context is the fact that in most cases resources for redesign of products are limited. Therefore, standardization or modularization efforts need to be segmented and prioritized. Consequently, it is necessary to identify a) unnecessary variety and b) the most promising candidates thereof for standardization or modularization efforts. To provide the expected support, we pursue a data-driven approach, which allows for analysis of documented product component variants across the portfolio. Analysing a complete product family requires consistent data that is comparable across product subsystems. In addition, the data needs to be available for all components. Although more and more companies apply e.g. functional modelling to represent product architectures, the most commonly available data domain is product structure information in the form of bills of materials (BOMs). We aim at developing an analysis toolbox that is generally applicable by modelling these hierarchical structures including components, assemblies and parts. With the modelled and analysed structures, we lay the foundation for identification of similar variants that are potentially unnecessary and thereby promising candidates for standardization within existing and highly variant product families.

1.3 Structure of the paper and research methodology

This paper is structured as follows. Section 2 reflects the state of the art in the fields of variant management and product structure information both in research and at MAN Truck & Bus (MAN), a global truck manufacturer. Existing approaches to analyse product structure information and to quantify similarities in product families are presented and an intermediate conclusion is drawn. Section 3 presents the graph-based approach to identify standardization candidates in existing product families. In section 4, the approach is applied to a selection of product components at MAN, namely driven rear axle variants as an initial validation. Section 5 gives a conclusion and outlook on ongoing and future research.

We conducted our research within a methods group for modular systems with three researchers and two product architecture- and variant management experts at MAN. Initial studies included discussions with experts, reviews of literature and of product structure information at the company. The developed approach was implemented in Soley Studio, and the initial case study was performed with product structure data of well-known components at MAN to allow for direct validation of generated results.

2 STATE OF THE ART

2.1 Variant management approaches and their implementation at MAN

Variant management, a discipline of complexity management, emerged with the continuously growing demand for diversity in individual specifications of mass produced products, or mass customization. For

companies in markets characterized by mass customization, it is crucial to manage the variety of products and product components to remain profitable and to optimize the portfolio for maximum profit (see (a) in Figure 1). The main target herein is to balance customer-perceived external variety of the portfolio against internal variety of documented product components, processes etc. (Tseng et al., 1996). Wildemann (2016) classifies three strategies of variant management: (1) reduction, (2) handling, and (3) prevention of (unnecessary) product variety. In the domain of engineering management and -design, this translates into approaches from the field of product family design such as standardization, modularization or the design of modular kits comprising platform-, modular- and product specific elements (Figure 1 (b)) (Jiao et al., 2007; Kreimeyer et al., 2014).

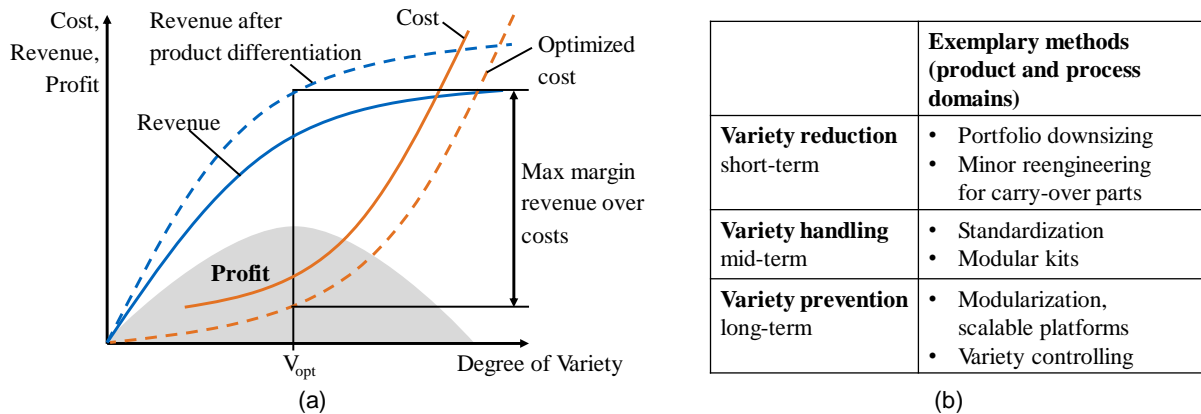


Figure 1. (a) Cost-, revenue- and profit effects of product variety (after Schuh, 2005), (b) variant management strategies and exemplary methods (after Wildemann, 2016)

