

Visualizing Social Documents as Traces of Collaborative Activity in Enterprise Collaboration Platforms

Julian Mosen
 University of Koblenz-Landau
 Institute for Information
 Systems Research
 Faculty of Computer Science
 Germany
julianmosen@uni-koblenz.de

Susan P. Williams
 University of Koblenz-Landau
 Institute for Information
 Systems Research
 Faculty of Computer Science
 Germany
williams@uni-koblenz.de

Petra Schubert
 University of Koblenz-Landau
 Institute for Information
 Systems Research
 Faculty of Computer Science
 Germany
schubert@uni-koblenz.de

Abstract

Enterprise collaboration platforms are large-scale information infrastructures that provide a wide range of tools and functionality to support collaborative work in organizations. These collaborative activities leave digital traces in the form of social documents, which can be analyzed to understand how employees work together to coordinate their joint work. In this paper, we present the findings of a research project to visualize the structure of social documents to prepare them for analysis as traces of collaborative activity. Using the representation of social documents defined in the Social Document Ontology (SocDOnt), we draw on concepts from graph theory to develop a method for the graphical visualization of social documents. Applying this method to analyze the social documents in an operational enterprise collaboration platform, we identify and display different types of social documents and define their characteristic structure. Our findings provide the necessary foundation for conducting computational ethnographies of collaborative work.

1. Introduction and Motivation

Enterprise collaboration platforms, such as IBM Connections and Jive, are complex, large-scale information infrastructures, implemented in organizations to provide an integrated platform to support employee collaboration and the coordination of digital work [16]. Spanning multiple global regions, business divisions and workgroups, enterprise collaboration platforms support the work of many thousands of employees, often widely dispersed in both space and time [31]. Rich in enterprise social software functionality (e.g. wikis, blogs, social profiles, activity streams, likes, tags) [23] enterprise collaboration platforms have become a core

platform for the digital workplace [31]. However, to date there are few studies that provide a holistic investigation of enterprise collaboration platforms to understand how they are being used by organizations and their employees to support collaboration and the coordination of work. There are two potential reasons for this lack of progress.

First, to conduct such studies and to understand how employees are using such systems to support collaborative work requires access to real-world data from operational systems. Gaining access to such systems can be problematic as they are closed systems, privately owned by organizations and installed behind firewalls. Through our long-term research program we have such access through both the current and historical data generated in our own operational collaboration platform (UniConnect) and also from extensive data provided by the companies participating in our university-industry research program [30].

Second, enterprise collaboration platforms potentially contain millions of digital artifacts (social content), created to support the collaborative and coordinative work of, hundreds, often thousands of employees. To analyze these large volumes of data requires both an understanding of the semantic structure of the social content within the system and appropriate methods to identify and visualize collaborative activity.

Enterprise collaboration platforms provide many ways for people to work together to collaborate and coordinate their joint work. These collaborative activities leave digital traces in the collaboration platform in the form of *social documents* [11], which are comprised of items such as blog posts, wiki articles, forum topics, likes, tags and comments, that are created and enhanced as people collaborate on joint work. In recent years there has been renewed interest in the study of digital documents to understand the ways they mediate interaction, communication and collaboration between people and with technologies [11, 12, 17, 19, 21, 25, 32].

Research to examine documents and documentary practices has a long history in the form of ethnographic studies of work [10, 20]. However, whilst these studies contribute significantly to our understanding of technology-mediated collaboration and the coordination of work they have a number of limitations for our proposed work on enterprise collaboration platforms. First, in terms of scope, they are often focused on single-site implementations, in specific locations within relatively narrow timeframes [8, 18]. Second, they apply ethnographic methods of co-presence and participant-observation [2] to study specific document types and activities [14, 27]. Whilst these methods provide very detailed accounts of collaborative work they are not suited, or perhaps better stated, not intended to examine the evolving interactions between large numbers of participants taking place across large-scale information infrastructures where the scale and scope of the problem space is much larger.

Our research to investigate collaborative activity in large-scale distributed enterprise collaboration platforms requires us to examine collaborative activity at both the micro-level of the individual documents and tasks as well as at larger scales across working groups and the platform as a whole; in order to follow the collective work practices of potentially thousands of users as they use a diverse range of tools and functionality to support their collaborative work. In addition, we are interested in understanding the ways work practices are inscribed and how social documents and collaboration platforms evolve over time. To achieve this requires alternative ethnographic approaches and methods that accommodate these conditions of both scale and scope. For this we turn to the emerging field of computational ethnography [1, 3] which allows for the study of larger scale infrastructures [1]; enables the study of everyday practices in information spaces [6] going beyond formal accounts to interrogate what people do [13] and accommodating the digital traces of both human and non-human actors. Of particular promise for our research examining collaborative work and work practices in enterprise collaboration systems is Geiger and Ribes [8] work on *trace ethnography* “that exploits the proliferation of documents and documentary traces” in distributed large-scale technology-mediated systems [8]. Trace ethnography uses the documents and digital traces (such as transaction logs, conversation transcripts, version histories) laid down in sociotechnical environments to “provide rich insights into the interactions of users, allowing us to retroactively reconstruct specific actions at a fine level of granularity” and “allowing researchers to carefully follow coordination practices, information flows, situated routines, and other social and organizational phenomena across a variety of scale.” [8:1].

