

Dependency Graph Based on User Taxonomy and Related Parameters for more Efficient Collaborative Work

Franca A. Rupprecht^{1,a*}, Francois M. Torner^{2,b}, Jörg Seewig^{2,b}, Achim Ebert^{1,a}

¹Computergraphics & HCI, University of Kaiserslautern, Germany

²Institute for Measurement and Sensor-Technology, University of Kaiserslautern, Germany

^a[rupprecht, ebert]@cs.uni-kl.de, ^b[torner, seewig]@mv.uni-kl.de

Keywords: Computer-supported collaborative work, dependency graph, user taxonomy, visualization techniques, group performance, dependency parameters

Abstract. Collaboration, performed by a large group of experts of diverse fields and competences, is a time-demanding and complex process. It is crucial to provide tools to facilitate the identification and manipulation of interdependencies as well as the active collaboration process. Dependency graphs of the participants can help to improve processes, to plan tasks, and to identify potential for more efficient cooperation. Such a dependency graph comprises clear defined entities, which are linked with each other based on defined relationships [1]. In the course of this work, a taxonomy of users in industrial corporations will be introduced, which is needed to define the entities and relationships of the dependency graph and is easily adaptable to specific corporations. Such a taxonomy cannot be found in the literature, but is important for the design and development of software products under the rules of user centered design [2]. However, there is still the big challenge to display a meaningful relation between those entities and to give an easy understandable overview of the whole relationship with the goal to solve complex tasks and to improve a groups' performance. Therefore, a set of parameters will be introduced, which help to find out how good tasks and work packages are distributed within the network. State-of-the-art techniques are used to visualize and recognize interdependencies and information flow. Based on a case study, the findings of this work are embedded and combined in an interactive and intuitive user interface that facilitates planners to recognize and explore complex multi-dimensional networks.

Introduction

The conceptual design and critical assessment of complex systems often requires large teams, e.g. scientists, engineers and planners, who need to work together. The task of conceptual design typically involves several iterations before an acceptable solution can be obtained. Such a process is extremely time consuming. In the collaborative process team members often switch between tasks to achieve successful cooperation. Changes, performed by team members, effectuate the entire system and individual performed tasks. According to Johnson et al. [3] it is necessary that team-members understand this impact and adjust tasks and operations accordingly. Therefore, the availability of a common framework to support the decision-making process is highly required. Ideally, experts can work independently or jointly on certain aspects/subsystems - still focusing on the design goal that must be achieved. The system is supportive for the whole process by visualizing connections, dependencies, or system changes. More and more activities can be delegated to the system, but it is crucial to identify which activities and tasks are performed in the collaborative work and who performs those activities. Detailed task-models under the rules of user centered design and dependency graphs need to be modeled and visualized to enable the data exchange and to consider the dependencies between different users. Therefore, a user taxonomy has to be deployed and is used as fundament to describe those task models. Exemplary visualizations of a chosen dataset are shown in Fig. 1. Based on the task models and dependency graph, explicit visualization and interaction techniques can be created and integrated into a virtual collaboration environment.

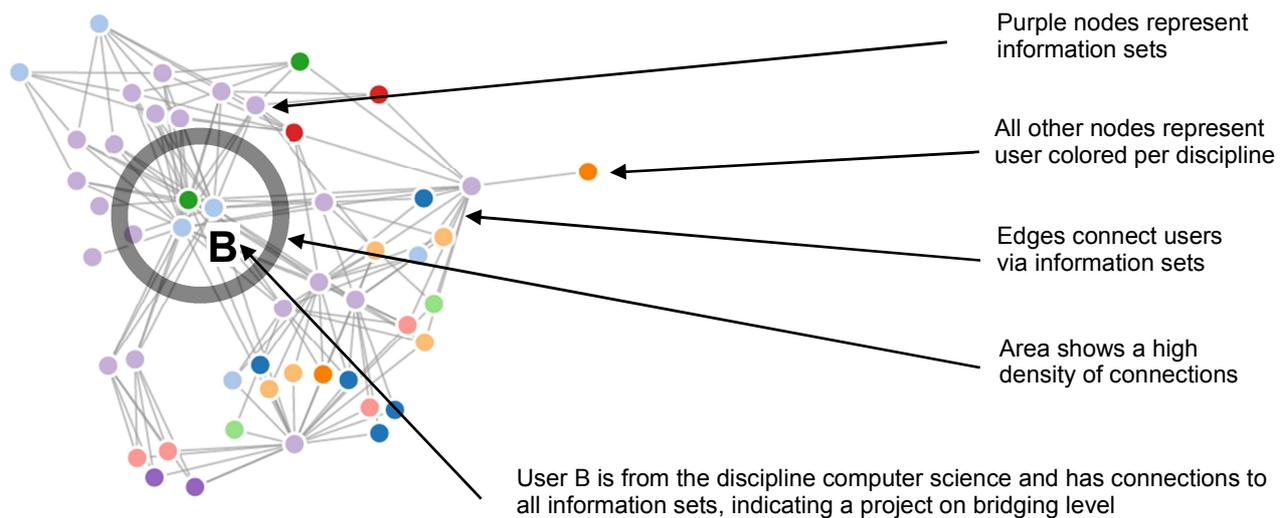


Figure 1 – Dependency graph of dataset, representing projects within the IRTG 2057 college [26], visualized with the use of a force-directed graph. The interactive 2-dimensional layout uses draggable nodes to facilitate the recognition of existing connections.

Such a collaboration environment is IN2CO (intuitive & interactive collaboration), a human-centric visualization framework for intuitive and collaborative data exploration and manipulation has been developed therefore [4]. Specifically, its contribution is the integration of ubiquitous technologies and existing techniques to explore data and dependencies in collaborative decision-making for co-located and distributed participants. The general collaborative framework is aimed to support collaborative work in an efficient manner when bringing diverse areas of expertise together. In this publication, a user taxonomy is introduced, which is used to clearly define entities and relationships in a system in order to model a network of their relations. Afterwards, a set of dependency parameters is proposed, to display a meaningful relation between objects and to give an easy understandable overview of the whole relationship with the goal to solve complex tasks and improve a groups' performance. Following, several methods will be introduced and analyzed, that give the possibility to visualize class data and show relations, connections and interdependencies of group members. The visualization techniques used in this paper are: Force directed graphs, Sankey diagrams, ring based radial visualization with hierarchical edge bundling. Those differing visualization techniques are applied to make a complex system architecture easily understandable for users, fading out unnecessary information, illustrated in a case study. The given data of the case study, containing different user groups that are acting in a virtual, industrial corporation, is related to the IRTG 2057 college. The system itself is represented by dependency parameters, using three different types of visualization. The representations are embedded in an interactive application, which allows users to explore the recognized dependencies and links and which is described succeeding.

The findings and gathered information of the investigation can be reused to model dedicated users and activities with a high level of detail and give insights of systems architecture. Together with the deducible dependency graph, explicit visualization and interaction techniques can be created and integrated into the IN2CO framework, which can substantially enhance efficiency and user-centered aspects of the distributed collaboration environment for design, simulation and analysis efforts, and this is our main contribution.