Existing methods for product family design usually follow the logic of product architecture processes, which addresses the definition of product functions and their embodiment in physical components based on customer requirements (Ulrich, 1995; Yu et al., 1999). At MAN, one major embodiment of variant management lies in the early phase of development processes. The so-called process of component variant planning focuses on components of the full vehicle (Schumacher et al., 2015). Market requirements are translated into variant driving product characteristics, which are configurable by the customer (external variety). These customer requirements together with technical restrictions are multiplied into the required number of variants for each component (internal variety). This variant planning process creates an overview of internal variants per component in early development stages. It follows the logic of product architecture methodologies and processes found in literature. However, the domain of product functions is not explicitly modelled at full vehicle level. One of the main drivers for the lack of generalized functional modelling at MAN is the effort to find common levels of abstraction and terminology across engineering disciplines. Methods for reengineering towards modularization at full vehicle level are not implemented at MAN. Such methods are only employed at component level. This differentiation into two levels, i.e. full vehicle and component, is also pursued in MAN's modular kit system and in the way of documenting product structure information.

2.2 Product structure information at MAN

Other than functional product models, product structure information is readily available for all components at MAN, as it is in the vast majority of manufacturing companies. There are modular kits defined at two levels. At full vehicle level, cross-product configuration of main components is realized in a modular kit. This means that all component variants resulting from the component variant planning process are attributed with configuration rules. It is determined, which component variant can be built into which vehicle model. The second level of modular kits is the component level. All components, e.g. axles, cabin or frame, and their variants are designed as modular kits for themselves, comprising sub-components, assemblies and parts (Kreimeyer et al., 2014). For each component variant, a bill of materials (BOM) is documented with all subassemblies and parts. The BOM-type MAN employs are the multi-level modular BOM. It allows for reduced documentation effort of component variants that are configured from a modular kit (Jiao et al., 2000). A subassembly is referred to in each component variant BOM where it is reused via its ID. As soon as a single part changes within one (sub)assembly,

the assembly gets a new ID. The taxonomy for defining part IDs furthermore allows to classify parts into e.g. raw materials or standardized parts such as bolts and nuts.

2.3 Modelling and analysing product families and -structures

Analysing product families for variant management purposes requires **modelling the product family** at an appropriate level of abstraction. Existing modelling approaches from product family design focus on the domains of product architecture, i.e. product functions and -structure, and include perspectives from further downstream the product lifecycle, e.g. production (Pirmoradi et al., 2014). Due to limited space, only the basic modelling approaches of product architecture domains are reviewed here. They can be clustered into three basic types of modelling: (1) functional and flow diagrams with graphical representation, (2) matrix-based modelling, and (3) graph-based modelling. As a form of type (1), Harlou (2006) presents a method to visualise the variety of an existing product family. Functional relations and product structures are modelled, indicating the usage of functional and physical elements across the product portfolio. Recurring patterns in the product family architecture can be defined as standard designs, ready for reuse across several or all products of the product family. Matrix-based approaches based on e.g. design structure matrices (DSMs) are applied to model lateral relations of a system, whereas BOM data contains mostly vertical, i.e. hierarchical relations (Eppinger and Browning, 2012). There are approaches that apply DSMs to model more than one hierarchical level of a product (Tilstra et al., 2012) or use BOM data to identify platform elements in an existing product family (Steva et al., 2007). They are, however, either enhanced by functional relations between the components or they become unsuitable for highly variant families of complex products (Kreimeyer and Lindemann, 2011). Graph-based modelling is a more promising approach for large data sets of complex products (Lindemann, 2016). Especially multi graphs allow for intuitive modelling of vertical and lateral relations in products and components. Multi graphs are merged graphs of multiple domains, thus featuring several classes of nodes (e.g. products, assemblies, parts) and edges (e.g. vertical, lateral relations). Both nodes and edges can be attributed with information such as cost or quantity, which allows for algorithm-based calculation of indices, pattern recognition or graph transformation (Heckel, 2006).

Based on modelled product structures, **analyses** can be conducted in a product family. An important set of information for reduction of internal variety of an existing product portfolio is a comprehensive assessment of parts commonality and similarity of variants. There are several commonality indices (CI) that can be calculated from BOM- and additional data focusing on different characteristics. Table 1 gives an overview and remarks on applicability as well as limitations for the most relevant indices. We selected indices mainly requiring hierarchical product structure data as provided by BOMs. When modelling BOMs as hierarchical graphs, also denoted as trees, the All-Path-Tree-Edit-Distance-Algorithm (APTED) by Pawlik and Augsten (2016) can be considered. It calculates the tree edit distance, which is the minimum-cost sequence of node edit operations that transforms one tree into another. The output of the APTED-Algorithm can be converted into a similarity index (noted as APTEDI) of two trees as shown in Section 3.

