Identifying Bridge Users: the Knowledge Transfer Agents in Enterprise Collaboration Systems

Priscilla Jimenez-Pazmino
University of Notre Dame
pjimenez@nd.edu

Trenton Ford
University of Notre Dame
tford5@nd.edu

Ronald Metoyer
University of Notre Dame
rmetoyer@nd.edu

Nitesh Chawla
University of Notre Dame
nchawla@nd.edu

Abstract

In recent years enterprise collaboration systems (ECS) integrated with social network capabilities have become popular tools for supporting knowledge management (KM) strategies and organizational learning. Increased usage has resulted in higher interest in understanding and classifying the roles that ECS users adopt online. Previous research has studied user role identification by considering: the degree of participation in an ECS, the user interactions with shared content, the user role in the ECS network, and the user KM-role observed within an interaction. Although all of these factors provide insights into ECS user engagement, they fail to fully consider the knowledge sharing perspective. In this paper, we define "bridge users" within the context of KM and present a framework for identifying them using semantic analysis of user-generated content. Further, we present results and observations from tests of our pipeline on the ECS of a large multinational engineering company with more than 100k users.

1. Introduction

Increasingly, organizations are using Enterprise Collaboration Systems (ECS) as Knowledge Management (KM) platforms to support their organizational KM strategies [1, 2]. KM and organizational learning managers are interested in increasing knowledge-sharing activities and identifying success metrics related to the usage of these KM platforms [3]. Current approaches to studying these platforms rely on enterprise social network analysis [4, 5] and social collaboration analysis [3, 6, 7], which focus on user-to-user interactions, and not users' interactions within and between distinct communities. From either perspective, users are the key agents of knowledge-sharing activities within an organization; therefore, identifying the type of users and their knowledge-sharing behaviors within an ECS coupled with how that behavior supports knowledge flow is an important focus for KM researchers.

Research on ECS users supports common user categories such as creators, contributors, lurkers, inactives, non-users and users without accounts [8]. Prior research has also studied how Communities of Practice (CoPs) adopt these KM technologies and how the subsequent utilization affects the community structure over time [9]. These studies are useful to highlight levels of member participation, the collaboration within a CoP, and how the technology is adopted. However, none of these studies illuminate the question of how to identify knowledge transfer (KT) instances enabled by individual users who engage with multiple CoPs, we call this type of user a *bridge user*.

Within the context of this research, bridge users are defined as individuals who are pivotal to knowledge flow within an ECS network. Bridge users are community members who communicate with members in different, often siloed, communities to enable knowledge transfer between them. The behavior has been shown to boost innovation within organizations [10]. Thus, identifying bridge users is an important task whose results highlight the members who facilitate the knowledge flow across communities. Identifying bridge users is a challenging task though, it involves being aware of the knowledge a user has either learned in one community and made known to another, or that they originated and distributed to multiple communities.

Accurate identification of bridge users could be achieved by self-reporting knowledge transfer activities and contributions; however, this task becomes time-consuming for both the ECS users and the KM officers (for the purpose of this paper, we will refer to KM officers as the staff who plan and evaluate the KM strategies implemented in an organization).

To provide insights about the identification of bridge users, we studied the Enterprise Collaboration Network (ECN) of a large multinational engineering company. The ECN data included user interactions such as posts, questions, and replies, along with the user-generated text

content related to each interaction type. Further, the ECN also contained user-defined communities to which users could join and post within. Using this dataset and considering the gaps in the role identification literature, this paper answers the following question:

> RQ: Can we identify cases in which an ECS user applies/transfers knowledge learned in one community of practice to another community?

To answer this question, we developed a pipeline that utilizes structural analysis of a communication network and semantic analysis of user-generated text content content to identify bridge users, and we present the results of our pipeline's efficacy for the ECN of a large multinational engineering company. This work suggests that the identification of bridge users can be automated through the pairing of text analysis and structural analysis of an ECN.

2. Background and Related Work

Organizations rely on ECSs to support collaboration among employees. In this section, we describe the application context and summarize the approaches that other researchers have taken to study user roles within enterprise collaboration systems.

2.1. Context of Application

By exercising KM, organizations seek to provide their members with professional environments that enable the effective utilization and transference of their knowledge [11, 12]. Knowledge sharing is one of the key activities in knowledge management (KM) [13, 14]. Knowledge sharing is the product of situated learning systems that, within the professional context, are observed in Communities of Practice (CoPs). CoPs are platforms that allow organization members who share passions and interests, to exchange knowledge, and learn from one another [15, 16].

KM platforms, a type of ECS [17], are digital environments that combine software features such as member profiles and groupware components (e.g. group mailing list, scheduling events, and digital spaces) to provide members and CoPs with virtual shared spaces and tools to search for subject area experts, engage in discussion threads, and find answers to questions. Organizations provide their members with KM platforms to support knowledge sharing between members and within CoPs [1, 2]. The utilities of KM platforms are very well known [18, 19, 20, 21]. They enable users to find useful information and resources; helping users to have a more extensive network of professional contacts; providing repositories

for content created by users and CoPs; and facilitating learning, knowledge sharing, problem solving, and innovation. However, when KM officers try to assess the usage of these systems their approaches are limited to quantifying and identifying the types of usage (e.g. project, organizational, or task-oriented) [3], and to characterizing and identifying the members of CoPs (e.g. identifying roles of leaders and learners, or degree of participation) [22]. So far, little has been done to identify users that engage in knowledge transfer.

