Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

Metrics for Analyzing Social Documents to Understand Joint Work

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

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

Florian Schwade
University of Koblenz-Landau
Institute for Information Systems Research
Faculty of Computer Science
Germany
fschwade@uni-koblenz.de

Abstract

Social Collaboration Analytics (SCA) aims at measuring and interpreting communication and joint work on collaboration platforms and is a relatively new topic in the discipline of Information Systems. Previous applications of SCA are largely based on transactional data (event logs). In this paper, we propose a novel approach for the examination of collaboration based on the structure of social documents. Guided by the ontology for social business documents (SocDOnt) we develop metrics to measure collaboration around documents that provide traces of collaborative activity. For the evaluation, we apply these metrics to a large-scale collaboration platform. The findings show that group workspaces that support the same use case are characterized by a similar richness of their social documents (i.e. the number of components and contributing authors). We also show typical differences in the "collaborativity" of functional modules (containers).

1. Introduction

The research presented in this paper is part of a long-term research program that has been following the implementation and adoption of enterprise collaboration platforms in user organizations for the past ten years [46]. Collaboration platforms support a wide range of work practices and have, in recent years, been enhanced with "social features" (see below) that facilitate new ways for people to work together, to share information and to collaborate on shared tasks. Common to all these work practices and tasks is that they are mediated by what have been defined as social (business) documents [12].

Whilst collaboration platforms provide a solution for the digital workplace through the digital support of communication and the codification of information and knowledge [32], the adoption of Enterprise Social Software (ESS) still proves challenging [9] and its actual use in organizations has, to date, not been widely examined.

Our study of social documents builds upon and extends research in the area of Social Computing, more specifically in the field of *Social Collaboration Analytics (SCA)* [35], which is concerned with measuring the use and the benefits achieved from using enterprise collaboration platforms. SCA is a newly emerging field that applies methods from the computer sciences to query the databases of collaboration software [35]. Its aim is to examine and better understand how collaboration software is actually used to support collaborative activity in organizations. As we will show in our literature review, studies that actually measure and interpret the use of collaboration platforms are still rare and the field of SCA lacks established frameworks, methods and terminology [36].

Most of the SCA studies available in the academic literature use *transactional data* (event logs) for the analysis of user activity. In this paper, we turn our attention to a different data source (*content data*), the "social content" that is created and enhanced with the help of "social features" as defined in the terminology framework for *Enterprise Social Software (ESS)* [35]. ESS is a software type that provides typical groupware functionality for the three classical Cs in CSCW, i.e. communication, collaboration and coordination [7]. *Social features* have their origin in (public) Social Media and include subscribing (following) information or people, commenting or tagging content as well as simple annotations such as recommendations or likes [34]. *Social Media Analytics* is the term suggested by Stieglitz et al. [42] for the study of these public, openly accessible platforms.

Previous studies based on data collected from practitioners have shown that industry needs better tools to measure and understand user activity. A survey among 24 collaboration professionals from leading user companies in German-speaking countries showed that

URI: https://hdl.handle.net/10125/63769
978-0-9981331-3-3

HICSS

"SCA is more than just counting and collecting numbers. More complex metrics that assess the cooperativity of ESS are necessary" [37]. In this paper, we will address this concern by providing, applying and interpreting metrics on social documents that help us understand how employees contribute to social content and in doing so jointly work on documents. It is our aim to provide *platform managers* (who are responsible for the entire collaboration infrastructure of an organization) and *workspace managers* (who are coordinating specific workgroups) with improved analytics tools to better understand user engagement in the digital workplace.

The paper is organized as follows: we begin with an examination of related work in the field of Social Collaboration Analytics to motivate the need to identify key concepts and provide theoretical and analytical grounding. This is followed by a detailed description of our method of analyzing the structure of social documents: the data sources, metrics and how these metrics were successfully applied to a large-scale, integrated collaboration platform. We conclude the paper with a discussion of our contribution and an outlook on future research.

2. Motivation and Research Design

Enterprise Social Software (ESS) is still a relatively new form of collaboration software in companies and its adoption and use are still under investigation [9]. Organizations are investing heavily in ESS [18] and there is a need to understand if and how users appropriate the new technology in their everyday work practices [31].

The literature on SCA distinguishes three possible data sources [35]: The first two accrue from the actual use of the system: (1) *transactional data* (event logs) are automatically recorded for each user activity and (2) *content data* that is created by the users (social documents). In addition to these, (3) *organizational data* (user profiles, roles, locations, etc.) can be used to group or filter SCA results.

An in-depth review of the literature in the field of SCA showed that the majority of studies use the first type, namely *transactional data*, for the analysis of collaboration platforms. In this paper, we analyze the second type (social content), and more specifically, the *structure of social documents*.

Many of the previous applications of social analytics focus on a *specific type of functionality* and most of them use data from public Social Media platforms such as Twitter [1, 15], Facebook [17] or Instagram [43]. Many of these platforms provide some form of API to their content and event logs and are thus suited (and encourage their use) as sources for data analytics.

The focus of our study, however, is on social software in *organizations*. Again, as for Social Media, the majority of studies in organizations focus on *specific functional modules* of ESS (e.g. only microblogs or only Wikis). We take a broader approach, which is focused on the analysis of large-scale integrated Enterprise Collaboration Systems (ECS) that include a range of different functional modules (blogs, microblogs, forums, wikis and tasks).