Social collaboration analytics [23] is a growing field, however to date work is largely focused on social network analysis to identify, for example, who is working with whom and how social networks form [24]. Whilst important in understanding collaborations between people, these studies do not tell us anything about the ways that people collaborate, the artefacts they use and the new work practices that evolve as employees work together over time. Our focus is not on the social networks that form (though this is a part of the work) but on how people work together, and how they use the functionality and affordances of collaboration technologies to support their work and develop new work practices. To do so, our goal is to examine the traces of collaborative activity inscribed in social documents.

To achieve this goal first requires an understanding of the structure and nature of social documents. In our previous work [29] we conducted an in-depth analysis of the structure of social documents and derived a generic model (the Social Document Ontology – SocDOnt) to describe their structure. In this paper, we build on this foundational work on the semantic structure of social documents to develop methods for their visualization and analysis in order to better understand collaborative activity in enterprise collaboration platforms.

The remainder of this paper is structured as follows. In section 2, we present the Social Document Ontology and the foundational concepts related to the structure of social documents. Our research process is presented in section 3, which describes the preliminary findings of an empirical study to visualize social documents in an enterprise collaboration platform and its contribution to the final visualization. Based on the data from UniConnect, an enterprise collaboration platform used in practice, section 4 demonstrates the visualization of social documents and presents our findings on their typical structures for group workspace, containers and collections. Finally, section 5 contains conclusions and an outlook on future research.

2. The Structure of Social Documents: Key Concepts and Terminology

Social documents as digital traces of collaborative activity are complex, compound documents created by the use of social software and composed of heterogeneous components [11]. For the investigation of their structure, Hausmann and Williams [11] derived a *conceptual information model* and provided a first structural description of their nature and a list of possible components, such as versions, comments, attachments, tags

and likes. Focusing on a more detailed and technical description, the *Social Document Ontology* (SocDOnt) provides an ontology for the generic modelling of social documents and their structure on both, a micro-level (composition) and a macro-level (relations between and aggregations of social documents) [29]. As SocDOnt is based on and aligned with well-established ontologies from the field of web science and semantic web, it makes use of standardized terminology and existing concepts and provides the theoretical foundation for the graph representation and visualization of social documents presented in this paper. As these terms and concepts are a prerequisite for the understanding of the research process and the findings described in section 3 and 4, the key concepts from SocDOnt are introduced in the following.

Drawing on prior research, SocDOnt describes a social document as composition of single *items*, and is a single piece of user-generated content, such as a microblog post or a comment. The initial item of each social document is the *intellectual entity* and represents the core element, which is the center of collaboration on a micro-level. A good example for the intellectual entity is a wiki article, which is the core of collaborative activities, such as subsequent edits, changes or comments. Items that are created in the context of an intellectual entity and contribute any kind of meaning or content are *components* of a social document. Depending on their complexity, SocDOnt distinguishes between *simple components* (e.g. tags and likes) are non-intellectual items, containing a small amount of information or a meaning, intellectual components (e.g. comments and attachments) have an intellectual property and generally contribute more complex information.

Containers serve as a storage location for social documents and their components. For each type of intellectual entity there is a certain type of container (e.g. a forum or a wiki), where it is created and stored. Depending on the content type, sub containers (e.g. folders) are used for further structuring. *Spaces* describe the workspace and location, where containers and social documents reside and social software features are provided for their creation and manipulation. Typical spaces in enterprise collaboration platforms are group workspaces and individual user workspaces.

Social Documents that are associated to each other by an intellectual connection form a common *collection*. In contrast to containers, which are concrete instances that are created automatically or manually before a social document can be created, collections arise implicitly over time by social documents linking and referencing each other, having a parent-child relationship or addressing the same matter of fact. While a social document is stored within exactly one container, it can be a

part of multiple collections at the same time. Thus, the identification of collections, which a social document is part of, can be a challenging task.

3. Research Approach and Data

The aim of this paper is to draw on previous research on the structure of social documents to i) develop a method for visualizing social documents and ii) to identify and display different types of social documents and define their characteristic structure. The work is part of a wider research project to understand the structures of social documents and to use them to analyze collaborative activity. The overarching research project is organized in five research phases guided by Design Science Research (DSR) [26]. This study on the visualization of social documents and their structure is based on four research steps derived from the CRISP-DM approach for data mining [4]. Figure 1 shows the research approach. The first column shows the activities of the wider research project and the highlighted items in the other two columns show the steps and artefacts described in this paper.