Table 1. Comparison of Commonality Indices and APTED Index

Index	Remarks	Limitations
Degree of Commonality Index (DCI) (Collier, 1981)	<ul style="list-style-type: none"> • Most traditional commonality index • Reflects the average number of common parent items per distinct component part • Easy to collect data 	<ul style="list-style-type: none"> • Difficult to interpret due to varying upper limit for different families (no fixed boundaries of index)
Component Part Commonality Index (CIC) (Jiao and Tseng, 2000)	<ul style="list-style-type: none"> • Based on the DCI • Takes into account production volume, quantity per operation, and the cost of component part • Difficult to collect data 	
Total Constant Commonality Index (TCCI) (Wacker and Treleven, 1986)	<ul style="list-style-type: none"> • Normalized version of DCI • Fixed boundaries: $0 < TCCI < 1$ • Easy to collect data 	<ul style="list-style-type: none"> • Focusing only on the percentage of common/unique components rather than cost factors
All Path Tree Edit Distance Index (APTEDI) (based on Pawlik and Augsten, 2016)	<ul style="list-style-type: none"> • Considers included components plus the assembly structure within the product • Possibility to adjust transformation costs • Fixed boundaries: $0 < APTEDI < 1$ • Easy to collect data 	<ul style="list-style-type: none"> • Poor runtime for large BOMs • No data about the applicability of the APTED-Algorithm for BOMs

2.4 Intermediate conclusion and research gap

To reduce unnecessary inner variety in historically grown product families it is required to gain transparency of documented variants and the differences between them. Analysis results can serve as input for standardization or modularization efforts as part of variant management in engineering design. Predominant methods mainly focus on functional modelling or a combination of functional and structural modelling of product families. In manufacturing companies however, product structure data in the form of BOMs still is the best-maintained data type available for the complete portfolio of products and components. Existing methods are not capable of utilizing this structure information on a large scale, i.e. for analysis of highly variant product families. A data-driven approach is required that is generally applicable in industrial practice. Therefore, we present an approach that models and analyses all hierarchy levels of product structures based on BOM data, aiming to fulfil the need of identifying promising candidates for product or component standardization in a complex portfolio.

3 APPROACH

As part of our approach we introduce two methods for product structure analysis. The goal is to increase transparency regarding the existing inner product variety in a historically grown product portfolio and to facilitate the derivation of appropriate measures to reduce it. Possible measures include the adaption of the degree of product standardization as well as the use of common parts or carry over parts. As matrix-based modelling or flow-diagrams are unsuitable for modelling hierarchical relations of highly variant, complex products such as commercial vehicles, we employ graph-based modelling in combination with a state-of-the-art tree matching algorithm for similarity calculations. This allows us to handle large amounts of product structure data. Both proposed methods, the "Assembly Analysis" and the "Area analysis", allow the identification of parts and assemblies that are potential standardization candidates. The starting point of this graph-based data analytics approach is a set of multi-levels BOM data, which is modelled in a graph as tree structures. The introduced methods analyse the similarity between multiple tree structures. We implemented both methods with the software Soley Studio.

3.1 Method 1: Assembly analysis

The assembly analysis consists of four steps which are shown in Figure 2. The steps and results are explained below.

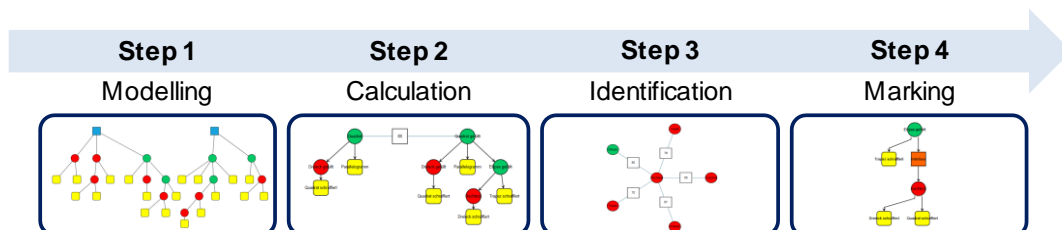


Figure 2. Steps of the Assembly analysis

Step 1: Modelling

The first step starts with the acquisition of the multi-level BOM data. As the assembly analysis is highly automated, it does not matter if the data set represents variants of products, components or subassemblies. In the following, we assume that the BOM data is available and well maintained. In order to transform the BOM data into a tree-like graph in Soley Studio, a meta model is required. In the meta model, the classes and attributes of all nodes and edges are defined. The examples given throughout this paper show the analysis of component variants (root nodes) consisting of multiple subassemblies and parts (branches and leaves). After defining the meta model, the data is imported to Soley Studio and visualised as a graph.