These applications are behind the firewalls of companies and only allow access for employees and trusted partners. Research on these systems requires the active collaboration and consent of the user companies, which might be one of the reasons that few studies exist. Examples are studies on *Enterprise Social Networks* for platforms such as Yammer/Communote [30] and Jive [26].

Research objectives. This paper addresses two research objectives as follows:

(1) to *develop metrics for SCA* based on the structure of social documents (as defined by SocDOnt).
(2) *to evaluate these SCA metrics* by applying them to real-world data and analyzing and interpreting the findings.

Research design. The research was conducted in three phases. *Phase 1* was an in-depth literature review following principles suggested by [44]. In *Phase 2*, metrics were developed based on the structure of social documents as traces of (joint) user activity. SocDOnt, an existing ontology for social documents [45], was used to develop metrics that span multiple containers for social documents. In *Phase 3*, we applied these metrics to a (purposefully selected) sample of 12 workspaces on a large-scale integrated collaboration platform with more than 3000 users representing more than 40 organizations. The findings were then interpreted to gain insights into the collaborative activity on this platform and to show how the metrics can be used to characterize specific forms of use (e.g. the joint work in projects).

3. Literature Review: Social Collaboration Analytics

A structured literature review on analytics in the area of Enterprise Social Software using the search terms and databases described in [36] retrieved 220 publications. After analyzing the abstract and scanning the content, 85 papers remained for detailed analysis. Among these, only 62 studies actually measure and present analytics results. Using the terminology framework by [35] we grouped these studies according

Table 1. Exemplary studies that apply analytics to social software

Category	Software type	Content type	Data sources	Refs
Enterprise Social Network (22)	MedNet (BW) (3) Yammer (11) "Inhouse development" (1) Unspecified „ESN System" (7)	Contact requests (6) Messages (14) Remark: most likely these are microblogs; no clear specification if 1:1 (chat) or 1:n (microblog)	Transactional (5) Content (13) Unclear (4)	[3, 29, 33]
Enterprise Collaboration System (18)	HP WaterCooler (2) MS SharePoint (3) Jive (3) IBM Connections (7) "Inhouse development" (2) Not specified (1)	Studies were not limited to specific content types. However, each content type was analyzed separately (→ localist study).	Transactional (8) Content (1) Transactional and content (1) Unclear (8)	[6, 22, 41]
Groupware (11)	BSCW (7) Lotus QuickPlace (1) Not specified (3)	Files (8) Chat messages (1) Folders (1) Blog posts (1) Poll posts (1) Not specified (3)	Transactional (8) Transactional and content (2) Unclear (1)	[5, 13, 24]
Enterprise Social Software (11)	Wiki (1) Conferencing system (1) Social bookmarking (2) File sharing (4) Jira (1) Not specified (2)	E-Mails (1) Instant Messages (1) Meetings (2) Blog posts (1) Bookmarks (2) Wiki pages (2) Discussion topics (1) Tags (1) Files (4) Tickets (1)	Transactional (9) Transactional and content (2)	[8, 23, 39]

to the software category that was under examination. Most publications (22, 35%) examine *Enterprise Social Networks* (microblogging and relationships between employees) followed by 18 publications (29%) that analyze data from integrated *Enterprise Collaboration Systems*. 11 studies (17%) investigate traditional groupware and the remaining 11 studies (17%) focus on *Enterprise Social Software* (portfolio applications).

Out of the 62 studies, only three suggest deriving metrics from *documents* [5, 8, 10]. Whilst most studies are applying existing metrics to data, only two studies develop *new* metrics [13, 28].

Table 1 provides an overview of the software categories, software types, content types and data sources identified in these studies. The last column contains some exemplary references.

The data source for SCA, referred to as *digital traces* or *digital records* [2, 11, 16], is of central importance for SCA. Behrendt et al. define digital traces as "digitally stored, event-based, chronological records of activities of actors, which result in direct or indirect actor relations or content in different data formats" [2:4]. Unfortunately, most of the SCA studies that we reviewed do not contain a precise definition of their source data, which prevents the use of the same setup in a similar (and potentially comparative) study.

In the studies on content data that we identified, authors use the terms *content* or *content data* [14, 27] and *communication data* [3, 29] to characterize their data sources. Thirteen (21%) studies do not contain a description of the data at all. Most studies (31, 50%) make use of transactional data, fourteen studies use content data for analysis. A closer inspection of this group revealed that eleven of the studies using content data examine Yammer, which does not provide transactional data so content data is the only available source. To circumvent this limitation, the authors of these studies reconstructed user actions from the available content data. Only five studies combine transactional data and content data. For most of the studies, the reader has to guess or infer the used data type.

Overall, the literature review showed that the majority of studies would benefit from a clear description of the software, data and methods of the analysis. We provide an example of such a description for our own SCA study in Table 2.

As previously outlined, only two studies develop *new metrics* for SCA, both of them address the concept of "collaborativity" (i.e. how intensely users work together) in a workspace. Jeners and Prinz [13] develop an *activity index* for measuring the activity of collaborative workspaces. Otjacques et al. [28] propose the *Coopadex* as a metric for measuring the average use of a collaboration workspace. Both metrics serve the same purpose and we applied a similar idea to our social documents (see below).

Bøving and Simonsen [5] suggest *"collaborativity metrics"* from documents and divide documents into three groups: (1) no edits, (2) edited by one user, (3) edited by several users. For each of these groups, they calculate the average *lifespan of documents* and the average *number of participating users*. The authors argue that such document-centric metrics provide better information on *collaborativity* and the *lifecycle* of documents in collaboration systems, which is in line with the research presented in this paper.