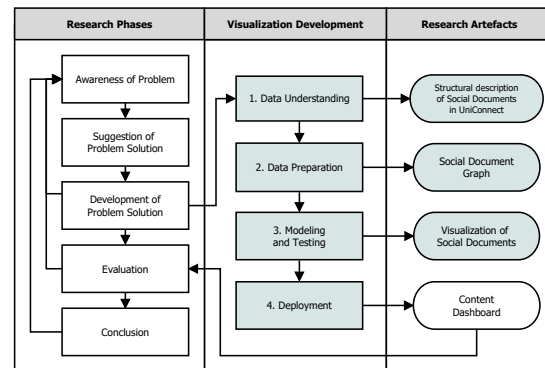


Figure 1. Research approach

Research Step 1: Data understanding. In the first step we examined the inherent structure of the social documents contained within *UniConnect*, a large-scale enterprise collaboration platform. *UniConnect* is an academic collaboration platform with 3500 users and more than 1200 group workspaces (communities) and is based on the commercial collaboration software *IBM Connections*. IBM Connections is currently one of the largest integrated collaboration platforms on the market [9]. For our examination, we had full access to all data in UniConnect and investigated social documents from the perspective of all architectural layers: presentation, application and database.

On the *presentation layer*, we examined the user interface and derived the terminology used within

UniConnect for the description of social documents (intellectual entities and components), containers and spaces and mapped these terms to the corresponding concepts from SocDOnt [29]. Table 1 shows the mapping of concrete classes from SocDOnt to the types of spaces and containers available in UniConnect. A more detailed mapping of intellectual entities to their corresponding containers is included in Table 2.

Analysis of the *application layer* enabled us to identify all types of social documents and containers and to derive a list of functionalities available for the generation and manipulation of intellectual entities and components.

Table 1. Mapping of concepts defined in SocDOnt to concepts from UniConnect

High-level concepts <SocDOnt>	Concrete classes <SocDOnt>	Terminology of UniConnect
Space	Organisational Platform	Platform
	Group Workspace	Community
	User Workspace	User Profile
Container	Folder	Files and Folders
	Message Board	Forum
	Microblog	Status Updates
	Task Container	Activity
	Weblog	Blog
	Wiki	Wiki

To achieve this, we designed an examination tool (shown in Table 2), that uses concepts from SocDOnt [29] and is guided by the first three dimensions of the *Social Collaboration Analytics Framework* (SCA) [23]. For each module in UniConnect, we identified the *space* (SCA: *where*) and the *container* (SCA: *content type*) in which a user can create *items* (SCA: *content components*). The result of this investigation is shown in Table 2. Illustrated by the example of a blog post, it can be read as follows: A blog post can only be created within a blog, which cannot have subordinated blogs (sub containers), but exists within the scope (space) of the platform (organizational workspace) or a community (group workspace); changes to a blog post are not tracked by versions; a blog post cannot have attachments but it can have comments, which can be also commented on (threaded) and it can have likes (recommendations) and tags.

Within UniConnect we identified three types of spaces, six types of containers, three types of sub containers and seven types of intellectual entities. While top containers are created only once and automatically during the initialization of group and user workspaces, sub containers are always created manually by the user. Only a few top containers (blog, forum, wiki), which can exist outside of group and user workspaces, can be created manually and multiple times in the space of the platform. Within one container type, a social document can be composed of up to six different item types (intellectual entity, version, attachment, comment, like and tag). The single rows in Table 2 show that some types of social documents can have similar compositions (e.g. forum topics and status updates), but there is no consistency regarding the potential components of social documents across all containers. The versioning feature is limited to files and wiki pages; tasks and entries do not support recommendations; blog posts and files cannot have attachments; a comment of a file or a wiki page cannot be commented itself. A possible explanation of this inconsistency is that there is no requirement for some features (e.g. why would we need to attach an attachment to a file?). We also identified inconsistencies for the storing of attachments. While attachments of status updates are always stored within the top container “files” (and never in a sub container “folder”) of a user workspace or group workspace, attachments of tasks, entries, forum topics and wiki pages are stored within the container of their intellectual entity. This leads to the situation that only attachments of status updates can have versions, be commented and liked. It is likely that these inconsistencies are a consequence of the integration of third party applications as modules in the underlying collaboration software IBM Connections. The analysis revealed five types of social documents with different structural characteristics (c.f. dots in Table 2). Forum topics and status updates have identical characteristics.

On the *application layer*, we did not find any concrete forms of collections, but there are some features that implicitly point to their existence; tasks and entries can be nested hierarchically within their container (activity) by having subtasks and subentries, wiki pages can be extended by child wiki pages. Some of the intellectual entities listed in Table 2 (column “type” having an asterisk (*)) can have attachments, which can result in an association of two social documents, where one is a file. These relationships are indicators for the identification of collections, which we will use later in section 4.

On the *database layer*, we investigated the technical implementation of the social documents by examining how single components are stored in the relational databases of UniConnect. UniConnect is based on a proprietary, closed source software product and to gain access

Table 2. Examination tool for the structure of social documents applied to UniConnect

Space	Container		Social document items					
			Intellectual entity		Intellectual component		Simple component	
	Top container	Sub container	Type	Version	Attachment	Comment	Like	Tag
Community, User Profile	Activities	Activity	Task*, Entry*		•	•*		•
Platform, Community	Blog		Blog Post			•*	•	•
Platform, Community	Forum	Forum**	Forum Topic		•	•*	•	•
Community, User Profile	Files	Folder*	File	•		•	•	•
Community, User Profile	Status Updates		Status Update		•	•*	•	•
Platform, Community	Wiki		Wiki Page*	•	•	•	•	•

* can be nested ; ** within a community only

to its architecture we applied reverse engineering techniques. We analyzed the database schema of UniConnect and identified the relevant databases, tables, fields and their relationships. Guided by the classification of data sources for Social Collaboration Analytics from [23], we looked at *transactional data* (1 database), user-generated *content data* (6 databases) and *organizational data* (2 databases).