2.2. User Roles in ECSs

The study of ECS user characteristics has long been of interest since users act as creators and transmitters of knowledge within the system. Previous research on the types of users in ECSs has focused on classifying them considering: 1) the user membership or degree of participation in a CoP or knowledge network [8, 16, 23]; 2) the user's actions taken related to the content (e.g. creator vs consumer) [24, 25]; 3) the social role of the user within the network (e.g. coordinator, debater, seeker, helper) [26, 27]; and 4) the KM-related role of the user observed within the interaction (e.g. information brokers) [28].

The user typology based on degree of participation or membership usually considers the frequency of user participation or level of commitment within a CoP. As a result users are classified as core members, active participants, occasional participants, peripheral participants, inactives, non-users, or users without accounts [8, 19]. Other typologies consider the user's action related to the content, classifying users as knowledge creators, knowledge sharers, or knowledge users [24, 25]. Other researchers followed similar approaches to study the social roles of users in online communities [26, 28]. Some of these roles are initiator, debater, coordinator, seeker, helper, expert, and socializer. Only a few researchers have studied users emphasizing KM-related roles [28], specifically identifying brokers as a user type who ease searches for knowledge by referring experts to information seekers. These user types emphasize user engagement to indicate how much, how, and who is using the KM platform, but it does not inform us about the users who engage in various CoPs exercising knowledge transfer.

2.3. Network Analysis Approaches in the Identification of User Roles

Previous work [26, 29, 30, 31] has studied the identification of key users (e.g., influential users or knowledge actors) within ECNs using network analysis techniques. Researchers use measures such

as betweenness centrality to find structural bridges and siloes within networks [32]. Most of those methods tried to identify these key users based on network structure alone. Unfortunately, the bridges found using structure only methods do not necessarily serve as bridge users in the KM context as these methods do not consider the content being transmitted. Other work [30] has used ECS content (unstructured data) such as user posts and replies to construct a network; however, these methods relied on human evaluation to determine whether text content was a knowledge transfer. Although some of these methods were able to correctly identify bridge users, due to the large human time expense the size of a network severely limits their applicability. To extend applicability, this research develops an automated process for identifying bridge users.

3. The Role of Bridge Users

Consider the following scenario. An employee X actively engages in two CoPs inside an organization. One of the CoPs, $CoP\text{-}A$, focuses on sharing knowledge about digital tools that can be used at the workplace. The second CoP, $CoP\text{-}B$, focuses on the use of technology and innovation for improvement of business services and products. Employee X decides to share via the KM platform knowledge learned in $CoP\text{-}A$ that could support $CoP\text{-}B$ organizational processes. Similar knowledge flow takes place when employee X answers questions posted in $CoP\text{-}B$ with knowledge learned in $CoP\text{-}A$. In both cases, the active participation of employee X facilitates knowledge flow among employees who belong to two different CoPs. This knowledge-sharing behavior could take place offline or online. In any case, employee X is important to the organization. Employee X is acting as a bridge user.

Bridge users are pivotal to the structure of the ECN, thus to the knowledge flow. A *bridge user* is defined as a user who shares or disseminates similar knowledge in more than one community. In some instances, the action performed by a bridge user corresponds to the "alert" action type defined in [8], where the user notifies others about existing content or knowledge-related event. However, when the user is not only notifying but transferring knowledge from one CoP to another, this collaborative action becomes more meaningful.

In some instances, the observed collaboration is highly related to the particular job a user performs within an organization. This is the case with "boundary spanners", who are employees who work across unit boundaries [33] and their job description includes frequent interactions across group boundaries that results in knowledge sharing. Although, the knowledge sharing activity can be perceived as the same of a bridge user, the boundary spanner role is a human resource strategically allocated by the company which focuses mainly on searching for external information, making connections, or overcoming differences across unit boundaries [34, 35]. Thus, a bridge user is a valuable human resource in the sense that performs voluntarily collaborative interactions within member-led communities which are not necessarily within organizationally defined units. Similarly, the concept of "knowledge broker" in organizations refers to individuals that facilitate the transfer of knowledge among organizational units, however the main difference is that these individuals do not perform their role as part of a group but as external agents [36, 37], examples of knowledge brokers are identified as part of some job such as IT professionals or unit evaluators [38, 39, 40].

In the KM context, bridge users are key employees because they are holders of information or knowledge. By their collaborative action within and among CoPs via the KM platform, they demonstrate their ability and willingness to share with others. In a KM platform, we are looking to identify similar content that has been shared in different CoPs. In this instance, there are two ways to identify bridge users: 1) a user who shares related content in two different CoPs (in this case, the original source of the knowledge is unknown), and 2) a user who shares in a CoP what they have learned from another users' post in a different CoP.