Related Work

The whole significance has been defined in the set of issues of *joint activity*, in which already extensive studies were performed, which will be elucidated in the following. *Joint activity* is exemplary defined by Johnson et al. [5]: "In joint activity, individual participants share an obligation to coordinate, to a degree sacrificing their individual autonomy in the service of progress

toward group goals.” In contrast, coordination is defined as ”managing dependencies between activities” [6]. Within interdependent activities conflicting interests are present. In order to achieve common goals, conflicting interests have to be coordinated to capture discrepancies before they become serious with the help of common grounds [7]. Those are supported by continuously informing others about changes that have occurred outside their views [5]. Really interesting findings have furthermore been explored by Johnson et al. who stated, that not all team members must be fully aware of the entire scope of an activity; but every participant needs to be aware of the interdependence in-between their activities [4]. The awareness of tasks and performed activities within a collaborative work session influences the coordination and performance of tasks in a positive manner. Due to established common knowledge and impact awareness, team members can work together effectively and adjust their activities as necessary [8]. Van der Veer et al. [3] determined, that a high level of details in task modeling is needed for collaborative work in order to design activity assignments and optimal support by the system. Similar research is being pursued in the field of robotics, where monitoring processes using an overview level is crucial. One well-known example is supervisory control, where a user allocates tasks to machines and monitors execution performance [9, 10]. One solution providing the desired insight is done via an additional display monitoring the current status of a process [11].

The overall goal of this paper is to prepare a given dataset in a way that dependencies and activities are extracted and become recognizable. Initially, independent projects are examined and overlapping can be found. First, we introduce a user taxonomy that describes users in a domain, considers different aspects and tries to classify those users. Based on the taxonomy, users and activities can be dedicatedly modeled with a high level of detail as required. The visualization techniques used in this paper are: Force directed graphs, Sankey diagrams, ring based radial visualization with hierarchical edge bundling. Force directed graphs have the attributes of evenly distributed vertices, edges with uniform lengths, and reflecting in order to create higher level of aesthetic and better readability for the user [12]. This approach has been applied in uncountable cases and served as basic for more enhanced visualizations like in [13, 14, 15, 16]. Traditionally, Sankey diagrams are applied to visualize energy flows or material flows. Those diagrams represent quantitative information about flows, their relationships, and their transformation depict as directed, weighted graphs [17]. The possibly most famous Sankey diagram is Charles Minard's Map of Napoleon's Russian Campaign of 1812, which was created 1869, even before the actual Sankey diagram has been introduced in 1898 [18]. Sankey diagrams are applied in a wide range of fields, such as in [19, 20, 21]. The third visualization technique used in this work is ring based radial visualization with hierarchical edge bundling as characterized by [22]: Nodes are positioned around the circumference of a ring; line segments (edges) are used to connect the nodes; additional nodes optionally appear in ring's interior. Ring based radial visualizations are commonly used to depict relationships among disparate entities. The edge bundling is used to reduce the visual complexity. Radial visualizations are widely used and enhanced by e.g. [23, 24, 25].

User Taxonomy in Industrial Corporations

Industrial corporations rely on interdisciplinary collaboration of integrated complex design and decision-making. The conceptual design and critical assessment of complex systems generally requires large teams of scientists, engineers and planners who collaborate, bringing together different aspects and knowledge from different areas. The conceptual design process is extremely time-consuming, typically involving several iteration steps to improve conceptual designs before a generally acceptable solution is obtained. In this taxonomy, users in industrial corporations are described based on three categories: System-level, discipline, and tasks.

System-level. Industrial corporations can be adapted to a layer model, where every layer is considered as one subsystem that is in an exchange relationship with the neighboring levels. Design decisions on the lowest level escalate to the highest level and vice versa. The following levels can be distinguished as shown in Fig. 2: factory level, machine level, process level, and the bridging

level, which combines all aspects and brings those levels together. This level model is related to the whole IRTG 2057 [26] structure, where every participant is assigned to one of those given levels. More details on the levels will be given.

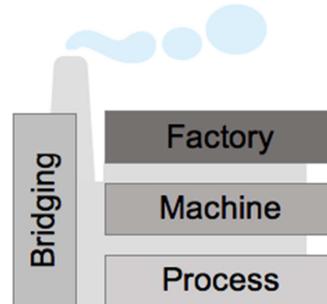


Figure 2 – System-levels in industrial corporations related to the structure of the IRTG 2057 college with the title “*Physical Modeling for Virtual Manufacturing Systems and Processes*” for the successful realization of activities and tasks in an analogically industrial project.

The *factory level* comprises physical properties, features of the factory, and of the manufactured products. Exemplary, factory-level transactions describe the material flow within the whole manufacturing process or conditions of the indoor environment. On the *machine level* the focus lies on machine tools and their components, tooling systems, and measuring instruments. Machine-level transactions specify manufacturing processes on the workbench level and include physical characteristics like deformation, stress, and temperature. On the *process level* single machining processes themselves (e.g. cutting, grinding, and milling) find consideration. Material properties as well as material behavior of the work piece and the machining tools under machining conditions are investigated. Changes made on the lower levels have an influence on the overall production program, output, and quality to a higher degree. The exchange of transactions between levels and the connections between each other as well as the consideration of cross-references is investigated and developed on the *bridging level*. Correlations and interdependencies between the levels, as exemplarily shown in Fig. 1, are identified and connections are ensured in order to obtain a comprehensive view of the manufacturing system.

Disciplines. Industry is defined as the type of trade that is distinguished by the production and the processing of material goods. Mainly performed disciplines in many industrial corporations are mechanical engineering, industrial engineering, electronic engineering, and computer science. An example of disciplines and their corresponding sub-disciplines is depicted in Fig. 3. This categorization was used for the case study conducted in this research, but it can easily be adapted to the disciplines that are important for different kinds of corporations.