Benhiba et al. [4] propose three types of social artefacts but do not demonstrate their concepts with actual data. Their social artefacts distinguish between content and activities as lenses on collaboration. As outlined above, most previous research is based on user activity logs (transactional data). In this paper, we follow the approach by Bøving and Simonsen [5] and use documents (content data) to identify and analyze collaboration. The following section describes how we used existing approaches from the literature to derive new SCA metrics that use the structure of social documents as the basis for analysis.

4. SCA: Analyzing the Structure of Social Documents

Enterprise collaboration platforms are large-scale highly-integrated information infrastructures comprising an ecosystem of tools and functionality to support collaborative work [16, 20]. The main difference between these platforms and earlier forms of collaboration systems is the native integration of *social software* (e.g. wikis, blogs, social profiles, activities, likes, tags etc.) which enhance functionality for collaborative work [30]. Enterprise collaboration platforms are typically implemented in large organizations to provide a platform for the digital workplace, supporting collaboration between many thousands of employees, who may be widely dispersed across the organization [47].

IBM Connections is one of the few commercial software products currently on the market that can be used to build an *integrated* enterprise collaboration platform. Table 2 shows the characteristics for the platform used in our study.

However, due to limitations in skills regarding IT operations and budget, most medium- to small-sized companies build their own platform following a *portfolio approach* where they combine software from different vendors in order to provide the required range of functionality for their digital workplace. The downside of the portfolio approach is that each separate software application has (if at all) its own analytics tool, which is limited to the analysis of data from this particular software. This has made it (so far) impossible to derive a company-wide (platform-wide) view.

It is therefore not surprising that most of the above mentioned studies are limited to a single type of software (e.g. blogs or wikis) [30] or to specific activities (e.g. knowledge sharing or project management) [19]. Monteiro et al. argue that this localist focus, often on small group interaction is potentially problematic "in light of the kinds of large-scale, integrated and interconnected workplace information technologies [...] increasingly found within and across organizations today" [20].

Table 2. Description of the collaboration platform

Software platform:	UniConnect (based on IBM Connections)
Users:	Managers, researchers and students from Universities, companies and public agencies in the DACH area.
Number of users:	3500
Selected time period:	2014-2019
Number of workspaces	1200
Content:	34700 social documents with 137744 items
Examined databases	FORUMS, WIKIS, BLOGS, SNCOMM, FILES
Data type examined:	Social documents
Metrics used:	Listed in Table 3

For our research investigation, we had full access to all data (content and log files) of an operational instance of IBM Connections (UniConnect), an integrated collaboration platform with 3500 users and more than 1200 communities. Our data source contained around 34.700 social documents with 137.744 items (see Table 2).

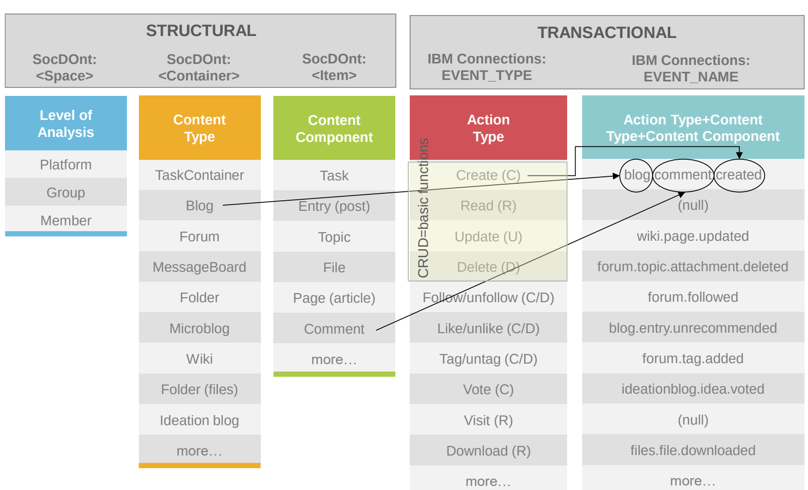

Figure 1. Dimensions of SCA

4.1 Structural vs transactional view

As described above, our research objective is to understand joint work around social documents. Our goal is to measure and understand the interactions around a document over time and develop a measure for the degree of collaboration. The structure of social documents represents how people communicate, share information and coordinate which links back to original research in the area of CSCW [7]. Thus, understanding the structure of these documents provides additional insights into collaboration.

With the three categories of social artefacts, Benhiba et al. [4] indicate a distinction between a *structural* perspective and a *transactional* perspective. Figure 1 shows these two perspectives with their dimensions for SCA.

The *transactional view* (right side of Figure 1) represents what a user has done on the platform. The *action types* (4th column) contain user activities (create, read, update, etc.). In SCA, these user activities can be interpreted (e.g. according to their level of engagement). The categories consumption, contribution and creation, for example, allow the identification of different users types e.g. creator, contributor, lurker, inactive and non-user [38].

In IBM Connections, there are 58 different basic functions; not all of them can occur in every functional module (container). The event log (METRICS) records a combination of the content type, the content component and the action (e.g. blog.comment.created) in a special field (EVENT_NAME, 5th column) and is thus ideally suited for analysis.