Our analysis of the content data identified that there is exactly one database for each type of container storing social documents. During the analysis, we found similarities in the architecture of these databases, which underlines our previous observation in the application layer that some modules offer similar or different features for the creation of social content. For example, the architectures of the databases storing files and wikis are the same, which means that the corresponding modules provide similar features for the creation and combination of social content. The same holds for the modules and databases of activities and forums. For each of the content databases we developed an entity-relationship diagram describing which tables are used to store containers, the single components of a social document and their relationships. The investigation of organizational databases was necessary to identify the spaces in which the containers and social documents are created. The most important result of our database investigation was the understanding of where the components of social documents are technically stored. Thus, we were able to build the necessary database queries to extract the social documents and their structure for the data preparation, described in the following.

Research Step 2: Data preparation. The aim of our second research step was the identification of the relevant data stored within the content databases of

UniConnect and its transformation into a manageable format. For the data extraction we built several database queries to collect the containers, intellectual entities and components of each social document for a specified space. By executing the database queries, we received up to 39 result sets including one set for each item type per top container type and three types of sub containers (6 item types x 6 top container types + 3 types of sub containers= 36 result sets). These result sets contain all attributes and relations of the social content items and the containers for the given space. Next, we transformed the result sets into a comprehensive and uniform graph representation, which was guided by the Social Document Ontology (SocDOnt). For this, we define the *social document graph* as a directed graph $G = (V, E)$ with a set of vertices (nodes) described by $V(G) = \{v_1, v_2, \dots, v_n\}$ and a set of edges described by $E \subseteq V \times V$. Each vertex represents a social document item (intellectual entity, intellectual component or simple component) or a sub container. We excluded top containers in the set of vertices, because they are commonly created automatically by the system and not explicitly by the user. Each edge of the social document graph represents an association of two items being a component, a child or a reference of each other. For a more precise modelling, all edges are directed and thereby indicate the direction of an association. The direction of an association has an important meaning and allows, for example, to show which comment responds to which comment. Whilst we were able to extract all associations from the data representing components and children, identifying and extracting the references is a challenging task. Most references are not stored within separate fields of the database tables but are contained within the

item's content as hyperlinks. As their extraction is more complex and requires more computational power, references have only been extracted for some items (status update attachments referencing files). Given that a graph is *connected*, there is a path between every pair of vertices [5, 28]. A *connected component* of a graph is a subset of the graph's vertices "such that (1) every node in the subset has a path to every other and (2) the subset is not part of some larger set with the property that every node can reach every other" [5:26]. Considering the different kinds of edges (component of, child item of, references) the concept of connected components is well suited for the description of social documents, containers and collections within the social document graph, which will be further elaborated in the following.

Research Step 3: Modeling and Testing. After preparing the data and transforming it into the social document graph, a graphical representation for its visualization was developed. The main goal was to provide a graphical representation that allows us to comprehend the composition of real social documents within UniConnect and to identify their characteristic structures. For each type of social document item (vertices of the graph) and association (edges) we specified certain symbols and different types of arrows, which are introduced and illustrated in the following example of a wiki.

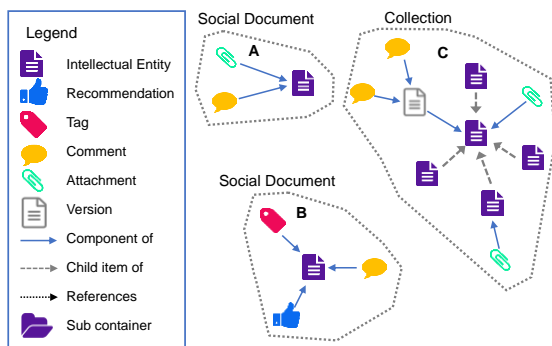


Figure 2. Graphical representation of the social document graph

Figure 2 represents an extract of a wiki containing seven wiki pages (intellectual entities), one previous version of a wiki page (version), three attachments, four comments, one tag and one recommendation. While the solid lined arrows describe that an item is a component of another (e.g. an attachment is a component of a wiki page), the dashed lined arrows represent that an item is a child of another (e.g. a wiki page being a child of another wiki page). Dotted lined arrows indicate that items are referencing each other (e.g. via hyperlinks). The arrowhead of each line indicates the direction of an association. While the sections A and B each represent one

wiki page with their components, section C contains five wiki pages that are related to each other. The wiki page in the center of section C has one attachment and a previous version, which has been commented twice. Beside of these components, the centered wiki page has four sub wiki pages. The social document graph in Figure 2 contains three connected components, which are highlighted for illustration by dotted lines. While the sections A and B both represent one compound social document, the section C contains five social documents (wiki pages). The fact, that all four wiki pages are sub-pages of the fifth, aggregates these social documents to one common collection.