4. Research Setting

This work was conducted as part of a larger project that aimed to evaluate the use of a KM platform and identify key users. To study bridge users, we examined KM platform user contributions (posts) within a large multinational engineering company. Our dataset consists of ECN data from 2008-2018 and includes contributions (posts/documents) within CoPs from engineering divisions only. For each contribution captured by the ECS, the user that posted the message, the date and time the message was posted, the content of the message, and the CoP where the message was posted are all available. The initial dataset consisted of approximately 117k users and 91k text contributions.

5. Proposed Methodology

For this research, we developed a pipeline (shown in Figure 1) which takes as input a network containing preestablished communities, users, and their associated
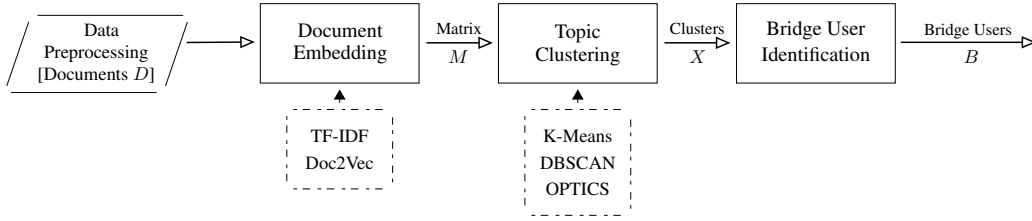
Figure 1. Bridge User Identification Pipeline. Documents D contains the set of documents belonging to potential bridge users as described in 5.2. Each document $d \in D$ contains metadata such as contributing user, posting date, and CoP in which the document was published. Matrix M contains the document embeddings produced using either TF-IDF or Doc2Vec. Set X contains clusters x produced by either K-Means, DBSCAN, or OPTICS clustering algorithms. Set B contains pairs of bridge users and their bridging contributions.

posts and produces as output a subset of users who, by transferring knowledge between communities, serve as bridges. Our process extends previous research that identifies bridge users using network structure alone through the inclusion of text content from users' network contributions in the identification process. It uses a similar strategy to the method proposed in the 2014 work by Berger et al. [30], but leverages semantic analysis techniques to offload much of the human intervention proposed previously. At a high level, the pipeline has four main parts: (1) data preprocessing, (2) document embedding, (3) topic clustering, and (4) bridge user and bridging contribution identification. The specific algorithms used for parts 2 and 3 of the pipeline are interchangeable depending on their applicability to a dataset of interest. For this reason, part of our analysis of results involves the comparison of document embedding methods and topic clustering methods as they relate to the pipeline's precision in the bridge user identification task.

5.1. Data Scheme

This pipeline requires a document corpus (set of CoP contributions) along with the contribution's associated metadata: user, predefined CoPs, and a posting date. CoPs in our context are user-led groups created to serve as discussion spaces for Q&As, how-tos, learning, and ideation around topics of interests.

In our network data, CoP contributions are user-generated text-based posts within the CoP. CoP contributions can be either initial posts or replies to other posts created by users who are members of the CoP. For ease of description, we attribute each document d_i from the set of documents D with the user who contributed the document ($d_i.user$), the posting date that the document was contributed ($d_i.date$), and the CoP to which the document was contributed ($d_i.CoP$).

5.2. Implementation

Algorithm 1 Bridge User Identification Pipeline

> **Input**: documents D
> **Output**: bridge users B

1: Set $B \leftarrow \{\}$
2: Matrix $M \leftarrow$ DOCUMENTEMBEDDER(D)
3: Clusters $X \leftarrow$ TOPICCLUSTERER(M)
4: **for all** Cluster $x \in X$ **do** $\triangleright$ $x \ni$ documents d
5:     **for all** document pairs $(d_i, d_j) \in x | i \neq j$ **do**
6:         **if** $d_i.CoP \neq d_j.CoP$ **then**
7:             **if** $d_i.date < d_j.date$ **then**
8:                 **if** $d_j.user \in d_i.CoP$ **then**
9:                     $B \leftarrow B \cup \{(d_i.user, d_i)\}$
10: **return** B

---

(1) Data Preprocessing

The first stage of the pipeline is to clean and reduce the size of our dataset using purely structural properties of bridge users within the network. This method is inline with previously discussed non-content based methods for finding bridge users [41]. In particular, we select users who are members of at least two communities and have made a contribution to at least one community. Similar to many social networks, the proportion of users who actively contribute to communities is relatively small when compared with the total number of users. In our case, the number of users fitting this criterion represents 7% of the entire user population (7,419 users out of $approx$ 117k users). We consider this subset of users as our set of potential bridge users.

(2) Document Embedding

State of the art topic analysis techniques require their input to be a matrix. Thus, we need to represent corpus

D, in our case all contributions from our potential bridge users, with a matrix (see Algorithm 1, line 2). We created the matrix representation of the corpus D using two methods: 1) Term Frequency-Inverse Document Frequency (TF-IDF) [42], and 2) Doc2Vec [43].

TF-IDF is the most often used approach for representing a set of documents as a matrix, where the columns represent documents and the rows represent words. Thus, the size of a TF-IDF matrix is $|D| \times |V|$ where $|D|$ is the total number of documents and $|V|$ is the total number of words contained in all documents in D. The TF-IDF value attempts to capture the importance of a word contained in any single document d_i by considering how frequently the word is used throughout the collection of documents. As an example, a word used frequently in a single document d_i and in no other documents would have a higher TF-IDF value than a word used frequently across many documents [42].