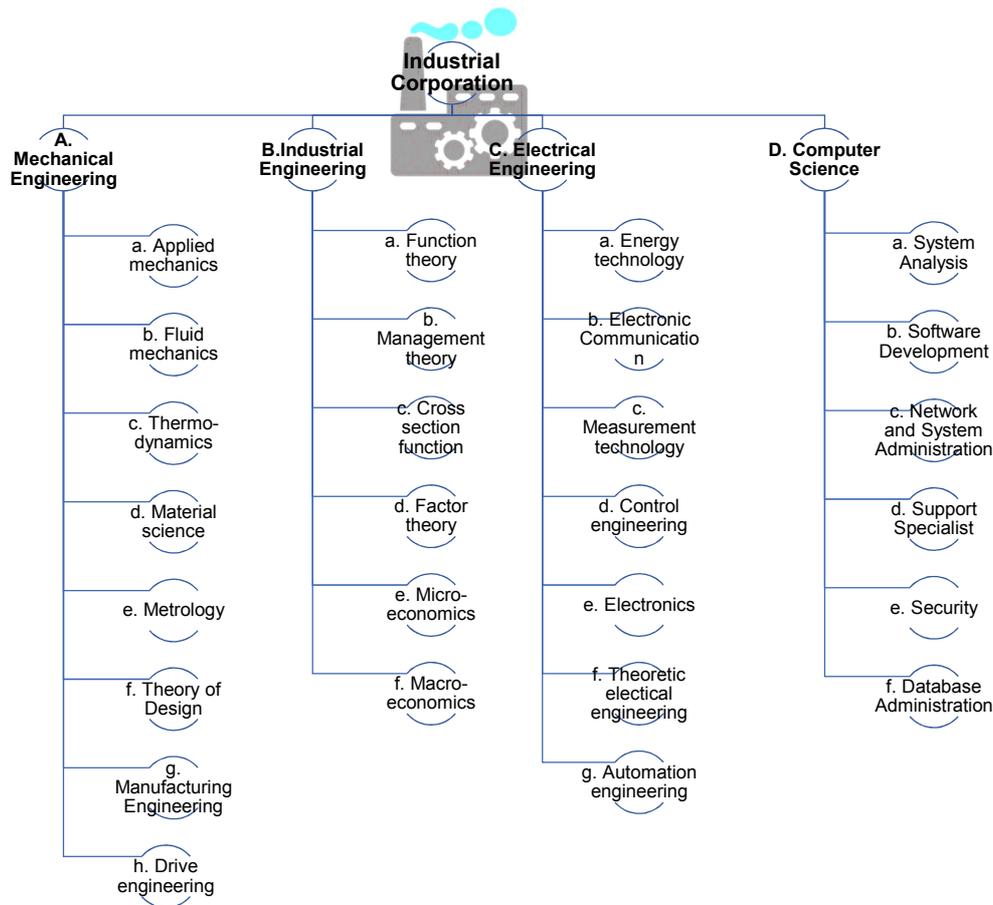


Figure 3 – Overview and classification of industrial disciplines within a typical industrial corporation. Segmentation of the individual categories in dedicated ranges of tasks for the realistic modelling of an industrial corporation. Classification based on [27, 28, 29].

Mechanical engineering is one of the main disciplines performed within an industrial corporation, performing the tasks of design, analysis, manufacturing, and maintenance of mechanical systems and assigns the principles of engineering, physics, and materials science, subdivided in the disciplines shown in Fig. 3, column 1. Mechanical engineering involves the design, production, and operation of machines. Business administration as sub-discipline of *industrial engineering* has the goal to describe, explain, and support decision-making processes, mainly performed by several experts. Exemplary fields of activities in industrial corporations are designing the product program, optimize factory layouts, insurance of part deliveries and many more. Sub-disciplines of industrial engineering are listed in Fig. 3, column 2. *Electrical engineering* deals with the research, development, and production of electrical appliance. Exemplary, electrical machines, components, and circuits need to be utilized in embedded systems and manufacturing systems. Fig. 3, column 3 lists the sub-disciplines of electrical engineering. For example, measurement technology and automatic engineering are disciplines that have a close relation to mechanical engineering and are very frequently adapted in industrial corporations. Especially the field of *computer science* experienced a strong growth in industry. The reason is a growing demand of applications and systems that are computer-driven and necessarily need to be developed by computer scientists (networking, embedded systems, data-management, etc.) The term “digital transformation” plays an important role for corporations to remain competitive and to open up new markets. Consequently, disciplines of computer science are deeply integrated into the daily business, as depicted in Fig. 3, column 4.

The proposed classification of industrial disciplines serves as overview for incorporated disciplines in industrial corporations, but does not describe the generality for all companies. The diagram merely comprises exemplifying activity fields, which is not generalizable for all

organizations. There might be corporations with a different discipline distribution, and differing activity fields.

Tasks. The tasks in an industrial corporation are centralized in a *design cycle* in Fig. 4 (left), which illustrates the iterative optimum quality assurance concept and comprises the tasks analysis, scenario building (Sc. Building), planning, design, and evaluation. All tasks are proceeded on all levels, leading to cross-references between disciplines, tasks, and levels.

Analysis deals with the observation of causalities, in which one decision is partly responsibly to an effect. The responsibilities and dependencies are ascertained and formalized. The *scenario building* phase addresses the task of developing processes, strategies, and guidelines based on those causalities. Those strategies are collected and classified into diverse use cases. In the *planning* phase, the current scenario is acquired and mapped with collected use cases. The different strategies are balanced and adapted to the current scenario. The chosen strategy is elaborated and transformed into an operating plan to achieve the overall business goals. *Design* addresses the actual task of manufacturing work-pieces, design of machines and their arrangements, composition of materials, and fabrication of products. Once all strategies are implemented and realized, performance indicators are acquired and analyzed within the *evaluation* phase. Quality insurance tasks and maintenance can be comprised in this phase. Based on the evaluation the current scenario is newly valued and the design cycle is reiterated.

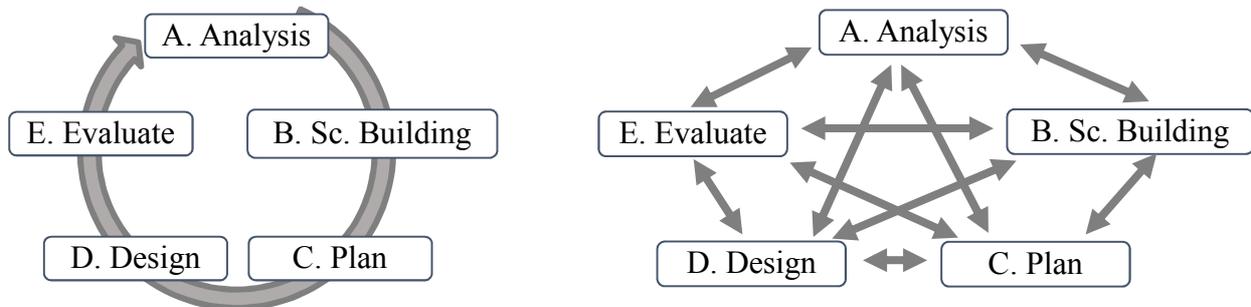


Figure 4 – Visualization of tasks within an industrial corporation. Left: Optimal design cycle with a perfectly structured task sequence. Right: Optimizable design process with many cross-references and feedback loops.

Typically, the design process in industrial corporations is not structured conducted as described. Interactions, adjustments, and spontaneous decisions take place, which might disturb the process structure and which must be controlled. To steer the process flow it is necessary to model it, to visualize it, and to positively influence it with the use of duly appropriated parameters, which will be described in this publication. While the left design cycle shows an optimal process, the right one (Fig. 4) in contrast represents a more realistic design process. The structure is highly; cross-references between tasks exist or that phases are continuously performed simultaneous to other tasks. Especially the task of evaluation and quality assurance is performed continuously over all tasks, phases, and disciplines.

Similar to the classification of disciplines this classification of tasks serves as overview for performed tasks in industrial corporations, but does not describe the generality for all companies. The diagram merely comprises activity fields, which can differ in chronological order, scopes, and defined sub-activities within the category.

Dependency Parameters

To visualize the existing dependencies within the network so-called Sankey plots are used. A Sankey plot is a well-known and commonly used method to highlight relations between several participants, objects or elements within a group. Cubic splines can be used to visualize those connections. Nevertheless, there is the challenge to display a meaningful relation between those objects and to give an easily understandable overview over the whole relationship to help solving

the upper described process (cf. Fig. 4, right). The line plots themselves quite often do not lead to an immediate understanding of a complex system as shown in Fig. 1. Therefore, it is mandatory to define a set of parameters that can be visualized, using color-coded line plots or transparency definitions within a Sankey diagram. This problem will be discussed considering a case-dependent two- or three-dimensional representation of subcategories and a set of elements in a main category. Considering several categories, whose relations between each have to be investigated the main category describes the entities which are in focus of the investigation and which are located in the center in the visualization. For example, the main category can represent a group of cooperating people where the focus is lying on their cooperation. System-levels, disciplines, and tasks are defined to be subcategories that need to be considered and help to optimize the upper process. The “system complexity” exemplarily is one parameter that characterizes the whole system. Before describing a set of parameters, view definitions will be made.