During this phase, a prototype for visualizing real data sets from UniConnect was developed, evaluated and refined.

Research Step 4: Deployment. The findings from the data preparation and modelling were then used for the implementation of the *Content Dashboard*, a prototype application, which extracts, visualizes and analyses the social documents of a specified space from UniConnect in real time. The Content Dashboard was deployed to UniConnect as an integrated application and serves our research by visualizing the social documents of group workspaces. For drawing the social document graph, the application uses a *force-directed layout algorithm*, which is well suited for visualizing connected components [7] and results in a human readable arrangement of the graph. In detail, we used an algorithm that is based on *ForceAtlas2* described in Jacomy et al. [15]. As a result, we now have a tool for the automated visualization of social documents stored in UniConnect. The Content Dashboard has been through several cycles of evaluation in workshops with practitioners and has also been tested on live data within an organizational context.

4. Visualizing Social Document Structures

This section presents the findings we derived from visualizing social documents and their structure within a real group workspace (community) from the collaboration platform UniConnect. All the following illustrations are based on real data and were derived from the prototype application described in section 3.

For the investigation of the structural characteristics of all types of social documents, we selected a group workspace that makes use of all applications (modules) of UniConnect and contains a large amount of content. Thus, we examined a long-term group workspace, which is actively used by researchers and practitioners for collaboration and communication in the context of a

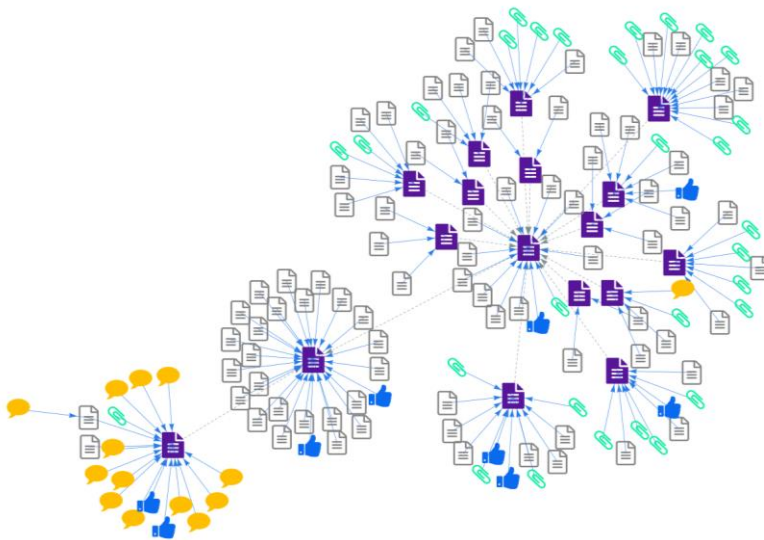


Figure 3. Intra-container collection: subordinated wiki pages and their components

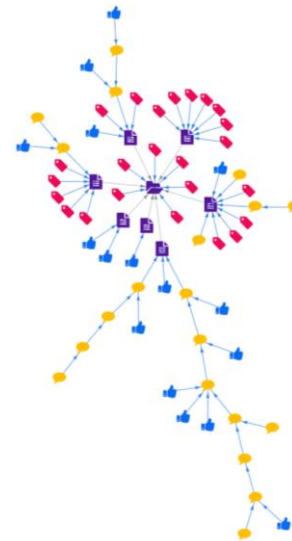


Figure 4. Forum with forum topics and threaded comments

collaboration project [30] for more than 5 years and currently has 86 members.

Micro-level: Social Documents. On the micro-level of the visualization, there are certain structural characteristics of social documents that depend on their type of intellectual entity. Some of these characteristics will be described in the following two examples.

Figure 3 shows a typical part of the wiki, which has been created within the group workspace. The intellectual entity furthest to the left (solid document symbol) represents the recent version of a wiki page, which has 2 previous versions, 13 comments, 2 recommendations and 1 attachment. As described in Table 2, comments of wiki pages cannot be nested and result in the typical circular arrangement of comments around a wiki page shown in Figure 3. In addition, the visualization clearly shows that there is one comment on a previous version of the wiki page. During the investigation of social documents from the presentation layer, we found that the user interface always presents the components of a certain version as if they were assigned to the latest version. Considering that the content of a wiki page can change over time, the lack of information about which version has been commented might lead to misunderstandings and it might be important to emphasize that the comment refers to something that has already been corrected in a recent version. While the front-end of UniConnect does not distinguish between components belonging to different versions, the visualization can be used to show the distinct association of components to their original version.

Looking at the forum topics of the group workspace, the visualization reveals a different structural characteristic for their social documents from that of wiki pages. Regarding our examination of Table 2, forum topics cannot have versions, but their comments can be nested. Figure 4 shows one forum (sub container) contained in the group workspace, having seven forum topics (initial forum posts). The lowest forum topic (document symbol) is the intellectual entity of a social document, which contains several comments and recommendations. In detail, the forum topic has been commented twice and recommended once. Both comments have been commented on their own and describe the root of two threaded discussions, which can become arbitrarily long and complex. While the visualization of this threaded discussion makes it easy to comprehend which comment is a response to which comment, this can be a challenging task in the user interface. These threaded comments are one of the characteristics for social documents that have a forum topic as their intellectual entity.