Doc2Vec is an increasingly popular method for embedding documents into a matrix representation. Internally Doc2Vec uses an unsupervised learning approach (a shallow neural network) to learn the best matrix representation that describes the documents in D. The size of the matrix is $|D| \times d$, where d is the number of latent features (hidden layer weights) learned by Doc2Vec. The advantage of this method is that it captures more nuanced similarity between documents and not simply exact word matches as TF-IDF does [43].

(3) Topic Clustering

The third step of the pipeline relies on the identification of documents containing similar topics. In our case, the documents are users' text contributions, but the intuition is the same; we would like to group together user contributions that are topically similar (see Algorithm 1, line 3). In this stage we do not consider whether the contributions take place in different CoPs, we only care if the contributions are similar in content. To accomplish this task, we utilize the matrix representations of the documents that were generated in stage 2 of the pipeline.

For either the TF-IDF or Doc2Vec matrices, we can attempt to capture this similarity using the same methods. We tested three well known methods for clustering documents: K-Means, DBSCAN, and OPTICS. These three methods aim to segment the data into groups of similar features and they produce output in the form of cluster associations for each document in D. The set of clusters returned is X, and individual clusters $x \in X$ contain documents from D. Details about how these clustering methods work are beyond the scope of this paper, but the referenced papers provide

well crafted explanations.

(4) Bridge User Identification

Using the set of document clusters generated in the previous stage, we must now generate the set of bridge users B. Each cluster x, in the set of clusters X, contains semantically similar documents, but whether the documents are bridges requires that we analyze the documents' metadata. As discussed previously, we would like to consider users who either post similar knowledge in multiple CoPs or users who may see a post in a community of which they are a member, and transfer that knowledge to another community.

To capture both of these knowledge transference types, we need to consider a document's contributor, CoP, and posting date. These attributes are associated with each document during the data preprocessing stage of the pipeline. Specifically, for each pair of documents in a cluster x we make sure that they were posted in different communities, determine which was posted most recently, and check if the member who posted the most recent contribution is a member of the group in which the older contribution was posted. If all three criteria are met for a pair of documents in a given cluster x, we store the user whose contribution was most recent and their bridging contribution in B (see Algorithm 1, line 4-9).

6. Analysis

In this section, we discuss how analysis of the clusters produced by the pipeline and the bridge users identified by the pipeline relates to overall pipeline performance.

6.1. Cluster Analysis

Most clustering methods require parameter tuning to produce clusters that best capture sufficiently similar data. Tuning these parameters requires some method of measuring the quality of a cluster. We selected a modified version of Akaike information criterion (AIC) called AICc [44] to tune K-Mean's *k* number of clusters, DBSCAN's *epsilon* parameter, and OPTICS's *xi* parameter. We chose AICc due to AIC's well documented utility for this task, and AICc's increased penalty against higher numbers of clusters.

Recall that the intuition behind clustering the document embeddings is that documents in the same cluster should be topically similar. We use AICc to find the best fit for the document embeddings, but to confirm that the embedding and clustering truly captured sufficiently similar documents in the same clusters we

manually code the documents in each cluster and report whether they actually contain messages that are similar. Here we apply binomial proportion confidence intervals (BPCIs) [45] to the results of the human coding from each cluster and report the precision of each pipeline combination.

6.2. Bridge User Analysis

The last stage of the pipeline produces pairs of users along with their bridging contributions. With this output, we manually analyze the content of each bridge to confirm that they contain valuable knowledge and determine the type of contribution. This analysis verifies that the contributions are truly transferring knowledge. We coded using a binary classification of either 0 or 1 (0 = false bridge; 1 = true bridge) by two coders.

Following this, we performed a thematic analysis to identify the common content types present in the documents of the dataset (pipeline output).

Finally, we investigate the distribution of users within the clusters produced by each of our pipeline combinations. This helps us to identify whether the source of the knowledge was the KM platform.

7. Results

Table 1. Document embedding dimensions.

Embedding Method	Output Size (rows;columns)
TF-IDF	(91,176; 101,003)
Doc2Vec	(91,176; 128)

Given the large size of the resulting matrix of the TF-IDF embedding, as shown in Table 1, coupled with the computational complexity of the K-Means and OPTICS methods the (TF-IDF, K-Means) and (TF-IDF, OPTICS) pipelines were impractical to test, therefore results for pipeline configurations including TF-IDF embeddings are limited to those using the DBSCAN clustering method. The optimal parameter values for the tested pipeline combinations are reported in Table 2, and the cluster size distributions for each tested combination are reported in Table 3.

Table 2. Clustering method parameters tuned using AICc clustering fit measure.

Embedding Method	K-Means	DBSCAN	OPTICS
TF-IDF	*NA*	$eps = 0.42$	*NA*
Doc2Vec	$k = 55$	$eps = 0.8$	$xi = 0.05$

Table 3. Cluster size distributions based on pipeline configurations and AICc optimized parameter tuning.