Step 2: Calculation

In the second step, the similarity between each pair of the imported assemblies is analysed based on their subassemblies and parts. In the following, we present two algorithms for the similarity calculation of unconnected tree structures. Each BOM tree ("A" and "B") represents a component variant.

Common Parts Algorithm (CPA)

The CPA compares two trees based on the nodes that constitute the lowest hierarchy level (leaf nodes). The nodes and edges of higher hierarchy levels are not considered. The part nodes of the two assemblies are compared based on their "BOM ID" attribute, which has been defined in the meta model. Hence, the return value of the CPA is a measure for the content similarity of two assemblies without taking into account their structural patterns.

$$CPA \text{ Return Value} = \frac{2 * \text{Number of Common Parts}}{\text{Sum of all Parts}} \in [0,1] \quad (1)$$

As with the TCCI, the CPA return value ranges from 0 to 1. A return value of 1 indicates that the two assemblies consist of the exact same parts, whereas a return value of 0 signals that there is no common part at all. This comparison is made once for each pair of assemblies automatically. In Figure 3 (a), the CPA is applied to compare the two BOM tree structures "A" and "B". Having 6 common parts (yellow) and 14 parts in total, the CPA returns 0.86 or 86% as (content) similarity value. Parts that do not exist in both trees are highlighted in red. The greyed-out assembly nodes are not compared by the algorithm.

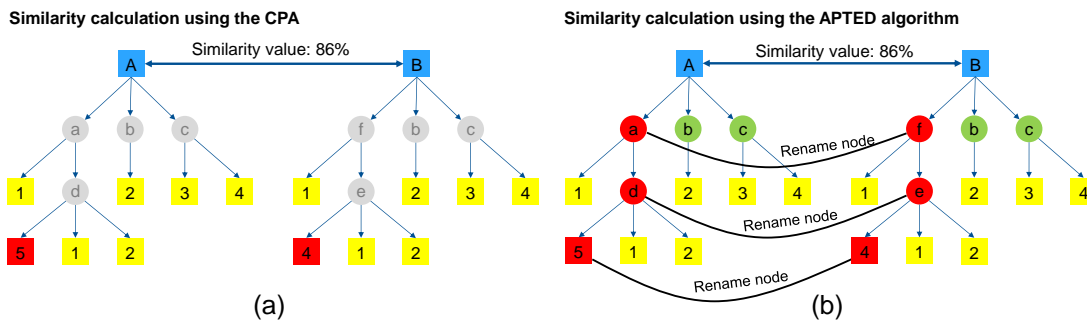


Figure 3. Similarity calculation using (a) the CPA and (b) the APTED algorithm

APTED Algorithm

The All-Path-Tree-Edit-Algorithm (APTED) by Pawlik and Augsten (2016) allows for a more precise similarity analysis of tree-like graph objects. In addition to the content similarity, the APTED algorithm also takes into account the structural similarity of all hierarchy levels. Consequently, the APTED algorithm unfolds its benefit of higher accuracy compared to the CPA especially in complex multi-level tree structures. The idea behind APTED is to identify the minimum-cost sequence of transformation operations when transforming a tree structure A into a tree structure B. Therefore, the user defines cost rates for the standard transformation operations "delete node", "insert node" and "rename node". For each operation, the respective single transformation cost (STC) incurs. The sum of all STCs gives the total transformation cost (TC). APTED is able to find the minimum-cost transformation operation sequence among the theoretically infinite number of sequences. In order to create comparability to the CPA, a return value between 0 and 1 is required. Thus, the incurring costs are standardized at the maximum possible transformation cost (MAX) depending on the user-defined cost rates. If the highest cost rate is assigned to the operation "insert a node", the MAX defines the case in which all nodes of each tree have to be inserted.

$$APTED \text{ Return Value} = \frac{2 * MAX - TC}{2 * MAX} \in [0,1] \text{ with } TC = \min(\sum STC) \in [0, \infty) \quad (2)$$

Figure 3 (b) shows how the algorithm proceeds when comparing two component variant trees "A" and "B". All three cost rates shall be set to 1. In order to transform tree "A" into "B", the minimum cost incurs by renaming the nodes "a", "d" and "5" to "f", "e" and "4", respectively. As the structure of both trees is equal, no nodes have to be inserted or deleted. The three renaming operations cause a TC of 3. The MAX represents the cost if all eleven nodes of both trees have to be inserted from scratch. Thus, we receive a return value of 0.86 or 86 %.