Macro-Level: Spaces, Containers and Collections. On a higher level, the visualization of social documents allows us to look at typical structures for spaces, containers and collections.

The typical *structure of a space* is presented in Figure 5 and shows all social documents contained in the group workspace introduced at the beginning of this section. The visualization shows that there are both, highly structured, complex social documents, that have many components, (mostly in the center of the figure) and simple social documents that consist of one or few items.

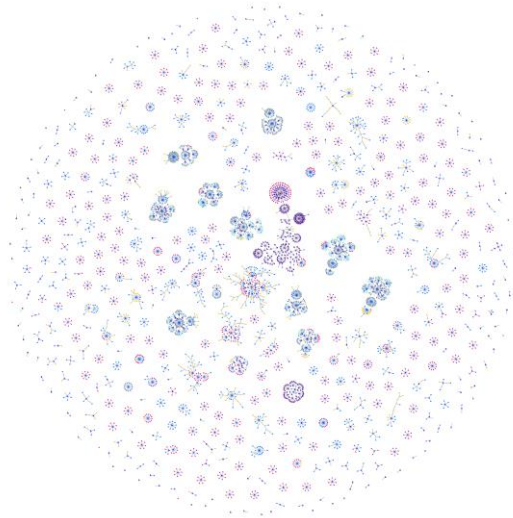


Figure 5. Visualization of all social documents within a group workspace of UniConnect

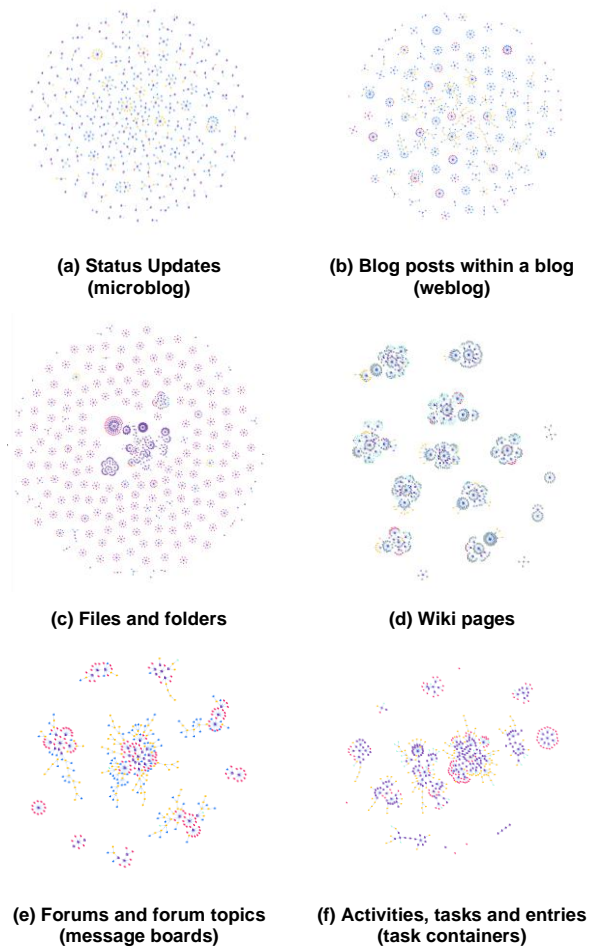


Figure 6. Group workspace displayed in Figure 5 broken down into top containers

The typical *structures of containers*, such as activities, forums, blogs, status updates, files and wikis, are shown in Figure 6. The subfigures of Figure 6 illustrate the same social documents that are contained in Figure 5 sorted by their type of top containers and respectively by the modules the content was created with. The status updates (microblog posts) shown in Figure 6a are the least complex social documents and have fewer components compared to other content types. Blog posts, shown in Figure 6b are more complex than status updates (microblog posts) and tend to have more components, especially more comments. Files (Figure 6c) can be grouped by folders and can have versions, which leads to a more complex structure and more components. Wikis and their contained wiki pages, (Figure 6d), can be highly structured with a high number of versions and are typically arranged hierarchically. Forums (message boards) and activities (tasks containers) are shown in Figure 6e and Figure 6f and typically have the most complex structure. While forum topics contain a high number of comments, which form long branches of threaded discussions, activities are structured by a high number of tasks, entries, subtasks and subentries.

As described in section 3, the types of components a social document is composed of depend on its type of intellectual entity, for instance, a blog post cannot have attachments, because IBM Connections does not offer this functionality. It is important to consider these differences when comparing social documents from different containers and for identifying collections.

In terms of collections, we identified two typical *structures of collections* with the visualization of the specified group workspace. One example of a collection is shown in Figure 3. The wiki page on the left side of the figure is connected to its parent wiki page by a dashed arrow. This parent wiki page has several previous versions, two recommendations and a second subpage, which is in the center of the right side of the figure and has several subpages itself. Assuming that a subpage is created by a user because it relates to its parent page, the resulting relationship represents an intellectual connection between both pages. Thus, both pages are part of the same collection. In the example of subordinated wiki pages being a collection, all parts of the collection are in the same container (the wiki) and represent an *intra-container collection*.