Clustering Method		Min	Max	Mean	STD
D2V K-Means	($k = 55$)	365	6,422	1,657.81	1,093.32
D2V OPTICS	($xi = 0.05$)	2	16	6.16	3.08
D2V DBSCAN	($eps = 0.8$)	2	292	7.93	31.92
TF-IDF DBSCAN	($eps = 0.42$)	2	24	8.25	7.56

Figure 2 shows clustering precision values and confidence intervals generated using the BPCI method (mentioned in Section 6.1) for the different pipeline combinations. Higher precision values mean that more of the documents in each cluster had similar topics. The confidence interval is influenced by the number of documents that were contained in each cluster. For instance, a cluster containing 10 documents where 95% are correctly clustered would have a larger confidence interval than a cluster containing 100 documents where 95% are correctly clustered. Although the (TF-IDF, DBSCAN) pipeline had a higher clustering precision, the number of results contained in the clusters was substantially lower than the number returned by the (Doc2Vec, DBSCAN) combination. The (Doc2Vec, DBSCAN) pipeline combination appears to perform best for our dataset. It can also be observed that OPTICS's performance was notably worse than both pipeline configurations using DBSCAN clustering.

We observe a similar trend in the pipeline's overall ability to successfully identify bridge users. Figure 3 shows the precision values for true bridges within clusters. As with the clustering results, pipeline configurations using DBSCAN performed better than pipelines using OPTICS clustering. Between the two DBSCAN variants, we observed that the precision of the Doc2Vec embedding is slightly lower than the TF-IDF embedding, yet the confidence interval for the TF-IDF pipeline is much larger meaning that the TF-IDF pipeline identified considerably fewer bridge users than the Doc2Vec pipeline. Overall, our results indicate that when parameters are tuned using AICc measures, the combination of Doc2Vec document embedding and DBSCAN clustering perform best on our dataset.

To verify the bridge users suggested by the pipeline, we coded each contribution as "true bridge" or "false bridge." Two coders attained "good" intercoder agreement as indicated by a Cohen's kappa of 0.75 (comparing at least 10% of the data). Any disagreements between coders were addressed and resolved, and then a single coder completed the task for the rest of the documents generated by the pipeline.

The results of the analysis and verification of bridge users (see Table 4) reveals that 75% of reported bridge users (output by the pipeline) can be
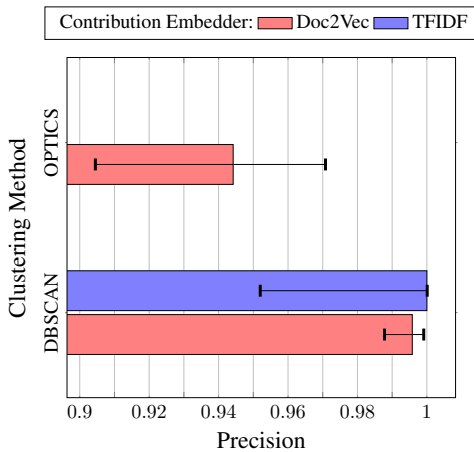
Figure 2. Comparing precision for different embedding and clustering method combinations at the task of topic clustering, with a 95% Clopper-Pearson confidence interval.
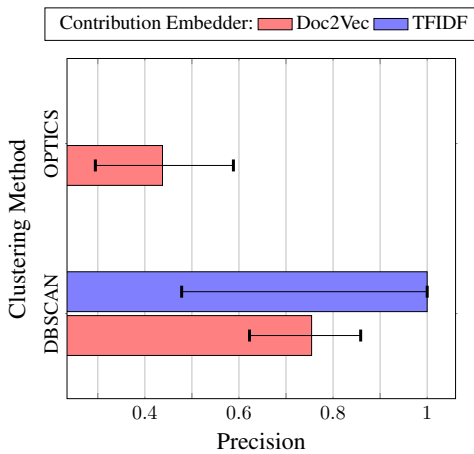

Figure 3. Comparing different embedding and clustering method combinations in the task of bridge user identification using precision, with a 95% Clopper-Pearson confidence interval.

considered as true bridges (verified by the coders), which correspond to 1% of total active users in the ECS. That percentage corresponds to the results of Doc2Vec-DBSCAN. We are reporting on those values because Doc2Vec-DBSCAN shows better results using the binomial proportion confidence intervals (BPCIs), shown in Figure 3. The low percentage of bridge users is not surprising considering the definition of bridge user that requires a user to actively learn or share knowledge (post in the KM platform) in at least two different CoPs. Additionally, literature [8] consistently reports low percentages of contributors within an ECS.

Additionally, looking at the content of clusters

Table 4. Number of bridge users by pipeline implementation

	Output as Bridges	True Bridges
D2V-DBSCAN	57	43
D2V-OPTICS	48	21
TFIDF-DBSCAN	5	5

generated from the pipeline, we can learn about the type of content shared to different CoPs via the KM platform by bridge users (results are shown in Table 5). We identified different message themes that communicated key knowledge, among them:

- *Events*, for documents that contain announcements related to invitations to talks, seminars, conferences, or abstracts submissions.

- *Sharing knowledge*, for documents that contain information about skills, expertise, or problem-solving statements.

- *Sharing experience*, for documents that contain user narratives of personal experience inside CoPs, projects, or organizations.

- *Sharing instructions (by the KM team)*, for documents that contain instructions given by the KM team to all platform users. Although the nature of this message can be classified as "sharing knowledge", doing so could misrepresent the total amount of knowledge sharing activity being supported by the KM platform.