Step 3: Identification

In order to make the calculation results applicable and to ease their interpretability, we filter the data graph in step three in order to find the (theoretically) most promising standardization candidates. The filtering criteria can be chosen on a case-specific basis as long as they are based on classes or attributes

defined in the meta model. As we are looking for promising standardization candidates among the assemblies, we need to identify assembly clusters of high similarity values holding a high savings potential. Thus, using the Pareto principle, our first filtering criterion is a minimum assembly similarity value of 80 %. In this way, we are able to focus on the remaining 20 % of assemblies which have the highest similarity values and therefore -at least theoretically- hold a high potential to be standardized. The second filtering criterion aims at the identification of the assemblies having the highest cost savings potential in case of standardization. Thus, the cost of the assemblies' parts would be an appropriate filter criterion. Whenever part cost information is not included, we propose to use the assembly size as a second criterion, i.e. the number of parts included in the assembly. This criterion is based on the hypothesis that the savings potential in case of the standardization of assemblies increases with their number of parts. To focus on the assemblies with the highest savings potential, we filter the top 20 % of the "biggest" assemblies. After filtering the data graph on the basis of the two above mentioned criteria, we receive multiple assembly clusters. These clusters only contain assemblies, whose similarity values are at least 0.1 and who are part of the top 20 % of the biggest assemblies, in terms of parts contained.

Step 4: Marking

The fourth step of the method aims at creating an intuitively interpretable graph visualization of the analysis results. In order to decide to which degree two or more assemblies can be standardized, we need to break down the similar assemblies to the level of parts. Hence, we reintroduce the filtered assemblies into their original BOM graph structure. To each assembly cluster from step three a marker node with a unique ID is assigned, which highlights their position in the product structure. The highly similar assemblies thus are easy to retrieve in the graph visualization of the component variants. In addition, the origin and structure of the filtered similar assemblies, as well as their position in the product structure are visible. Based on this graph visualization, experts can effectively assess whether (partial) standardization of the identified candidates is feasible in practice.

3.2 Method 2: Area analysis

The idea of the area analysis is to expand the similarity calculation and visualization beyond the "assembly-by-assembly" comparison approach. Now, we investigate whether the surrounding areas of similar assemblies are also promising targets of standardization measures - or in other words, whether the scope of standardisation should be expanded. We define an area as the network of two or more related assemblies within the same component variant, which are reachable via a user-defined amount of edges (path length). The area analysis starts from the previously identified assemblies that exceed the user-defined minimum similarity value. Compared to the assembly analysis, the area analysis only differs in step 2 "calculation". Initially, the similarity values between all assemblies from different component variants are calculated analogous to the assembly analysis using either the CPA or APTED algorithm. In this way, the workflow identifies all assemblies having a similarity value equal or bigger than a user-defined minimum (i.e. 90 %). If a pair of assemblies "a" and "b" exceeds this similarity limit, our analysis workflow identifies the areas around "a" and "b" according to the user-defined path length. All assemblies belonging to the same area are connected to their origin ("a" or "b") by edges and are visualized in the same colour. Thereupon, the analysis workflow calculates the similarity value for each pair of assemblies belonging to different areas. If the similarity value exceeds a user-defined minimum (i.e. 80 %), the workflow builds an inter-areal edge showing the calculation result between these assemblies. The resulting graph visualizes the areas and all similarity values. In addition, the user can filter the top 20 % of the biggest assemblies regarding number of parts included in order to focus on the assemblies with the highest savings potential. Analogous to the assembly analysis, the area analysis workflow allows for the reintroduction of the calculation results (areas, similarity values) in the original BOM hierarchy graph.

4 CASE STUDY AT MAN

This section summarizes the application of the assembly and area analyses to BOM data sets of 14 different rear axle component variants. The data was extracted from MAN's database system to an Excel table with 4250 rows, each resembling one node in the BOM tree. Thus, the average number of nodes per component variant tree was 304. First, we applied the **assembly analysis** using both the CPA and APTED algorithm for the pairwise calculation of assembly similarities. In the first step, we imported