Let us assume that there are $P \in \mathbb{N} \setminus \{0\}$ elements in the main category, which are in the focus and can be addressed by an index $p \in \{0, \dots, P-1\}$. Furthermore, let us assume that there are

$N \in \mathbb{N} \setminus \{0\}$ categories, which can be addressed by an index $n \in \{0, \dots, N-1\}$. Each category contains $K_n \in \mathbb{N} \setminus \{0, 1\}$ individual elements and is addressed using the variable $k_n \in \{0, \dots, K_n-1\}$.

System complexity. The number of connections between the categories and the main category defines the complexity of a system. A connection is described by the parameter $\delta_{k_n,p} \in \{0, 1\}$ where

$$\delta_{k_n,p} = \begin{cases} 1 & \text{if connection between main and sub } n \text{ exists} \\ 0 & \text{else.} \end{cases}$$

The maximum number of possible connections normalizes the parameter. The complexity then leads to a scalar value and can be defined as

$$\text{comp} = \frac{\sum_{n=0}^{N-1} \left(\sum_{k_n=0}^{K_n-1} \sum_{p=0}^{P-1} \delta_{k_n,p} \right)}{P \cdot \sum_{n=0}^{N-1} K_n} \in [0; 1] \tag{1}$$

comp = 100 % is the maximum possible complexity and comp = 0 % is the minimum complexity of the given system. This parameter does not contain any information on how homogeneous connections are distributed and therefore it is not very appropriate to visualize this parameter or make a judgment of the system’s balance (work balance).

User’s specialization regarding one category. Nevertheless, it is a good idea to have a closer look on the distribution of connections between a chosen category n and the main category. That is why it is necessary to focus on how many connections of an element within the main category exist for each subcategory. The parameter

$$v_{n,p} = 1 - \frac{\left(\sum_{k_n=0}^{K_n-1} \delta_{k_n,p} \right) - 1}{K_n - 1} \quad \text{where} \quad \left(\sum_{k_n=0}^{K_n-1} \delta_{k_n,p} \right) \geq 1 \tag{2}$$

can be visualized using color coding. The parameter makes the assumption that there is at least one group-related connection for each element available. This parameter gives a value for the degree of specialization of an element of the main category with respect to a chosen subcategory. Per definition $v_{n,p} = 0$ is a low degree of specialization and $v_{n,p} = 1$ is a high degree.

User's specialization regarding all categories. Since $v_{n,p}$ is a subcategory based parameter it might be interesting to define a mean value that is calculated from this parameter. Summing up all categories and considering the number of categories leads to

$$\mu_p = \frac{1}{N} \cdot \sum_{n=0}^{N-1} v_{n,p} = \frac{1}{N} \cdot \sum_{n=0}^{N-1} \left(\left(\sum_{k_n=0}^{K_n-1} \delta_{k_n,p} \right) - 1 \right) / (K_n - 1) \quad (3)$$

which gives the user an idea of the mean value. A low degree of specialization means that a selected user has connections to many subcategories. Consequently, the work balance is not well distributed and a system adjustment should be considered. Otherwise, it might be desired that a user works on several subcategories (e.g. he is a participant of the bridging level). To avoid an overload of this user other subcategories should be taken into account as well. This can be done calculating the empirical standard deviation of this parameter, which is described below.

Empirical standard deviation of a user's specialization. Consequently, the empirical standard deviation can also be visualized which gives the applicator an idea of the homogeneity for each object and leads to the information if there is a high deviation between degrees of specialization for all elements of the main category or not. The empirical standard deviation is defined as

$$s_p = \sqrt{\frac{1}{N-1} \cdot \sum_{n=0}^{N-1} (v_{n,p} - \mu_p)^2} \quad \text{if } N \geq 2. \quad (4)$$

If there is a low degree of specialization and a low standard deviation, this means that a user might be overloaded because he has connections to many subcategories. This must be avoided to improve the system's work balance and consequently the process flow (cf. Fig. 4).

Group coverage. So far the described parameters have had their focus on the main category. It might be interesting to set the focus on the subcategories as well. The group coverage is defined as a parameter that can be used to analyze how big the support of the main category actually is. This might be interesting within process planning if a big group must solve several kinds of tasks. This is the case within the IRTG 2057 group. The group coverage may be calculated with the equation

$$w_{n,k_n} = \frac{1}{P} \cdot \sum_{p=0}^{P-1} \delta_{k_n,p} \in [0;1] \quad (5)$$

If w_{n,k_n} is a big value this means that many elements of the main category have a connection to the selected category. In conclusion, if the main category contains people working in a group and if there are tasks defined within the subcategory, this value gives an idea on how good the coverage of the tasks actually is and might help to improve decision-making processes.

Mean group coverage and empirical standard deviation. In analogy to the defined parameters μ_p and s_p the category related mean value μ_n and s_n are defined as

$$\mu_n = \frac{1}{K_n} \cdot \sum_{k_n=0}^{K_n-1} w_{n,k_n} \quad (6)$$

and

$$s_n = \sqrt{\frac{1}{K_n-1} \cdot \sum_{k_n=0}^{K_n-1} (w_{n,k_n} - \mu_n)^2} \quad \text{if } K_n \geq 2. \quad (7)$$

Exemplary data set

The exemplary dataset is constructed based on actual projects within the International Research and Training Group (IRTG 2057, “*Physical Modeling for Virtual Manufacturing Systems and Processes*”). Although production planning is not a new research topic, this field still offers potential for optimization. With the use of theoretical computer models, production is planned from different points of view: from a single machine, up to a complete factory. However, the computer models lack actual physical properties, which enable to calculate key properties of a production line. Product quality or the energy consumption are calculated and targeted improvements can be performed. Like already mentioned, this program covers disciplines from mechanical engineering, industrial engineering, electrical engineering, and computer science allocated on given levels: process, machine, factory, and bridging (cf. Fig. 2). According to the discipline classification (cf. Fig. 3) and tasks (cf. Fig. 4) from above, users can be identified (cf. Table 1).

Table 1 – Identified users in the dataset related to IRTG 2057 college.

ID	User	Discipline	Task	S.-Level	ID	User	Discipline	Task	S.-Level
1	A	Applied mechanics	Scenario building	Process	16	Q	Computer science	Analysis	Process
2	B	Computer science	Analysis	Bridge	17	R	Fluid mechanics	Design	Machine
3	D	Manufacturing Eng.	Design	Process	18	S	Applied mechanics	Scenario building	Process
4	E	Computer science	Analysis	Factory	19	T	Applied mechanics	Design	Machine
5	F	Theory of design	Scenario building	Process	20	U	Theory of design	Evaluate	Process
6	G	Computer science	Design	Factory	21	V	Material science	Plan	Process
7	H	Computer science	Evaluate	Factory	22	W	Fluid mechanics	Analysis	Process
8	I	Management theory	Plan	Factory	23	X	Material science	Scenario building	Machine
9	J	Measurement technology	Analysis	Machine	24	Y	Theory of design	Scenario building	Process
10	K	Function theory	Plan	Factory	25	Z	Applied mechanics	Design	Machine
11	L	Function theory	Plan	Factory	26	AA	Theory of design	Evaluate	Process
12	M	Material science	Design	Process	27	AB	Material science	Plan	Process
13	N	Management theory	Evaluate	Factory	28	AC	Applied mechanics	Analysis	Process
14	O	Manufacturing Eng.	Plan	Machine	29	AD	Theory of design	Evaluate	Factory
15	P	Computer science	Scenario building	Factory	30	AE	Measurement technology	Design	Machine

The following types of information, which are needed and provided by a user, can be examined.