The *structural view* (on which we focus in this paper) on the left side of Figure 1 represents the content on the platform. We use the terminology from an established ontology in the field of Web Science (sioc) and its further development into SocDOnt (Social Document Ontology) [45]. The *space* (1st column) defines the level of analysis (the whole platform, selected group workspace(s) or a single user space). The content type (2nd column) is defined by its *container*, representing the physical place where content is stored. The content components (*items*, 3rd column) are the elements that form a (compound) social document.

Figure 2 shows graphical representations of social documents with their components. The intellectual entity is the item that initiates a social document. It becomes a *compound* social document when the first component is added. To give an example: a *user creates a blog post (intellectual entity) documenting the experiences at a conference she attended yesterday. A colleague reads the post and likes it to inform others in his network about the experience report. This brings the post to the attention of a third colleague who adds a comment asking if she met one of his most important customers at the event. To facilitate a later search, he also tags it with the name of the company hosting the event.* The (compound) social document now consists of four items, the intellectual entity (initiating post), another intellectual item (comment) and two simple items (like and tag).

Collections is the term used for social documents that are linked to each other usually through a hierarchical relationship (e.g. Wiki page/subpage). Collections with multiple authors are a very good indicator for interaction between users and thus collaboration [13]. The special structure of Social Documents described above also allows us to analyze how documents grow over time.

Not all items of a social document are equally “valuable” to an organization. Depending on its actual content, an *intellectual item* is the most valuable form of social document. It can consist of rich text and images/videos and is likely to contain a form of information or documentation that can be read and interpreted by others. Simple features (such as like or tag), on the other hand, are used to raise awareness or to classify content (and thus facilitate search). Social Collaboration Analytics on social documents can analyze the components of a document and help determine its potential “value”. It can also be used to create a *collaborativity index* similar to the ones suggested by [13] and [28] described above.

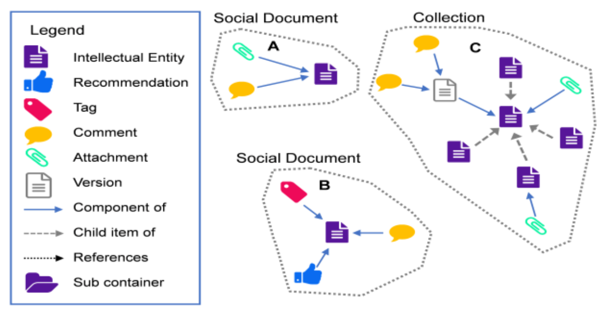

Figure 2. Components of Social Documents [21]

4.2 Developing Metrics for Social Documents

Guided by the literature discussed above and based on the structure of documents defined in the Social Documents Ontology (SocDOnt) [45], we developed the seven metrics listed and described in Table 3.

Table 3. Metrics for Social Documents (SD)

Metric	Explanation
# Compound SDs	Number of documents that consist of multiple components
∅ Contributors / SD	Average number of users who contributed at least one component to the SD
∅ Size / SD	Average number of components per SD
# Collections	Number of SDs consisting of multiple SDs
Content type distribution	Distribution (percentage) of the content types
∅ Components per member	Average number of contributed components per user
Rate of content growth	Rate of increase of components over time

The metrics can be calculated and displayed on the workgroup level with the help of a tool (*Content Dashboard*). The development and evaluation of this tool has been documented in [21].

The key concept behind the calculation of metrics for a social document is the use of its graph structure. By modelling and visualizing social content as a graph (Figure 2), where nodes represent (social document) items and edges represent their links/associations, social documents and collections can be identified as *connected components.* Put simply, a connected component is a subset of nodes, in which every node is connected to each other directly or via a path of their neighbors [40]. While it is easy for humans to find and count connected components in graph drawings of a manageable size, identifying these objects programmatically is a known problem in Computer Science and can be solved with a breadth-first search algorithm [40]. We make use of this algorithm for counting the number of social documents and collections by considering the different types of associations within the social content: compositions, parent child associations and references. We apply the breadth-first search algorithm on our graph twice: 1) We apply the algorithm on our graph containing only the set of edges that represent compositions. As a result, we obtain connected components that represent social documents. 2) We merge all sets of edges together (compositions, parent child associations and references) and apply the algorithm again. As a result, we obtain connected components that represent collections. Based on the programmatical identification of social documents and collections we were able to calculate the actual values for the metrics presented in Table 3.

4.3 Application of Metrics

As mentioned before, we evaluated the metrics on UniConnect, a collaboration platform hosted by our University for users from different organizations (see Table 2). The UniConnect platform is a large-scale integrated enterprise system that provides comprehensive features such as task management, blogs, files, forums, status updates (tweets), Wikis, joint editing of documents and more, on one unified platform. With the help of the Content Dashboard [21] we visualized the structure of the content (intellectual entities and components) and calculated the metrics listed in Table 3. Figure 3 shows the applications, intellectual entities and components that were analyzed.

We selected a *sample of 12 group workspaces*, which are used for four different use cases: Organizational Unit (2), Class (5), Community of Interest (2) and Project (5). We purposefully chose workspaces where we were able to look at the actual user activity in the frontend to be able to validate our results.