the multi-level BOM data into Soley Studio and stored it as data graph. In order to gain a preliminary overview of the data set, we plotted the assemblies of all component variants in a 2D portfolio (see Figure 4 (a), wherein the horizontal axis represents the assembly size (i.e. number of parts included). The vertical axis represents the assembly relevance, which is defined as the sum of all parts that would be affected by the standardization of an assembly. Hence, all affected parts from all assemblies in different component variants are taken into account. The assembly relevance can therefore be considered as the lever of the desired standardization effects throughout the existing product families. The portfolio chart is based solely on BOM data and helps to identify the most promising assemblies regarding the achievable cost savings potential in case of standardization. These assemblies are located in the top right part of the chart. In our rear axle data set, three of the most promising assemblies are A1, A2 and A3. These are different variants of axle gearings. After the filtering process in step three of the assembly analysis, three clusters of similar assemblies remained, which have both a similarity value $\geq 80\%$ and are part of the top 20% regarding the assembly size. Both algorithms produce very similar results. Thus, only the result of the assembly analysis using the CPA is shown in Figure 4 (b). The left cluster contains the three assemblies, which we identified as the theoretically most promising standardization candidates. Thus, the axle gearings A1, A2 and A3 should be analysed in depth to assess their practical standardization potential. To support this analysis, we reintroduced the calculation results in the original product structure in the fourth step and deleted all part numbers which had already been standardized. The graph visualization in Figure 4 (b) allows for the detailed comparison of the three assemblies in question through all levels of the BOM hierarchy.

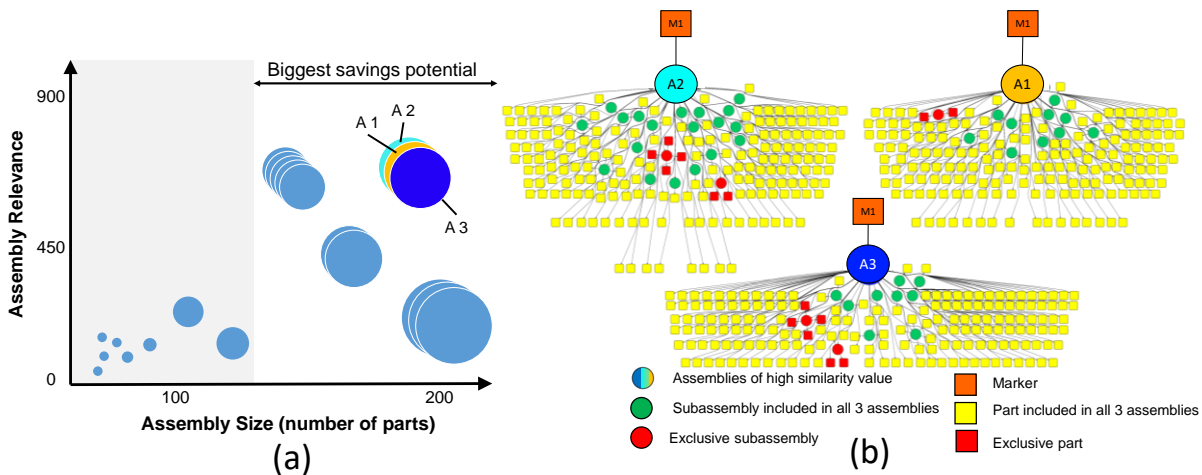


Figure 4. 2D portfolio to identify assemblies with a high savings potential (a) and reintroduction of the calculation results in the original product structure (b)

Our assembly analysis workflow enriched the graph with a marker ("M1") which clearly identifies this cluster of similar assemblies. Furthermore, subassembly nodes that are part of all three assemblies are highlighted in green, whereas the parts included in all three assemblies are yellow. The red nodes represent subassemblies (circles) or parts (squares) which are exclusive for each assembly. In this case, the red elements represent different crown and bevel gears that realize functional differentiation of the axle gearings. They cause an intended variety, which is perceptible and valuable from the customer's perspective. Thus, the red nodes do not represent standardization candidates. Instead, the green subassemblies and the yellow parts - which constitute more than 80% of the nodes - are potential standardization candidates. The practical feasibility of their standardization has to be assessed in further analyses by experts. As the graph visualization provides an intuitive representation of the potential standardization candidates and exclusive parts, it is a valuable support for such further analyses.

In addition, we applied the **area analysis** to the same data set. As the similarity calculation is based on the same algorithms as in the assembly analysis, the results are essentially identical. Using the area analysis, we identified the same three axle gearing assemblies A1, A2 and A3 as most promising assemblies, and found the same exclusive parts and subassemblies as by means of the assembly analysis (red elements in Figure 4 (b)). Going beyond the isolated analysis of the assemblies, the area analysis quantifies the similarity of the areas surrounding the most promising assemblies and creates a visualization of both the areas and the calculation results in the original BOM graph. Hence, the area

analysis supports the in-depth assessment of standardization possibilities in whole areas of product structures, whereas the assembly analysis focuses on the similarity calculation and visualization of assembly pairings. The result of the assembly analysis is a subset of the result of the area analysis. Both methods' graph visualizations allow for a detailed breakdown of all discovered similarities to the level of a single part node.