Figure 7 provides a simple example for a collection that is spread *across* containers. The left side of the figure shows a status update published in the group workspace with four components (a recommendation, a comment, a tag and an attached file).

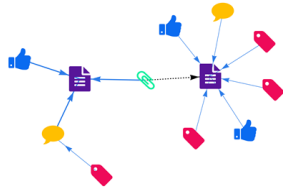


Figure 7. Inter-container collection: file (stored in file container) used as an attachment to a status update

The attachment of the status update is represented on the right side of the figure and has two recommendations, three tags and one comment. The dotted arrow directed from the attachment of the status update (paperclip) to the file (document symbol) indicates that the attachment of the status update is only a reference to an intellectual entity of another social document. In this case, the referenced intellectual entity is a file, which is stored in a different container to the status update. This example represents an *inter-container collection*. In general, we would assume that an attachment always leads to inter-container collections, but as we described in section 3, UniConnect has an inconsistency regarding the storage location (container) of attached files. If a file is *not* stored in the files container, UniConnect *does not* provide the social software features (commenting, recommending, tagging) for it, which finally leads to a file that is not really a social document [11]. In addition to inter-container collections, we also identified *inter-space collections*, which contain social documents that reference each other, but are located within different spaces. In UniConnect, typical inter-space collections arise if a user shares files, which are stored in the location of his personal user space and references these files within a social document of the group workspace. These inter-space collections require special attention, as they help to discover content that is stored outside a space.

It is an important insight that the inherent structure of different types of social documents leads to characteristic shapes. Their visualization is idiosyncratic for the different containers, to give an example, wiki pages have an inherently hierarchical organization and their visualization resembles the blossoms of flowers. Microblogs, as the simplest and least complex form of social documents, are characterized by speckled dots and look more like a Christmas bauble. As a consequence, experienced users of the Content Dashboard can identify the type of content by looking at the shape of its visualization.

5. Conclusion and Future Research

Considering social documents as digital traces of collaborative activities, this paper presents an approach

for visualizing these traces in UniConnect, an operational enterprise collaboration platform. The motivation for this deeper analysis of their structure and nature is to gain a better understanding of collaboration on multiple levels. To achieve this, required us to first understand and prepare the data obtained from UniConnect and to develop an approach for representing and visualizing the data. The visualization approach was then implemented as part of the Content Dashboard, which we use to examine concrete social document structures and which also provides a data source for further analysis, e.g. the study of typical use cases (as described in [22]). Such a visualization approach will also be useful for practitioners, enabling them to better understand and comprehend the usage of different features of their collaboration platform.

The development of both, the graph representation and the visualization of social documents, allowed us to analyze and understand the composition of hundreds of concrete instances of social documents at the same time and to identify characteristic structures for different types at a very large scale. We combined concepts from graph theory, allowing us to identify social documents and collections as connected components, and existing models, such as SocDont, which provided a theoretical description for the structure of social documents and presented a practical application of these abstract models. Thus, the main result of our work is a method and a format for visualizing social documents that allows us to look at real data from an integrated collaboration platform and identify typical structures on two levels.

On the *micro-level* of social documents, we examined typical structures related to the type of intellectual entity involved. On the *macro-level* we visualized containers and workspaces and examined the existence of *inter-container collections* and *inter-space collections* and provided preliminary examples. The ability to extract these structures is fundamental to conduct in-depth analysis of collaborative activities in future. The more accurate and comprehensive identification of collections will be part of our future research.

Inevitably, the structures of social documents we have identified to date are influenced and partly rely on the specific collaboration system's functionalities. The study of non-integrated collaboration systems and collaboration portfolios consisting of multiple standalone applications will be interesting for our future investigations of inter-container and inter-space collections.

6. References

- [1] Abramson, C.M., J. Joslyn, K.A. Rendle, S.B. Garrett, and D. Dohan, "The promises of computational ethnography: Improving transparency, replicability, and