- Process requirements
- Material composition
- Material properties
- Material behavior
- Process properties
- Process behavior
- Nominal-actual comparison
- Machine properties
- Machine behavior
- Machine mechanisms
- Machine setting guidelines
- Production program alternatives
- Optimized production program
- Decision making support
- Facility properties
- Facility layout guidelines
- Indoor environment properties
- Indoor condition guidelines
- Quality parameter
- Service guideline
- Sustainability indicators

Two users are connected with each other if they receive or provide the same kind of information. There are two types of connections available: indirect and direct dependencies. A connection is

assigned as direct, if a user receives information from another user. Indirect connections mean, that information are shared by those users and can be manipulated by both of them. Furthermore, Fig. 5 represents the information flow and data exchange between other users and a common database system among all levels based on the described dataset. For a better readability, the users are connected via graphical elements, which represent the information type (also highlighted in Fig. 5). The cylindrical shapes depict the continuous write and read transactions towards and from the database system.

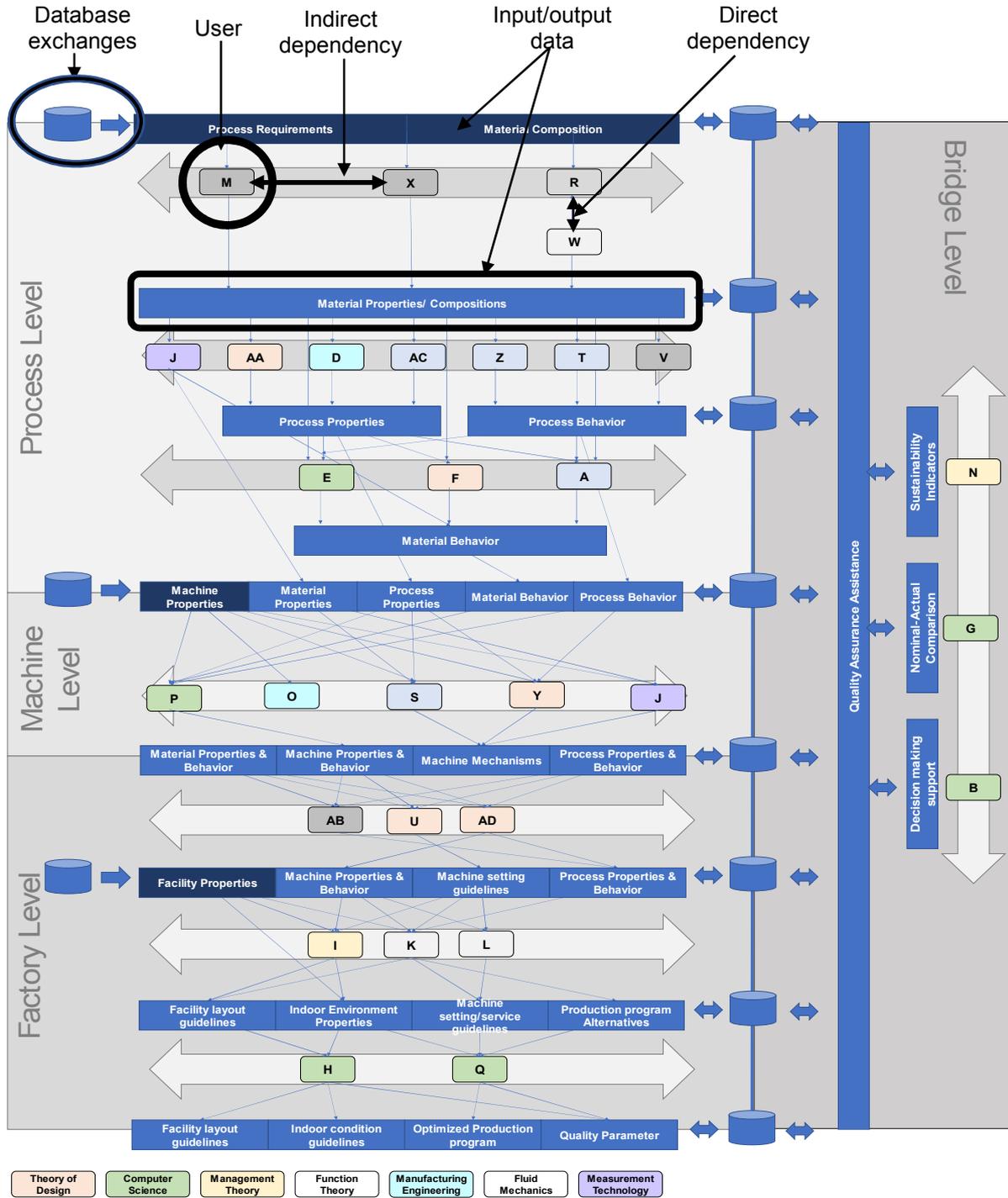


Figure 5 – Linkage of users with databases for detailed information exchange within an industrial environment. The whole information flow is a very complex network, which is hard to visualize and to understand.

Although the diagram might be overwhelming for the reader, we decided to include it to demonstrate the high complexity of the dataset and the consequential challenging design decisions

for the system architecture and task modeling. With the use of the defined dependency parameter we can confirm this statement. The complexity of the system is $\text{comp} = 0.411$, implying that the overall system complexity is within an acceptable range, which can be handled. For more reliable statements further analysis need to be considered. Exemplary specialization of user B for each individual category are: $v_{B,0} = 1$, $v_{B,1} = 0$, $v_{B,2} = 1$. It can be deduced, that user B has only one connection to category 1 and category 3 which means that he is specialized to one of the given subcategories (of each category, levels and disciplines). Regarding the second category (tasks), the user is not specialized but rather completely interdisciplinary. Hence, the user shows a mean specialization value over all categories of $\mu_{p=2} = 0.667$. The empirical standard deviation of a user B's specialization is $s_p = 0,5774$, implying that the degree of specialization has a comparable big variance over all considered categories. The user might have a good work balance. The mean group coverage values per category are $\mu_{n=0} = 0.111$, $\mu_{n=1} = 0.4933$, and $\mu_{n=2} = 0.25$. With the use of those values, the number of connections per category can be calculated: $\mu_{n=0} \cdot \dots \cdot n=0 = 0.111 \cdot 30 \cdot 9 \approx 30$. This number of connections indicates that each user is connected to exactly one subcategory. Consequently, category 2 (tasks) has 74 connections and category 3 (levels) 30 connections. The bigger the empirical standard deviation the higher is the disparity of the distribution of the users to subcategories. The calculated values $s_{n=0} = 0.0707$, $s_{n=1} = 0.0596$, and $s_{n=2} = 0.1774$ are really low, meaning the system has equal distributions among all categories.