Table 2 shows a comparison of the CPA and the APTED algorithm for both analysis methods. The main difference is that the CPA calculates the similarity between two assemblies solely based on the parts included. Thus, only the lowest level of the BOM hierarchy is taken into account, whereas the APTED algorithm also considers the structural similarity on all hierarchy levels of a tree. As each BOM describes just one component variant and not a whole product, the number of hierarchy levels of the analysed trees is low. This is the reason for the low maximum deviation of 3% in the calculation results. When analysing BOMs with a higher number of hierarchy levels, the results produced by the APTED algorithm are significantly more accurate. However, this accuracy is accompanied by a considerably higher run time ranging from about factor 5 (assembly analysis) up to about factor 17 (area analysis) for the given data set, which comprised up to six hierarchy levels per component variant BOM.

Table 2. Comparison of the CPA and the APTED algorithm

		Common Parts Algorithm (CPA)	APTED Algorithm
Assembly Analysis	Calculation results	<ul style="list-style-type: none"> • Almost identical • Calculated similarity values marginally higher using the APTED algorithm (maximum deviation: 3%) • Results are a subset of the results of the area analysis 	
	Run time [s]	0.96	4.59
Area Analysis	Calculation results	<ul style="list-style-type: none"> • Almost identical • calculated similarity values marginally higher using the APTED algorithm (maximum deviation: 3%) 	
	Run time [s]	1.46	25.30
Advantages		<ul style="list-style-type: none"> • Short run time • Simple programming • Good results for simple product structures 	<ul style="list-style-type: none"> • Content- and structural similarity • High accuracy for complex product structures • Multiple setting parameters (cost sets)
Challenges		<ul style="list-style-type: none"> • Content similarity only, no consideration of structural similarity • „Clean“ Bom data required 	<ul style="list-style-type: none"> • Higher run time • „Clean“ Bom data required • Determination of optimal setting parameters
Application		<ul style="list-style-type: none"> • Simple product structures • Rough analyses of complex product structures 	<ul style="list-style-type: none"> • Complex product structures (high number of hierarchy levels, high number of parts at each level)

5 CONCLUSION AND OUTLOOK

We presented an approach to increase transparency in existing product families by analysing product structure information. The aim was to provide improved input for standardization and modularization methods that reduce non value-adding variety in the portfolio, and thus to support effective variant management in an industrial context. A literature review, as well as industrial practice in the fields of variant management and product family design revealed that modelling BOM data holds great potential to develop a generalized and practical approach. We applied graph-based modelling of product structures for automated handling of large data sets and a state-of-the-art tree matching algorithm for similarity calculations. The approach helps with answering the following questions: Which assemblies or whole areas of the component variants are theoretically promising standardization candidates due to high similarity and high savings potential? How similar are these variants? Where are the similar assemblies and areas located in the product structure, and which subassemblies or parts do they contain? A case study conducted with real product data from MAN showed that the approach provides valid and practicable search results for standardisation candidates in a portfolio of complex product variants.

Henceforth, the developed approach can be used as valuable input for further standardization and modularization methods to reduce non-customer-value-adding variety in the portfolio and thus to support an automated and effective variant management in an industrial context. To exploit the full potential of the APTED algorithm, future research should tackle the appropriate definition of the transformation cost sets. A sensitivity analysis on adjusting transformation costs and scaling these with actual part costs would allow for more sophisticated similarity calculations. The biggest potential for future research lies in the possibility to extend the similarity analysis of bare product structure data by incorporating a mapping of component variants to customer requirements data, which represents the actual external variety. Especially the comparison of differentiating component features to meet the

requirements of different use cases to the internal variance discovered by the presented similarity calculation approach allows for a direct assessment whether varying elements represent demanded external variety and therefore add customer value. This enhanced insight allows for a systematic questioning of existing internal product variety in a historically grown portfolio of complex products and is a valuable input for the redefinition of platform and variable components.