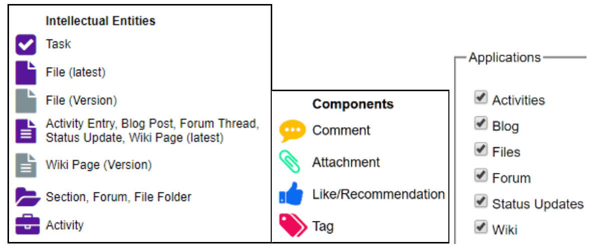

Figure 3. Intellectual entities, components and applications used in this study

Figure 4 shows the results of the metrics for the twelve selected workplaces over all applications (containers). The table shows the alias and the number of members (Mem) in the workplace. The column labeled *CDs* contains the number of compound documents, the column *Docs* shows the total amount of documents. *A/D* shows the ratio of authors per document and *C/D* the ratio of components per document. The last column labeled *Col* shows the number of collections. The highlighted fields show values greater than two for A/D and greater than five for C/D.

SD metrics all containers						
Alias	Mem	CDs	Docs	A/D	C/D	Col.
Class 1	167	54	114	1.9	3.0	12
Class 2	149	31	65	1.9	2.7	12
Class 3	145	54	100	1.8	2.5	12
Class 4	130	19	89	1.1	2.2	7
Class 5	114	18	95	1.2	1.9	8
CoI 1	10	67	151	1.5	2.9	7
CoI 2	6	52	149	1.3	2.4	15
OU 1	26	66	554	1.1	1.8	21
OU 2	4	26	100	1.1	1.7	17
Project 1	86	676	942	2.2	5.2	31
Project 2	8	91	113	1.6	5.2	8
Project 3	22	183	428	1.4	3.8	14
Project 4	20	172	446	1.5	3.7	65
Project 5	11	87	184	1.4	3.2	18

Legend: Mem=Members, CDs=compound documents, Docs=documents, A/D=authors per document, C/D=components per document, Col=collections

Figure 4. Distribution of content components

The two *Organizational Units (OU)* have 26 and 4 members, are longitudinal in nature (they have no fixed end date) and they serve an administrative purpose.

The five *Classes* have between 114 and 167 members (114, 130, 145, 149, 167) and content was added over one semester, mostly to make teaching material available (files), publish announcements and to discuss questions.

The two *Communities of Interest (CoI)* have no end date, as with the two OUs. They have six and ten members, and in each instance, they are used to discuss a specific joint topic of interest.

The five *Project* workspaces (members: 86, 8, 22, 20, 11) are used for cross-organizational project coordination. Three of these projects are finished, two are ongoing.

Figure 5 shows a sorted list with the values for components per social document (C/D).

4.4 Results and discussion

The results (Figure 5) show that, in our sample, the *number of components per SD* is an *indicator of the community type*.

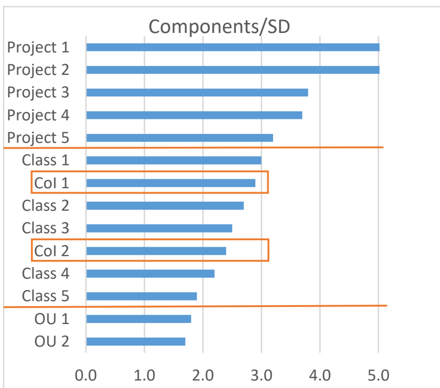

Figure 5. Components per SD in the 12 workspaces

The five workspaces that were used for project work have the highest average number of components attached to their documents. Workspaces used by organizational units (OU) showed the lowest richness in terms of their social documents. The middle group contains classes and general Communities of Interest (CoI, e.g. a Ph.D. Community that discusses literature). CoI have a *longitudinal nature* without a designated end. They are "ongoing" whereas in classes, content is *of limited temporal interest* and only added during one semester. At the end of this period, workspaces for classes turn into archives that are only used for exam preparation by the students who have not yet successfully finished the course.

These results indicate that the *purpose of a workspace* has an influence on the richness (number of components) of its documents. The workspaces that are output-oriented (typically projects) have more components attached to the intellectual entity. Not surprisingly, the documents in project workspaces are more complex (discussion, extension of content, etc.) in accordance with their use case. Organizational units, on the other hand, are mostly administrative and less interactive in nature with e.g. one person taking minutes during meetings or somebody occasionally announcing something to the others in a microblog. As a result, the content in these workspaces is not as rich as in project workspaces. The *longitudinal nature* of a workspace, on the other hand, seems to have little to no influence on the number of components per document.

The middle group contains classes (short-term, clear ending) as well as Communities of Interest (long-term, no defined ending). It is not surprising that classes are located in the middle of the figure. They typically have a small number of documents with high interactivity (discussion in the forum), which increases the average component size but also a lot of unidirectional communication (professor uploads files for students), which produce a value of 1 for A/D and C/D.

The values might appear low at first sight but one has to bear in mind that these are average numbers over *all* documents, including simple information shar-