Table 5. Shared-content types - D2V-DBSCAN implementation

Events	78% (209)
Sharing knowledge	8% (22)
Requests	4% (12)
Meeting time announcements	4% (10)
Questions	3% (7)
Sharing experience	1% (4)
Sharing instructions (by the KM team)	1% (4)

There were also sets of clusters that included messages with questions, requests (e.g., fill out surveys or participate in polls/ studies), or meeting time announcements without any other related information. Users linked to documents that belong to those themes were disqualified as bridge users since the content does not reflect a willingness to transfer key-knowledge.

Each contribution was assigned to one of these themes (see Table 5). Two coders attained "good" reliability after comparing at least 10% of the data (Cohen's kappa = 0.63), then a single coder completed

Table 6. Distribution of number of bridges in topic-related clusters. Clusters-TB: Clusters with true bridges

	D2V-DBSCAN		D2V-OPTICS		TFIDF-DBSCAN	
	All clusters	Clusters-TB	All clusters	Clusters-TB	All clusters	Clusters-TB
1	89	75	28	29	17	17
2	3	3	3	2	1	-
3	1	-	1	-	-	-
4	-	-	3	-	-	-
11	-	-	1	-	-	-

the task for the rest of the documents generated by the pipeline. In our dataset, "events" (78%) is the predominant content type shared by bridge users followed by a very low percentage of "sharing knowledge" activity (8%). This does not mean that knowledge sharing is not happening across the organization outside of the ECS. It only reflects the main use of the KM platform by the users under study. Nevertheless, our pipeline can identify documents and users linked to a collaborative action within an ECS.

Finally, in order to gain a deeper understanding of the bridge user's source of knowledge, we computed the frequency distribution of the number of bridges in topic-related clusters (see Table 6). These results show that the majority of clusters consist primarily of content shared by a single user. This means that the knowledge shared by the bridge user via the KM platform was likely acquired outside the KM platform. When we looked at the content in clusters with more than two bridge users, we noticed that those contributions were mostly invitations to seminars where it seemed those users were responsible for forwarding communications about them and a change in the steering team or leadership roles in the CoPs might have happened.

8. Discussion

The results of this work show a plausible automated solution for identifying bridge users and knowledge transfer behavior. First, it is possible to discover bridge users using our pipeline; however, its efficacy is data-dependent. In our dataset, there were no clear instances of a bridging contribution's knowledge source coming from the KM platform. However, it was possible to identify the users that share similar posts in different CoPs. Second, the topic clustering results allow us to identify how users utilize the KM platform. For our dataset, it appears that the KM platform is most used by bridge users to notify about key-knowledge related events. Finally, the bridge users list shows users who are engaging in collaborative actions within the ECS.

8.1. Implications for KM Platform Design

Answering the bridge user identification question presents many difficulties, many of which are a direct result of the limitations in current topic analysis techniques. Using semantic analysis to find implicit knowledge transference can be effective, but that effectiveness is bound by the ability to associate semantically similar text contributions. Given this limitation and the ease of an alternative solution, a resulting recommendation for designers and developers of KM platforms is to build in the UI features that allow users to self-report knowledge transference contributions. KM interfaces provide a shared space for capturing explicit knowledge but currently they do not allow content categorization that reflect whether the user-generated content is applied knowledge, knowledge gained from other contributions, or refers to sharing information about a knowledge-related event. This can be facilitated by implementing the following updates on KM platforms:

- Provide ways to categorize the knowledge that is being transferred or shared from one CoP to another one. This helps to understand what knowledge employees are more likely to transfer to different contexts and to share across CoPs.

- Keep track of bridge users and make the results accesible to KM evaluators. This helps to acknowledge employees' contributions. Also, identifying bridge users allows the study of real-world demographic characteristics and factors that can be used to facilitate users' collaborative actions or create incentives to increase bridge like behavior.

- Provide real-time access to statistics related to transference of knowledge. This can incentivize knowledge contribution codification.

8.2. Limitations

Our study evaluates the bridge user identification pipeline on the ECN dataset of a large multinational

engineering company. The literature suggests that ECS adopt the culture and usage patterns from the organization within which they are deployed [46]. This suggests that the semantic patterns observed and extrapolated from for bridge user identification may not exist as strongly in other ECN datasets. For this reason, a limitation of our research is that whether it is fully generalizable is still undetermined. Although, given the variety of ECN community variation within our dataset, we strongly believe that this pipeline can be reapplied elsewhere and will be useful to study other organizations' collaborative networks.

9. Conclusions and Future Work

Evaluation of knowledge-sharing activities within an organization calls for the identification of bridge users who facilitate knowledge flow. In this paper, we present a pipeline for automating the identification of bridge users that takes into account the semantic similarity of user-generated content. We tested different implementations of the proposed pipeline and validated the implementations by comparing the precision scores of their results.