Visualization Application

The whole dependency graph that is shown within Fig. 5 describes a very complex and hard to understand system. To get a more generalized overview of the given dependencies alternative visualization techniques need to be applied. Such an alternative visualization technique has already been shown at the very beginning (see Fig. 1). The given figure visualizes the identified dependencies using a force directed graph. The high amount of edges and crossing lines indicates a higher number of connections between all entities.

While Fig. 1 only shows the connections between the participants, Fig. 5 gives a detailed insight into the system structure, which is interesting especially for system's architects. It provides information on entities, data or material flow and the database structure. Nevertheless, the whole visualization might be overwhelming for other user groups, which do necessarily need such a deep insight into the architecture. Therefore, an appropriate visualization method for other user groups is introduced in Fig. 6. The given figure describes the relation between input and output data provided and needed by the users. Users who correspond to the same discipline are bundled as seen on the right-hand side of Fig. 6. The left-hand side of Fig. 6 in contrast represents database entities. Hovering over nodes highlights the connections. For example, the selected and highlighted user A has the green connections as input data and the red connections as output data. A red label combined with a green line indicates that this information type is both input and output data. Highlighted user N shows a high amount of connections to all database entities, which indicates that user N is on the bridging level, where it is necessary to have access to all available information. Hovering over a database entity highlights all users who are connected to that information type. Users that receive the information are colored in red, while users that provide the information are colored in green. Therefore, direct (green) and indirect (red) connections between users are indicated across the database entities.

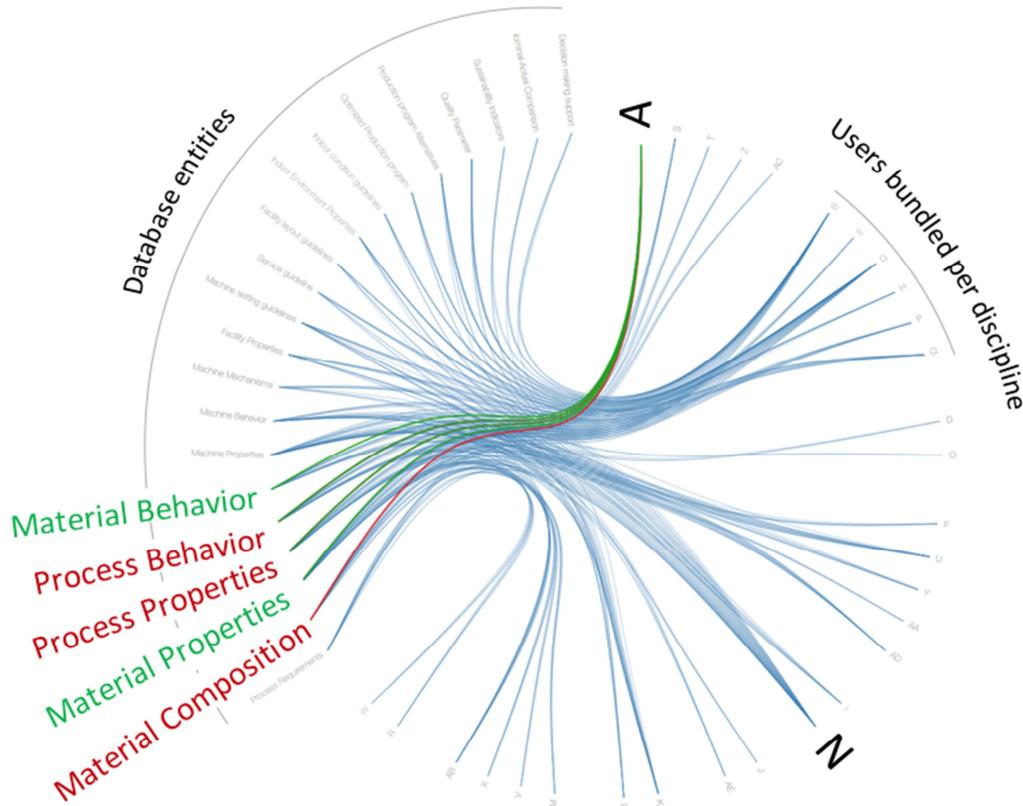


Figure 6 – Radial visualization of the dataset, which represents incoming information (red) and outgoing information (green). The nodes describe database entities and users.

In the course of finding simplified visualization methods for complex systems, an application to explore the dependencies and to get insights into parameters, which have been described in chapter “Dependency Parameters”, has been developed using Matlab. The result is an interactive environment, which is using the Sankey methodology, force directed graphs, ring based radial visualization with hierarchical edge bundling, and the deployed dependency parameter. The aim of this separate platform is to enable users to explore the recognized dependencies and links, which are deduced from the deployed user taxonomy. To support an efficient interaction with the model and easy information gathering, the application focuses the most important interdependencies by fading out irrelevant information and allows the user to switch between different visualizations without losing the focus. The initial view of the graphical user interface (GUI) is shown in Fig. 7.

The earlier described network visualizations (see Fig. 1 and Fig. 6) have been implemented, giving the possibility to set the focus on different aspects of the dataset. These are based on d3.js and Python, respectively embedded as web-view in the application. The user interface itself consists of a visualization area, selection buttons, and a parameter panel, and is shown in Fig. 7. Pressing the buttons aligns the visualization as depicted in Table 2. The dependency parameters, which have been discussed above, are calculated in the background and can be visualized individually using the interactive parameter panel (see Fig. 7, highlighted area). They provide a powerful tool to make the whole system structure better understandable using color-coded visualizations. Per default, a 3-dimensional Sankey diagram is shown. A 2-dimensional representation can be selected as well. In the center of this visualization is the focused category, which is surrounded by a given number of subcategories. Those are linked to the focused main group. In our example, the main group is represented by users (compare to Table 1), which are surrounded by the categories disciplines, tasks, and levels (compare to Fig. 2, Fig. 3, and Fig. 4).