Forum					Wiki					Microblog					Blog					Files				
CDs	Docs	A/D	C/D	Col.	CDs	Docs	A/D	C/D	Col.	CDs	Docs	A/D	C/D	Col.	CDs	Docs	A/D	C/D	Col.	CDs	Docs	A/D	C/D	Col.
0	0	0.0	0.0	0	9	20	1.2	2.2	1	0	11	1.0	1.0	0	4	17	1.2	1.4	1	5	29	1.1	1.8	6
1	1	5.0	10.0	1	23	34	1.3	11.4	1	9	26	1.6	1.8	0	2	4	1.8	1.8	1	31	489	1.0	1.2	18
6	6	2.3	3.7	2	4	4	2.3	15.5	1	0	0	0.0	0.0	0	4	17	1.2	1.4	1	4	68	1.0	1.1	4
1	1	2.0	5.0	1	8	8	1.8	11.8	1	0	1	1.0	0.0	0	4	18	1.2	1.3	1	6	61	1.0	1.2	4
3	3	3.0	6.3	1	20	25	1.1	6.1	3	11	18	1.7	2.3	0	13	20	1.8	2.8	1	3	80	1.0	1.1	8
35	35	2.9	4.7	8	0	0	0.0	0.0	0	2	5	1.4	1.4	0	6	15	1.4	1.5	1	11	45	1.1	1.3	3
25	29	2.9	4.6	7	0	0	0.0	0.0	0	0	2	1.0	1.0	0	2	4	1.5	1.5	1	4	30	1.0	1.2	4
1	1	2.0	4.0	1	47	51	1.8	5.7	2	12	15	2.5	3.5	0	1	2	1.5	1.5	1	6	82	1.1	1.2	3
42	48	3.0	5.5	8	0	0	0.0	0.0	0	3	3	2.3	2.3	0	5	12	1.5	1.5	1	4	51	1.0	1.1	3
0	1	1.0	1.0	1	44	52	1.6	6.2	1	1	3	1.3	1.3	0	6	11	1.5	1.9	1	33	82	1.4	2.4	8
4	4	2.0	6.8	4	45	184	1.2	3.5	6	23	36	2.0	2.7	0	27	33	3.3	4.2	32	46	161	1.2	2.5	3
0	0	0.0	0.0	0	101	104	1.9	9.4	1	10	13	1.6	2.5	0	13	32	1.6	2.2	1	47	257	1.2	1.9	6
0	0	0.0	0.0	0	32	33	1.5	8.6	4	18	29	1.7	3.2	0	10	17	1.8	2.9	1	1	1	1.0	2.0	0
35	39	3.1	9.2	9	135	138	2.5	8.5	12	170	186	2.5	3.7	0	96	105	5.6	6.9	2	242	477	1.2	4.0	6

Legend: CD=compound documents, Docs=documents, A/D=authors per document, C/D=components per document, Col=collections

Figure 6. Containers (sorted by degree of collaborativity)

ing (upload of a file) that does not call for, or require interaction. Files are not *born-social*; they are created outside the ESS e.g. with an office application and only "*become-social*" [12] once they are uploaded and users can start commenting, liking or tagging them.

Our analysis shows that files do not initiate a high degree of collaboration on the platform after they become social, most likely because collaboration on files takes place during their creation outside of our ESS and thus, they are already in a "finalized state" when they are uploaded.

As we can see in the example of a file upload, the average numbers on all content types can give us a first indication about the use case(s) of a workspace. For a better understanding of the collaborative activity around documents, it is necessary to take the analysis to the more detailed level of the single containers.

Containers provide the physical storage spaces for specialized applications. There are multiple applications available in an integrated collaboration platform and each of them offers different affordances to the user. The term *affordance* [25] is used in CSCW research to refer to the perceived and actual properties of a thing, or, in our context, the functionality that a user would expect from a functional module in an Enterprise Social Software. Since the affordances of forums, Wikis, microblogs, blogs and files are all different, we expect to see differences in the structure of their content.

Figure 6 shows the results of the same sample of workspaces but this time on the level of containers (forums, Wikis, microblogs, blogs and files). Values greater than 2 are highlighted in yellow and indicate a high average number of authors (A/D) or components (C/D) for the documents. Rows filled with only zero indicate that the respective container is not in use in this workspace.

Forum: The forum is the most "collaborative" container. Overall, it has the highest numbers of authors (A/D) and components (C/D) per social document (for the cases when the forum is used at all). The purpose of a forum is "discussion" in which multiple people add multiple components to the conversation around an ongoing topic, so it is not surprising that it has the highest average number of components.

Wiki: Documents in this container are the "richest", that is, they have the highest average number of components. This is to be expected because in IBM Connections this software module supports versioning, so every change to the intellectual entity creates a new version and thus, a new component of the social document. The high values in the result table of the Wiki are a reflection of its affordance of joint editing and information collection, a process in which multiple people (should) contribute.

Microblog: The *microblog* appears to be quite collaborative, which can be traced back to a high number of likes (recommends). The like is an awareness feature and a particularly important affordance of a microblog, which involves the exchange of short messages that are usually only of current interest. Most of the average values for the microblog are over 2 and up to 3.7, which shows that these short messages are on average recommended by 1-2 people. One exception (10) stands out in the data. The exploration of the source data showed that the value originated from an exercise class where a few very important posts had been recommended by a group of students to make sure that fellow students did not miss them.

Blogs: Blogs serve a similar purpose to microblogs (i.e. to share information with others) but for richer content (longer text, images) than a short message in a microblog. The values between 1.2 and 1.8 confirm the similarity of use. There are two exceptions (4.2 and 6.9). A closer examination of these workspaces showed that they are used for project management and the higher number of components for blog posts were caused by a possible (mis)use of this functionality. Some blog posts had stimulated an intensive discussion, an activity, which might have been better located in a forum. The high forum values confirm that there is an above average degree of discussion between the members in these two workspaces.

Files: Files have the highest number of *collections* because users frequently group them into folders. The numbers of authors per document (A/D) and the richness of components (C/D) is the lowest of all the containers. The values are evenly distributed between one and two with only a few exceptions, meaning that for most files there is no second author who contributes a component (not even a simple component such as a tag or a like). As mentioned previously, files are mostly of a documentary nature and their purpose is simply to share them with others once work on them has been finished. Our study confirms that inserting them into a social software does not stimulate "joint work" around them.