The pipeline output is a list of potential bridge users and their bridging contributions. We showed the efficacy of our method by verifying actual bridge users and the precision based on that output. Having a list of potential bridges can serve organizations in many ways. For example, KM officers can use this list to demonstrate the use of the KM platform and to acknowledge employees for their contributions to the knowledge network. Additionally, management could include "being a bridge user" as an employee quality in the list of metrics considered in the employee promotion evaluation process. The concept of bridge users can be used in the KM evaluations, for example, instead of only asking employees to recall the instances in which they contributed in a CoP, they could be asked to self-report instances in which they have transferred knowledge from one CoP to another.

Moreover, we performed a thematic analysis of the documents linked to potential bridge users and identified the types of contributions users share among different CoPs. The identified themes inform KM agents about the usage of the KM platform.

Future research remains to be done on the identification of bridge users. Currently, the identification of knowledge transfer relies on the semantic similarity of user-generated content. However, one can identify different types of knowledge that employees share; one could investigate the capture of additional information that could support the identification of the types of knowledge transfer that happens in a contribution. Also, the current pipeline output produces a list of bridge users without considering any ranking. More work should be done to classify and rank bridge user perhaps considering user's expertise or some notion of contribution value.

References

[1] P. Schubert and J. H. Glitsch, "Use cases and collaboration scenarios: How employees use socially-enabled enterprise collaboration systems (ecs)," *Int. Journal of Information Systems and Project Management*, vol. 4, no. 2, pp. 41–62, 2016.

[2] T. S. Teo, R. Nishant, M. Goh, and S. Agarwal, "Leveraging collaborative technologies to build a knowledge sharing culture at hp analytics.," *MIS Quarterly Executive*, vol. 10, no. 1, 2011.

[3] N. Jeners and W. Prinz, "Metrics for cooperative systems," in *Proc. of the 18th Int. Conf. on Supporting Group Work*, pp. 91–99, ACM, 2014.

[4] S. Behrendt, A. Richter, and M. Trier, "Mixed methods analysis of enterprise social networks," *Comput. Netw.*, vol. 75, pp. 560–577, Dec. 2014.

[5] J. V. Hacker, F. Bodendorf, and P. Lorenz, "A framework to analyze enterprise social network data," in *Enterprise big data engineering, analytics, and management*, pp. 84–107, IGI Global, 2016.

[6] R. L. Cross, R. L. Cross, and A. Parker, *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business Press, 2004.

[7] F. Schwade and P. Schubert, "Social collaboration analytics for enterprise collaboration systems: Providing business intelligence on collaboration activities," in *Proc. of the 50th Hawaii Int. Conf. on System Sciences*, 2017.

[8] F. Schwade and P. Schubert, "Developing a user typology for the analysis of participation in enterprise collaboration systems," in *Proc. of the 52nd Hawaii Int. Conf. on System Sciences*, 2019.

[9] M. Muller, K. Ehrlich, T. Matthews, A. Perer, I. Ronen, and I. Guy, "Diversity among enterprise online communities: Collaborating, teaming, and innovating through social media," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '12, (New York, NY, USA), pp. 2815–2824, ACM, 2012.

[10] A. Lungeanu and N. S. Contractor, "The effects of diversity and network ties on innovations: The emergence of a new scientific field," *American Behavioral Scientist*, vol. 59, no. 5, pp. 548–564, 2015. PMID: 26576061.

[11] M. Alavi and D. Leidner, "Knowledge management systems: issues, challenges, and benefits," *Communications of the Association for Information systems*, vol. 1, no. 1, p. 7, 1999.

[12] W. R. King, "Knowledge management and organizational learning," in *Knowledge management and organizational learning*, pp. 3–13, Springer, 2009.

[13] A. Aurum, F. Daneshgar, and J. Ward, "Investigating knowledge management practices in software development organisations–an australian experience," *Information and Software Technology*, vol. 50, no. 6, pp. 511–533, 2008.

[14] F. O. Omotayo, "Knowledge management as an important tool in organisational management: A review of literature," *Library Philosophy and Practice*, vol. 1, no. 2015, pp. 1–23, 2015.

[15] R. McDermott, "Learning across teams," *Knowledge Management Review*, vol. 8, no. 3, pp. 32–36, 1999.

[16] E. Wenger, R. A. McDermott, and W. Snyder, *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press, 2002.

[17] C. S. Greeven and S. P. Williams, "Enterprise collaboration systems: addressing adoption challenges and the shaping of sociotechnical systems," *Int. Journal of Information Systems and Project Management*, vol. 5, no. 1, pp. 5–23, 2017.

[18] I. Buunk, C. F. Smith, and H. Hall, "Tacit knowledge sharing in online environments: Locating bawithin a platform for public sector professionals," *Journal of Librarianship and Information Science*, p. 0961000618769982, 2018.

[19] L. Holtzblatt, J. L. Drury, D. Weiss, L. E. Damianos, and D. Cuomo, "Evaluating the uses and benefits of an enterprise social media platform," *Journal of Social Media for Organizations*, vol. 1, no. 1, pp. 1–21, 2013.

[20] A. Powell, G. Piccoli, and B. Ives, "Virtual teams: A review of current literature and directions for future research," *SIGMIS Database*, vol. 35, no. 1, pp. 6–36, 2004.

[21] C. Steinfield, J. M. DiMicco, N. B. Ellison, and C. Lampe, "Bowling online: social networking and social capital within the organization," in *Proc. of the 4th Int. Conf. on Communities and technologies*, pp. 245–254, ACM, 2009.