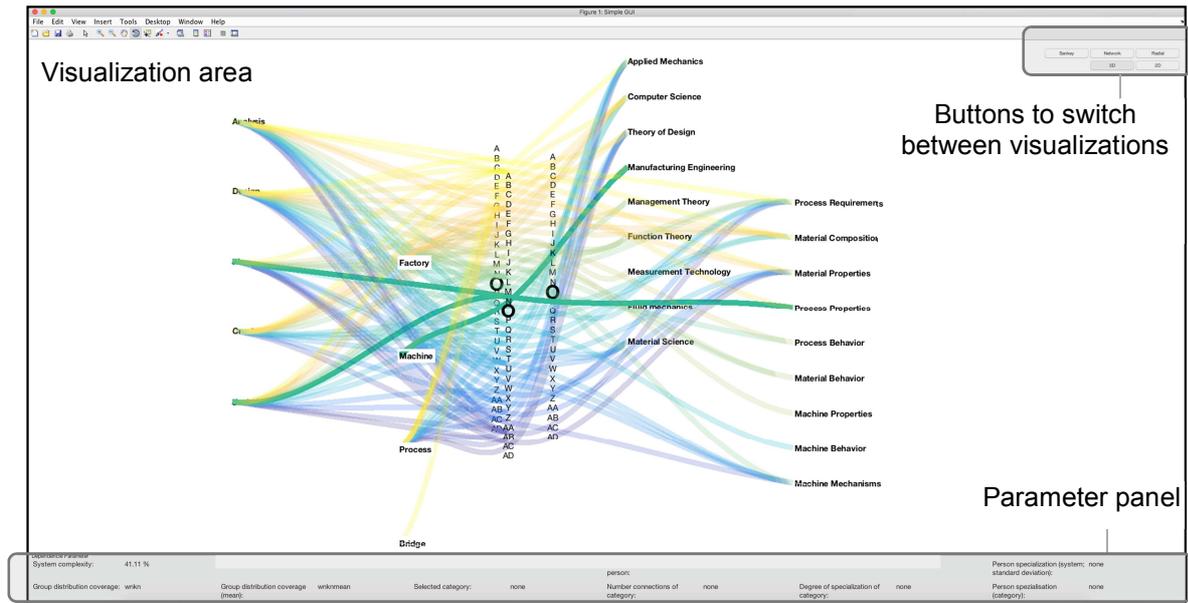


Figure 7 – The user interface of the application comprises mainly three components: the parameter panel, buttons to switch between different visualizations, and the visualization area. By initialization the dependency parameters are calculated and a 3-D Sankey diagram is plotted. Highlighted in the visualization are the connections of the selected user O.

Both the 3-dimensional and the 2-dimensional Sankey diagrams are interactively connected with the dependency parameters. Selecting one user or one category entity by clicking on the label highlights the label and all connected edges. The colors are defined per user evenly distributed over the complete spectrum of the parula-colormap, which gives a wide range of colors and is provided by Matlab libraries. Clicking on one of the defined dependency parameters results in a colorization of the Sankey diagram according to the calculated parameter, as shown in Fig. 8. The selected and highlighted labels and connections are retained unchanged when switching between the Sankey visualizations. In Fig. 8 the individualized parameter $v_{n,p}$ is displayed exemplarily, which gives the applicator a feedback on the users' specialization regarding disciplines, tasks and levels. The degree of the specialization of a user regarding one category is visualized with color-coded gradients from one connection (yellow) to the maximal number of existing connections (blue). While all users only have one connection to the other categories (system-levels and disciplines), users feature up to five connections to entities of the category “tasks”. The meaningful colorization makes it easy to see the number of existing connections and the degree of specialization of each user.

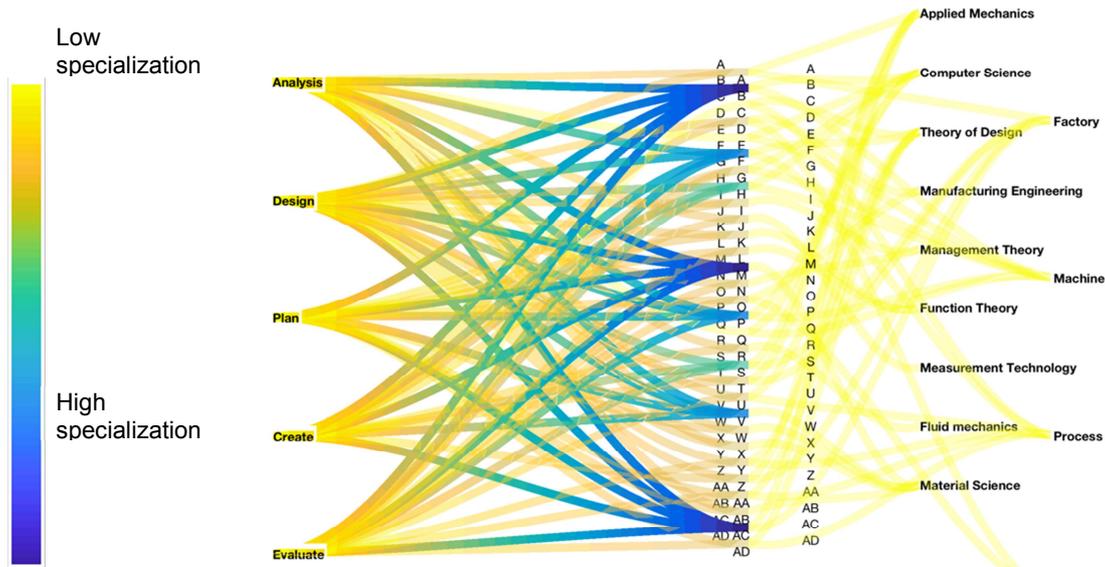
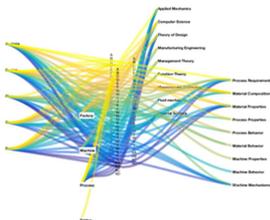


Figure 8 – Colorization of the diagram based on calculated specialization parameter $v_{n,p}$ according to Eq. 2. The visualization elucidated an inhomogeneous distribution of the group respective of the assigned scope of duties of the cooperating user. The color gradients from yellow to blue as well as the transparency elucidate the number of the existing connections.

Table 2 – Compilation of different visualization techniques, which have been implemented in an interactive, intuitive and easy to use graphical user interface, based on Matlab, D3.js, and Python.

Default view – 3D Sankey visualization

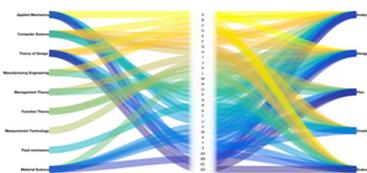


The 3D Sankey visualization shows all entities and connections in one view. By rotating the diagram the connections with all categories of one user can be explored. Any amount of categories can be visualized and be interactively explored.

Benefits & Shortfalls

- + All connections in one view
- + Visualization and comparing of individual parameters
- Might be overwhelming
- Impact depends on selected parameter

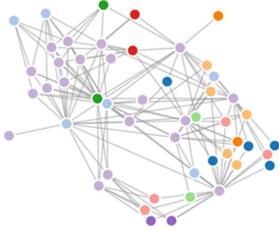
2D visualization of Sankey diagram



2D visualization sets focus on two categories. The links can be easily recognized and the user can observe the relevant data. User interdependencies are indirectly connected via categories. Entities and high-lighted parameters do not change when switching between visualizations.

Benefits & Shortfalls

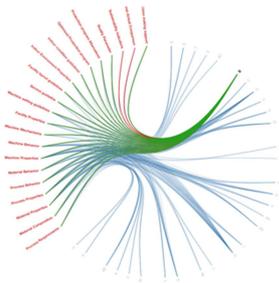
- + Selected features retain and focus remains
- + Only relevant data is visualized
- + Connections can be tracked
- No information on other connections
- No direct interdependencies between users

2D/ 3D graph visualization

Network visualizations are used to identify dependencies and connections at a glance. The highlevel visualization of connected nodes represents an alternative to the 3D graph visualizations. Entities with a very high or very low amount of connections can be easily recognized.

Benefits & Shortfalls

- + Can be combined with interaction capabilities
- + Densities are captured easily
- Details on the connections and entities are not recognizable
- No detailed information available

Radial visualization of information flow

Method is used to get insights into the information flow and connects users via existing information types. Input and output data of users are easily recognizable. Information flow and data transfer between provider and receiver can be highlighted interactively.