5. Conclusion and Future Research

The findings presented in this paper are an outcome of a larger research program on the characteristics and nature of ESS and the ways it is shaping the behavior and practices of joint work in the digital workplace. The paper presents a novel approach to the area of Social Collaboration Analytics: we focus on the *document* perspective (instead of the *user* perspective) and examine user activity around these artefacts. Our main contribution are the metrics for measuring collaborative activity and a preliminary demonstration of how these metrics can be calculated and interpreted to examine collaboration in a fully operational, large-scale integrated collaboration platform. The large-scale, integrated nature of our platform provides us with multiple content types for study and helps us address the limitations that Monteiro et al. [20] identified as "localist" studies that focus on a single-site implementation or a given system and enables us to study how users collaborate using multiple functional modules. We focus our analysis on the workspace level (involving clearly defined, mostly small to medium-sized groups) where the actual "joint work" takes place and not on the platform level common to many other studies. In our sample, these groups ranged from four to 167 people.

The social document graph is the basis for calculating the proposed metrics. We are currently planning to extend our work to include the analysis of hyperlinks that are contained in the content part of social documents, which will reveal additional relationships with other social documents. This will be especially relevant when we extend our examination from *integrated* systems (with a unified database) to a software *portfolio* (with differing database structures). It is our intention to use the ontology for social documents (SocDOnt) for the mapping of heterogeneous data structures of different systems, e.g. to analyze the Microsoft portfolio comprising Yammer, Skype and SharePoint.

Complementary to the study of the structure of social documents, we have started to experiment with the interpretation of the content on our platform, using text mining, sentiment analysis and a tone analyzer. This will further enrich the interpretation of the data on the platform with the final objective to create a dashboard that provides information on multiple facets of collaboration. In future work we are also planning to increase our sample size by including all active workspaces on our platform and feeding the results into SPSS for cluster analysis. This will then reveal typical workspace patterns and allow a platform-wide view of collaboration.

6. References

[1] Asur, S., and B.A. Huberman, "Predicting the Future With Social Media", *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, (2010), 492–499.

[2] Behrendt, S., A. Richter, and K. Riemer, "Conceptualisation of Digital Traces for the Identification of Informal Networks in Enterprise Social Networks", *25th Australasian Conference on Information Systems*, (2014), 1–10.

[3] Behrendt, S., A. Richter, and M. Trier, "Mixed methods analysis of enterprise social networks", *Computer Networks 75*, 2014, pp. 560–577.

[4] Benhiba, L., and M.A. Janati-idrissi, "Supercharging Enterprise 2.0 Analytics Modules with SNA Metrics", *International Journal of Computer Science Issues 10*(6), 2013, pp. 89–96.

[5] Bøving, K.B., and J. Simonsen, "http Log Analysis - an Approach to Studying the Use of Web-Based Information Systems", *Scandinavian Journal of Information Systems 16*, 2004, pp. 145–174.

[6] Damianos, L.E., and L.J. Holtzblatt, *Measuring Community Success: One Size does not fit all*, Bedford, MA, USA, 2010.

[7] Ellis, C.A., S.J. Gibbs, and G. Rein, "Groupware: some issues and experiences.", *Communications of the ACM 34*(1), 1991, pp. 39–58.

[8] Fan, S., X. Li, and J.L. Zhao, "Collaboration Process Pattern Approach to Improving Teamwork Performance: A Data Mining-Based Methodology", *INFORMS Journal on Computing 29*(3), 2017, pp. 438–456.

[9] Greeven, C.S., and S.P. Williams, "Enterprise collaboration systems: addressing adoption challenges and the shaping of sociotechnical systems", *International Journal of Information Systems and Project Management 5*(1), 2017, pp. 5–23.

[10] Hacker, J., R. Bernsmann, and K. Riemer, "Dimensions of User Behavior in Enterprise Social Networks", In R. Helms, J. Cranefield and J. van Reijsen, eds., *Social Knowledge Management in Action*. Springer International Publishing, Heidelberg, Germany, 2017, 125–146.

[11] Hacker, J., F. Bodendorf, and P. Lorenz, "A Framework to Analyze Enterprise Social Network Data", In M. Atzmueller, O. Samia and T. Roth-Berghöfer, eds., *Enterprise Big Data Engineering, Analytics, and Management*. IGI Global, Hershey, USA, 2016, 84–107.

[12] 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.

[13] Jeners, N., and W. Prinz, "Metrics for Cooperative Systems", *GROUP'14 18th International Conference on Supporting Group Work*, ACM (2014), 91–99.

[14] Kane, G.C., "Enterprise social media: Current capabilities and future possibilities", *MIS Quarterly Executive 14*(1), 2015, pp. 1–16.

[15] Kumar, S., F. Morstatter, and H. Liu, *Twitter Data Analytics*, Springer, New York, NY, 2014.

[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] Leung, C.K., F. Jiang, T.W. Poon, and P.-É. Crevier, "Big Data Analytics of Social Network Data: Who Cares Most About You on Facebook?", In *Studies in Big Data, Volume 27*. 2018, 1–15.

[18] MarketsandMarkets, *Enterprise Collaboration Market by Component, Solution, Service, User Type, Vertical, and Region - Global Forecast to 2021*, 2017.

[19] Millen, D.R., M.J. Muller, W. Geyer, E. Wilcox, and B. Brownholtz, "Patterns of media use in an activity-centric collaborative environment", *CHI 2005*, (2005), 879–888.