- validity for realist approaches to ethnographic analysis”, *Ethnography* 19(2), 2018, pp. 254–284.
- [2] Beaulieu, A., “Research Note: From Co-location to Co-presence: Shifts in the Use of Ethnography for the Study of Knowledge”, *Social Studies of Science* 40(3), 2010, pp. 453–470.
- [3] Beaulieu, A., “Vectors for Fieldwork. Computational Thinking and New Modes of Ethnography”, In L. Hjorth, H. Horst, A. Galloway and G. Bell, eds., *The Routledge Companion to Digital Ethnography*. Routledge, New York, 2017.
- [4] Chapman, P., J. Clinton, R. Kerber, et al., *CRISP-DM 1.0: Step-by-step data mining guide*, CRISP-DM Consortium, 2000.
- [5] Easley, D., and J. Kleinberg, *Networks, Crowds, and Markets*, Cambridge University Press, Cambridge, UK, 2010.
- [6] Entwistle, J.M., H. Blunck, N.O. Bouvin, et al., “Computational Environmental Ethnography: Combining Collective Sensing and Ethnographic Inquiries to Advance Means for Reducing Environmental Footprints”, *Fourth International Conference on Future Energy Systems*, ACM Press (2013), 87–98.
- [7] Fruchterman, T.M.J., and E.M. Reingold, “Graph drawing by force-directed placement”, *Software: Practice and Experience* 21(11), 1991, pp. 1129–1164.
- [8] Geiger, R.S., and D. Ribes, “Trace ethnography: Following coordination through documentary practices”, *44th Hawaii International Conference on System Sciences*, (2011), 1–10.
- [9] Gotta, M., N. Drakos, and M. Jeffrey, *The Future of Social Software in the Workplace*, Gartner Research Report ID: G00293232, 2017.
- [10] Harper, R., *Inside the IMF: An ethnography of documents, technology, and organizational action*, Academic Press, San Diego, 1998.
- [11] Hausmann, V., and S.P. Williams, “Issues for the long-term management of Social Business Documents”, *International Journal of Information Systems and Project Management* 4(3), 2016, pp. 45–61.
- [12] Heath, C., H. Knoblauch, and P. Luff, “Technology and social interaction: the emergence of ‘workplace studies’”, *British Journal of Sociology* 51(2), 2000, pp. 299–320.
- [13] Hsu, W., “Digital Ethnography towards Augmented Empiricism: A New Methodological Framework”, *Journal of Digital Humanities* 3(1), 2014.
- [14] Jackson, P., and J. Klobas, “Deciding to use an enterprise wiki: the role of social institutions and scripts”, *Knowledge Management Research & Practice* 11, 2013, pp. 323–333.
- [15] Jacomy, M., T. Venturini, S. Heymann, and M. Bastian, “ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software”, *PLoS ONE* 9(6), 2014, pp. 1–12.
- [16] Leonardi, P.M., M. Huysman, and C. Steinfield, “Enterprise Social Media: Definition, History, and Prospects for the Study of Social Technologies in Organizations”, *Journal of Computer-Mediated Communication* 19(1), 2013, pp. 1–19.
- [17] Luff, P., J. Hindmarsh, and C. Heath, eds., *Workplace studies: Recovering workpractice and informing system design*, Cambridge University Press, Cambridge, UK, 2000.
- [18] Monteiro, E., N. Pollock, O. Hanseth, and R. Williams, “From Artefacts to Infrastructures”, *Computer Supported Cooperative Work* 22(4–6), 2013, pp. 575–607.
- [19] Olsen, B.I., N.W. Lund, G. Ellingsen, and G. Hartvigsen, “Document theory for the design of socio-technical systems”, *Journal of Documentation* 68(1), 2012, pp. 100–126.
- [20] Østerlund, C., “Documents in Place: Demarcating Places for Collaboration in Healthcare Settings”, *Computer Supported Cooperative Work (CSCW)* 17(2–3), 2008, pp. 195–225.
- [21] Østerlund, C., and K. Crowston, “Boundary-spanning documents in online communities”, *Thirty Second International Conference on Information Systems*, (2011), 1–10.
- [22] Schubert, P., J. Mosen, and F. Schwade, “Metrics for Analyzing Social Documents to Understand Joint Work”, *53rd Hawaii International Conference on System Sciences*, (2020).
- [23] Schwade, F., and P. Schubert, “Social Collaboration Analytics for Enterprise Collaboration Systems: Providing Business Intelligence on Collaboration Activities”, *50th Hawaii International Conference on System Sciences*, (2017), 401–410.
- [24] Stieglitz, S., K. Riemer, and C. Meske, “Hierarchy or Activity? The Role of Formal and Informal Influence in Eliciting Responses from Enterprise Social Networks”, *22nd European Conference on Information Systems*, (2014).
- [25] Trace, C.B., “Documenting work and working documents”, *44th Hawaii International Conference on System Sciences*, (2011), 1–10.
- [26] Vaishnavi, V., and B. Kuechler, “Design Science Research in Information Systems Overview of Design Science Research”, *Design Science Research in Information Systems and Technology*, 2004.
- [27] Ventres, W., S. Kooienga, N. Vuckovic, R. Marlin, P. Nygren, and V. Stewart, “Physicians, Patients, and the Electronic Health Record: An Ethnographic Analysis”, *Annals of Family Medicine* 4(2), 2006, pp. 124–131.
- [28] West, D.B., *Introduction to Graph Theory*, Prentice Hall, 2000.
- [29] Williams, S.P., J. Mosen, and P. Schubert, “The Structure of Social Documents”, *53rd Hawaii International Conference on System Sciences*, (2020).
- [30] Williams, S.P., and P. Schubert, “Connecting Industry: Building and Sustaining a Practice-based Research Community”, *50th Hawaii International Conference on System Sciences*, (2017), 5400–5409.
- [31] Williams, S.P., and P. Schubert, “Designs for the Digital Workplace”, *Procedia Computer Science*, Elsevier B.V. (2018), 478–485.
- [32] Zacklad, M., “Documentarisation processes in documents for action (DofA): The status of annotations and associated cooperation technologies”, *Computer Supported Cooperative Work* 15, 2006, pp. 205–228.