Benefits & Shortfalls

- + Clear visualization
- + Contains relevant information in multiple dimensions
- Degree of simplification in some cases might be to high
- No direct connections recognizable without user interaction

Conclusion

In the course of this publication, a clearly defined dependency graph of users in industrial corporations was designed and demonstrated. Based on the introduced user taxonomy, clearly defined entities and relationships of that network can be modeled. A set of dependency parameters were proposed, to display a meaningful relation between objects and to give an easy understandable overview of the whole relationship with the goal to solve complex tasks and improve a groups' performance. A case study based on the dataset of projects within the IRTG 2057 college was performed. In the course of this observation, initially independent projects were examined and overlapping was found. Consequently, dependencies between users were clearly identified and connected across the corresponding information-flows. An overview of all existing relationships within that group is visualized in a complex all-embracing diagram. During the work of this publication, it turned out, that this way of illustration is too complex for getting insights into a specific entity. But furthermore, the visualization is not detailed enough to gather any information about the calculated dependency parameters as defined above. Therefore, we used several state-of-the-art methods that give the possibility to visualize classified data and show relations, connections and interdependencies of group members that perform a common task. Those different visualization techniques were applied to make complex system architecture easily understandable for users. The representations are embedded in an interactive application, which allows users to explore the recognized dependencies and links. To support an efficient interaction with the model and easy information gathering, the application focuses the most important interdependencies by fading out irrelevant information and allows the user to switch between different visualizations without losing the focus.

In future work, the findings and gathered information of the investigation will be reused to model dedicated users and activities with a high level of detail. Furthermore, the discovered insights on the input and output data can be used to install a database system and identify interfaces between entities within the system architecture. Both the models and the architecture design will be used to improve the collaboration tool IN2CO for more efficient collaborative work.

Acknowledgements

This research was funded by the German research foundation (DFG) within the IRTG 2057 "Physical Modeling for Virtual Manufacturing Systems and Processes."

References

- [1] J. A. Bondy and U. S. R. Murty, Graph theory with applications, vol. 290, Citeseer, 1976.
- [2] G. v. d. Veer, O. Kulyk, D. Vyas, O. Kubbe and A. Ebert, "Task Modeling for Collaborative Authoring," in *Proceedings of the 29th Annual European Conference on Cognitive Ergonomics*, New York, NY, USA, 2011.
- [3] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, B. Van Riemsdijk and M. Sierhuis, "The fundamental principle of coactive design: Interdependence must shape autonomy," in *Coordination, organizations, institutions, and norms in agent systems VI*, Springer, 2011, pp. 172-191.
- [4] F.-A. Rupperecht, B. Hamann, C. Weidig, J. C. Aurich and A. Ebert, "IN2CO - A Visualization Framework for Intuitive Collaboration," in *EuroVis 2016 - Short Papers*, 2016.
- [5] M. Johnson, J. M. Bradshaw, P. J. Feltovich, R. R. Hoffman, C. Jonker, B. van Riemsdijk and M. Sierhuis, "Beyond cooperative robotics: The central role of interdependence in coactive design," *IEEE Intelligent Systems*, vol. 26, pp. 81-88, 2011.
- [6] T. W. Malone and K. Crowston, "The interdisciplinary study of coordination," *ACM Computing Surveys (CSUR)*, vol. 26, pp. 87-119, 1994.
- [7] G. Klein, P. J. Feltovich, J. M. Bradshaw and D. D. Woods, "Common ground and coordination in joint activity," *Organizational simulation*, vol. 53, 2005.
- [8] H. Patel, M. Pettitt and J. R. Wilson, "Factors of collaborative working: A framework for a collaboration model," *Applied ergonomics*, vol. 43, pp. 1-26, 2012.
- [9] T. B. Sheridan, "Teleoperation, telerobotics and telepresence: A progress report," *Control Engineering Practice*, vol. 3, pp. 205-214, 1995.
- [10] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, V. a. B. M. Riemsdijk and M. Sierhuis, "Coactive design: Designing support for interdependence in joint activity," *Journal of Human-Robot Interaction*, 3 (1), 2014, 2014.
- [11] L. Bunch, M. Breedy, J. M. Bradshaw, M. Carvalho, D. Danks and N. Suri, "Flexible automated monitoring and notification for complex processes using KARMEN," in *Proceedings. 2005 IEEE Networking, Sensing and Control, 2005.*, 2005.
- [12] T. M. J. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, pp. 1129-1164, 1991.
- [13] D. Holten and V. a. J. J. Wijk, "Force-Directed Edge Bundling for Graph Visualization," in *Computer graphics forum*, 2009.
- [14] D. Egorov and A. Bezgodov, "Improved Force-Directed Method of Graph Layout Generation with Adaptive Step Length," *Procedia Computer Science*, vol. 66, pp. 689-696, 2015.
- [15] J. Hua, M. L. Huang and Q. V. Nguyen, "Drawing large weighted graphs using clustered force-directed algorithm," in *Information Visualisation (IV), 2014 18th International Conference on*, 2014.
- [16] W. Huang, P. Eades, S.-H. Hong and H. B.-L. Duh, "Effects of curves on graph perception," in *Pacific Visualization Symposium (PacificVis), 2016 IEEE*, 2016.
- [17] P. Riehmann, M. Hanfler and B. Froehlich, "Interactive sankey diagrams," in *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, 2005.
- [18] E. R. Tufte, "The visual display of quantitative information.," *Journal for Healthcare Quality*, vol. 7, p. 15, 1985.
- [19] M. Schmidt, "The Sankey diagram in energy and material flow management," *Journal of industrial ecology*, vol. 12, pp. 82-94, 2008.

-
- [20] G. Hu, X. Ou, Q. Zhang and V. J. Karplus, "Analysis on energy–water nexus by Sankey diagram: the case of Beijing," *Desalination and Water Treatment*, vol. 51, pp. 4183-4193, 2013.
- [21] C. Deng, H. Li and Y. Shao, "Research on the drawing method of energy Sankey Diagram based on Java," in *Advanced Communication Technology (ICACT), 2014 16th International Conference on*, 2014.
- [22] G. M. Draper, Y. Livnat and R. F. Riesenfeld, "A survey of radial methods for information visualization," *IEEE transactions on visualization and computer graphics*, vol. 15, pp. 759-776, 2009.
- [23] C. Tominski, H. Schumann, G. Andrienko and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Transactions on visualization and Computer Graphics*, vol. 18, pp. 2565-2574, 2012.
- [24] D. Holten, "Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data," *IEEE Transactions on visualization and computer graphics*, vol. 12, pp. 741-748, 2006.
- [25] M. Beham, W. Herzner, M. E. Gröller and J. Kehler, "Cupid: Cluster-based exploration of geometry generators with parallel coordinates and radial trees," *IEEE transactions on visualization and computer graphics*, vol. 20, pp. 1693-1702, 2014.
- [26] B. Kirsch, "IRTG 2057 Graduate program," 2014. [Online]. Available: <http://www.irtg2057.de/>.
- [27] J. Weber, "Überlegungen zu einer theoretischen Fundierung der Logistik in der Betriebswirtschaftslehre," *Beiträge zu einer Theorie der Logistik*, pp. 43-65, 2008.
- [28] D. Piekenbrock and A. Hennig, *Einführung in die Volkswirtschaftslehre und Mikroökonomie*, Springer-Verlag, 2012.
- [29] K. Küpfmüller, W. Mathis and A. Reibiger, *Theoretische Elektrotechnik: Eine Einführung*, Springer-Verlag, 2013.