[20] Monteiro, E., N. Pollock, O. Hanseth, and R. Williams, "From artefacts to infrastructures", *Computer Supported Cooperative Work 22*(4–6), 2013, pp. 575–607.

[21] Mosen, J., S.P. Williams, and P. Schubert, "Visualizing Social Documents as Traces of Collaborative Activity in Enterprise Collaboration Platforms", *53rd Hawaii International Conference on System Sciences*, (2020).

[22] Muller, M., "Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media", *Conference on Computer Supported Cooperative Work,* (2012), 253–256.

[23] Muller, M.J., J. Freyne, C. Dugan, D.R. Millen, and J. Thom-Santelli, "Return On Contribution (ROC): A metric for enterprise social software", *11th European Conference on Computer Supported Cooperative Work (ECSCW)*, Spinger (2009), 143–150.

[24] Nasirifard, P., and V. Peristeras, "Expertise extracting within online shared workspaces", *WebSci'09: Society On-Line*, (2009).

[25] Norman, D., *The Design of Everyday Things: Revised and Expanded Version*, Basic Books, New York, 2013.

[26] van Osch, W., B. Bulgurcu, and A. Majchrzak, "The Effects of Group Visibility on Social Capital Formation in Enterprise Social Media", *Twenty-Sixth European Conference on Information Systems*, (2018).

[27] van Osch, W., C.W. Steinfield, and B.A. Balogh, "Enterprise social media: Challenges and opportunities for organizational communication and collaboration", *48th Hawaii International Conference on System Sciences*, (2015), 763–772.

[28] Otjacques, B., M. Noirhomme, X. Gobert, and F. Feltz, "Cooperation Indexes to Support Workspace Awareness", In Y.A. Dimitriadis, I. Zigurs and E. Gómez-Sánchez, eds., *Groupware: Design, Implementation, and Use*. Springer Berlin Heidelberg, 2006, 94–101.

[29] Recker, J.C., and D. Lekse, "A field study of spatial preferences in enterprise microblogging", *Journal of Information Technology 31*(2), 2016, pp. 115–129.

[30] Richter, A., and K. Riemer, "The Contextual Nature Of Enterprise Social Networking: A Multi Case Study Comparison", *European Conference on Information Systems*, (2013), 1–12.

[31] Riemer, K., P. Overfeld, P. Scifleet, and A. Richter, "Eliciting the anatomy of technology Appropriation Processes: A Case Study in enterprise social media", *European Conference on Information Systems*, (2012).

[32] Riemer, K., and P. Scifleet, "Enterprise Social Networking in Knowledge-intensive Work Practices: A Case Study in a Professional Service Firm", *23rd Australasian Conference on Information Systems*, 2012, pp. 1–12.

[33] Riemer, K., S. Stieglitz, and C. Meske, "From Top to Bottom: Investigating the Changing Role of Hierarchy in Enterprise Social Networks", *Business & Information Systems Engineering (BISE) 57*(3), 2015, pp. 197–212.

[34] Schubert, P., "Joint Work and Information Sharing in the Modern Digital Workplace: How the Introduction of 'Social' Features Shaped Enterprise Collaboration Systems", In K. Riemer, S. Schellhammer and M. Meinert, eds., *Collaboration in the Digital Age: How Technology Enables Individuals, Teams and Businesses*. Springer, Berlin, Heidelberg, 2018, 45–59.

[35] 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.

[36] Schwade, F., and P. Schubert, "Social Collaboration Analytics for Enterprise Social Software: A Literature Review", *Multikonferenz Wirtschaftsinformatik 2018*, (2018).

[37] Schwade, F., and P. Schubert, "A Survey on the Status Quo of Social Collaboration Analytics in Practice", *26th European Conference on Information Systems*, (2018).

[38] Schwade, F., and P. Schubert, "Developing a User Typology for the Analysis of Participation in Enterprise Collaboration Systems", *52nd Hawaii International Conference on System Sciences*, (2019), 460–469.

[39] Shami, N.S., M. Muller, and D. Millen, "Browse and discover: social file sharing in the enterprise", *ACM Conference on Computer Supported Cooperative Work*, ACM (2011), 295–304.

[40] Skiena, S.S., *The Algorithm Design Manual*, Springer London, London, UK, 2008.

[41] Steinfield, C., W. van Osch, and Y. Zhao, "Intra-Organizational Boundary Spanning: A Machine Learning Approach", *21st Americas Conference on Information Systems*, (2015).

[42] Stieglitz, S., L. Dang-Xuan, A. Bruns, and C. Neuberger, "Social Media Analytics", *Business & Information Systems Engineering 6*(2), 2014, pp. 89–96.

[43] Voronin, G., A. Baumann, and S. Lessmann, "Crowdturfing on Instagram - The Influence of Profile Characteristics on The Engagement of Others", *Twenty-Sixth European Conference on Information Systems 2016*, 2018.

[44] Webster, J., and R.T. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review", *MIS Quarterly 26*(2), 2002, pp. xiii–xxiii.

[45] Williams, S.P., J. Mosen, and P. Schubert, "The Structure of Social Documents", *53rd Hawaii International Conference on System Sciences*, (2020).

[46] 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.

[47] Williams, S.P., and P. Schubert, "Designs for the Digital Workplace", *Procedia Computer Science*, Elsevier B.V. (2018), 478–485.