[22] E. Wenger, B. Trayner, and M. De Laat, "Promoting and assessing value creation in communities and networks: A conceptual framework," 2011.

[23] S. Lee, Y. Seog Kim, and E. Suh, "Structural health assessment of communities of practice (CoPs)," *Journal of Knowledge Management; Kempston*, vol. 18, no. 6, pp. 1198–1216, 2014.

[24] M. J. Muller, J. Freyne, C. Dugan, D. R. Millen, and J. Thom-Santelli, "Return on contribution (roc): A metric for enterprise social software," in *ECSCW 2009*, pp. 143–150, Springer, 2009.

[25] R. Helms and K. Buijsrogge, "Knowledge network analysis: a technique to analyze knowledge management bottlenecks in organizations," in *16th Int. Wksh on Database and Expert Systems Applications (DEXA)*, pp. 410–414, IEEE, 2005.

[26] F. Probst, L. Grosswiele, and R. Pfleger, "Who will lead and who will follow: Identifying influential users in online social networks," *Business & Information Systems Engineering*, vol. 5, no. 3, pp. 179–193, 2013.

[27] J. V. Hacker, F. Bodendorf, and P. Lorenz, "A framework to identify knowledge actor roles in enterprise social networks," *Journal of Knowledge Management*, vol. 21, no. 4, pp. 817–838, 2017.

[28] Z. Xu, J. Ramanathan, and R. Ramnath, "Identifying knowledge brokers and their role in enterprise research through social media," *IEEE Computer*, vol. 47, no. 03, pp. 26–31, 2014.

[29] R. Cross, S. P. Borgatti, and A. Parker, "Making invisible work visible: Using social network analysis to support strategic collaboration," *California management review*, vol. 44, no. 2, pp. 25–46, 2002.

[30] K. Berger, J. Klier, M. Klier, and A. Richter, "Who is key?-value adding users in enterprise social networks," in *22nd European Conference on Information Systems*, pp. 1–16, 2014.

[31] J. Viol, R. Bernsmann, and K. Riemer, "Behavioural dimensions for discovering knowledge actor roles utilising enterprise social network metrics," *arXiv preprint arXiv:1606.00883*, 2016.

[32] M. Ramanathan, A. Zhang, Y.-R. Cho, and W. Hwang, "Bridging centrality: Identifying bridging nodes in scale-free networks," 2006.

[33] K. Mäkelä, W. Barner-Rasmussen, M. Ehrnrooth, and A. Koveshnikov, "Potential and recognized boundary spanners in multinational corporations," *Journal of World Business*, vol. 54, no. 4, pp. 335–349, 2019.

[34] J. Birkinshaw, T. C. Ambos, and C. Bouquet, "Boundary spanning activities of corporate hq executives insights from a longitudinal study," *Journal of Management Studies*, vol. 54, no. 4, pp. 422–454, 2017.

[35] A.-M. Søderberg and L. Romani, "Boundary spanners in global partnerships: a case study of an indian vendors collaboration with western clients," *Group & Organization Management*, vol. 42, no. 2, pp. 237–278, 2017.

[36] D. G. Ancona and D. F. Caldwell, "Bridging the boundary: External activity and performance in organizational teams.," *Administrative science quarterly*, vol. 37, no. 4, 1992.

[37] D. G. Ancona and D. F. Caldwell, "Beyond task and maintenance: Defining external functions in groups," *Group & Organization Studies*, vol. 13, no. 4, pp. 468–494, 1988.

[38] S. D. Pawlowski and D. Robey, "Bridging user organizations: Knowledge brokering and the work of information technology professionals," *MIS quarterly*, pp. 645–672, 2004.

[39] G. Currie and L. White, "Inter-professional barriers and knowledge brokering in an organizational context: the case of healthcare," *Organization Studies*, vol. 33, no. 10, pp. 1333–1361, 2012.

[40] K. Olejniczak, E. Raimondo, and T. Kupiec, "Evaluation units as knowledge brokers: Testing and calibrating an innovative framework," *Evaluation*, vol. 22, no. 2, pp. 168–189, 2016.

[41] F. Buccafurri, V. D. Foti, G. Lax, A. Nocera, and D. Ursino, "Bridge analysis in a social internetworking scenario," *Inf. Sci.*, vol. 224, pp. 1–18, 2013.

[42] S. Bolasco and P. Pavone, "Multi-class categorization based on cluster analysis and tfidf," *9th Int. Conf. on Statistical Analysis of Textual Data*, 2008.

[43] P. S. Karvelis, D. Gavrilis, G. K. Georgoulas, and C. D. Stylios, "Topic recommendation using doc2vec," *Int. Joint Conf. on Neural Networks (IJCNN)*, pp. 1–6, 2018.

[44] M. R. E. Symonds and A. Moussalli, "A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using akaikes information criterion," *Behavioral Ecology and Sociobiology*, vol. 65, pp. 13–21, 2010.

[45] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval Estimation for a Binomial Proportion," *Statistical Science*, vol. 16, no. 2, pp. 101–117, 2001.

[46] K. M. Markman, "A networked self: Identity, community and culture on social network sites," *New Media & Society*, vol. 14, pp. 1240–1242, 2012.