REFERENCES

- Braun, B., Kissel, M., Förg, A. and Kreimeyer, M. (2013), "Referenzmodell zur Analyse von Baukastensystematiken in der Nutzfahrzeugindustrie", in Proceedings of the DfX Symposium, Hamburg, Germany, Sep 19-20, 2013, TuTech Verlag, Hamburg, Germany, pp. 25–37.
- Collier, D.A. (1981), "The Measurement and Operating Benefits of Component Part Commonality", *Decision Sciences*, Vol. 12 No. 1, pp. 85–96.
- Eppinger, S.D. and Browning, T.R. (2012), *Design structure matrix methods and applications*, Engineering systems, MIT Press, Cambridge, Mass.
- Harlou, U. (2006), *Developing product families based on architectures: Contribution to a theory of product families*, Department of Mechanical Engineering, Technical University of Denmark, Lyngby.
- Heckel, R. (2006), "Graph Transformation in a Nutshell", *Electronic Notes in Theoretical Computer Science*, Vol. 148 No. 1, pp. 187–198.
- Jiao, J., Simpson, T.W. and Siddique, Z. (2007), "Product family design and platform-based product development", *Journal of Intelligent Manufacturing*, Vol. 18 No. 1, pp. 5–29.
- Jiao, J. and Tseng, M.M. (2000), "Understanding product family for mass customization by developing commonality indices", *Journal of Engineering Design*, Vol. 11 No. 3, pp. 225–243.
- Jiao, J., Tseng, M.M., Ma, Q. and Zou, Y. (2000), "Generic bill-of-materials-and-operations for high-variety production management", *Concurrent Engineering*, Vol. 8 No. 4, pp. 297–321.
- Kreimeyer, M., Förg, A. and Lienkamp, M. (2014), "Fostering Modular Kits in an Industrial Brownfield Environment", in Proceedings of TMCE 2014, Budapest, Hungary, May 19-23, 2014, Delft University of Technology, Delft.
- Kreimeyer, M. and Lindemann, U. (2011), *Complexity metrics in engineering design: Managing the structure of design processes*, Springer, New York.
- Lindemann, U. (2016), *Handbuch Produktentwicklung*, Hanser, München.
- Pawlik, M. and Augsten, N. (2016), "Tree edit distance. Robust and memory-efficient", *Information Systems*, Vol. 56, pp. 157–173.
- Pirmoradi, Z., Wang, G.G. and Simpson, T.W. (2014), "A review of recent literature in product family design and platform-based product development", in *Advances in product family and product platform design*, Springer, pp. 1–46.
- Schuh, G. (2005), *Produktkomplexität managen: Strategien, Methoden, Tools*, Hanser, München.
- Schumacher, M., Märkl, S., Gilbert, M. and Kreimeyer, M. (2015), "Variantenplanung in der Produktarchitektur: Erfahrungen und Best Practices aus der Nutzfahrzeugentwicklung", in Proceedings of Tag des Systems Engineering, Herzogenaurach, Germany, Oct 25-27, 2015, Hanser, Munich, Germany, pp. 125–139.
- Steva, E.D., Rice, E.N., Marion, T.J., Simpson, T.W. and Stone, R.B. (2007), "Two Methodologies for Identifying Product Platform Elements Within an Existing Set of Products", in Proceedings of DETC2006, Philadelphia, PA, Sep 10–13, 2006, ASME, New York, pp. 811–821.
- Tilstra, A.H., Seepersad, C.C. and Wood, K.L. (2012), "A high-definition design structure matrix (HDDSM) for the quantitative assessment of product architecture", *Journal of Engineering Design*, Vol. 23 No. 10-11, pp. 767–789.
- Tseng, M.M., Jiao, J. and Merchant, M.E. (1996), "Design for Mass Customization", *CIRP Annals - Manufacturing Technology*, Vol. 45 No. 1, pp. 153–156.
- Ulrich, K. (1995), "The role of product architecture in the manufacturing firm", *Research Policy*, Vol. 24 No. 3, pp. 419–440.
- Wacker, J.G. and Treleven, M. (1986), "Component part standardization. An analysis of commonality sources and indices", *Journal of Operations Management*, Vol. 6 No. 2, pp. 219–244.
- Wildemann, H. (2016), *Variant Management: Guidelines for the Reduction, Control and Avoidance of Complexity in Products and Processes*, TCW, Munich, Germany.
- Yu, J.S., Gonzalez-Zugasti, J.P. and Otto, K.N. (1999), "Product Architecture Definition Based Upon Customer Demands", *Journal of Mechanical Design*, Vol. 121 No. 3, p. 329.

ACKNOWLEDGEMENTS

We would like to thank Thomas Gmeiner and Peter Grüner from Soley GmbH for their technical support in terms of algorithm implementation. The project was funded by MAN Truck & Bus AG and TUM.