

# Recommending Social Media Content to Community Owners

Inbal Ronen\*, Ido Guy\*\*, Elad Kravi\*\*\*, Maya Barnea\*

\*IBM Research-Haifa, Israel {inbal, mayab}@il.ibm.com

\*\*Yahoo! Labs, Haifa, Israel idoguy@acm.org<sup>1</sup>

\*\*\*Department of Computer Science, Technion, Haifa, Israel ekravi@cs.technion.ac.il

## ABSTRACT

Online communities within the enterprise offer their leaders an easy and accessible way to attract, engage, and influence others. Our research studies the recommendation of social media content to leaders (owners) of online communities within the enterprise. We developed a system that suggests to owners new content from outside the community, which might interest the community members. As online communities are taking a central role in the pervasion of social media to the enterprise, sharing such recommendations can help owners create a more lively and engaging community. We compared seven different methods for generating recommendations, including content-based, member-based, and hybridization of the two. For member-based recommendations, we experimented with three groups: owners, active members, and regular members. Our evaluation is based on a survey in which 851 community owners rated a total of 8,218 recommended content items. We analyzed the quality of the different recommendation methods and examined the effect of different community characteristics, such as type and size.

**Categories and Subject Descriptors:** H.3.3 [Information Search and Retrieval]: Information filtering

**Keywords:** Enterprise, group recommendation, online communities, recommender systems, social media.

## 1. INTRODUCTION

Online communities are increasingly playing a central role in the proliferation of social media within the enterprise [13]. A growing number of blog posts, status update messages, wiki pages, shared files, and forum topics are created and shared as part of enterprise communities, making them a key entry gate to enterprise social media [35]. Matthews et al. [31] report the existence of 111,577 communities with 487,941 distinct members within IBM's social media environment, indicating that almost every employee is a member of at least one community. Our own measurement within IBM indicates that over half of the social media activities performed at the time of this research were in the context of a community. This is a dramatic change compared to just three years ago, when only 10.35% of the activity belonged to enterprise online communities [17].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

SIGIR'14, July 6–11, 2014, Gold Coast, Queensland, Australia.  
Copyright © 2014 ACM 978-1-4503-2257-7/14/07...\$15.00.  
<http://dx.doi.org/10.1145/2600428.2609596>

Online communities pose many potential benefits to an enterprise, such as promoting collaboration and knowledge sharing [14,41], enhancing coordination and execution of different tasks [25], nurturing innovation [14,35], and more (see [31] for more details). Community *owners*, sometimes referred to as moderators or leaders, are a key factor in facilitating these benefits [8,24,25,26,42]. They are responsible for starting up the community, keeping it alive and engaged, and adding relevant members and content.

In spite of the owners' critical role in fostering the success of these communities, literature on systems that assist community owners is sparse. In a recent paper, Matthews et al. [31] state that virtually no research has been done on tools for community owners. In that work, the authors introduced a tool that surfaces various metrics to help owners maintain a healthy community. Xu et al. [42] recently presented a visualization tool enabling owners to make sense of activity that takes place within their communities.

In this work, we propose a recommender-systems approach that suggests relevant content items owners may want to share with their communities. Sharing content with a community can help increase participation and engagement. As mentioned by Xu et al. [42], intervention through contributing content is one of the most important roles of community owners, as they guide members towards achieving community goals. Sharing content that is external to a community can also help identify more common ground for the members, increase their knowledge base, and help members sift through the ever-growing overload of social media information. Furthermore, authors of the content items could be potentially invited to join the community as new members.

Previous work has addressed recommending social media content items to individual users of enterprise social media [18]. The task of recommending to community owners is different, since the recommendation must also take into account the characteristics and needs of the community as a whole, rather than just the owner's individual interests. Upon receiving a recommendation, owners must evaluate the recommendation both from their own perspective and from the community's perspective. The latter aspect is tied to the area of group recommendation [10,21,29], which aims at recommending items to groups rather than to individuals, taking into account the preferences of the group as a whole (more details in Section 2). To the best of our knowledge, this is the first study to examine this type of recommendation to community owners.

We experimented with various techniques for producing recommendations, which take into consideration both the

---

<sup>1</sup> Research was conducted while working at IBM Research

community's content and its members. The first method we used was based on the community's content, as reflected in its title, summary, and tags. The second method experimented with approaches of defining a community's interest profile based on the aggregation of individual interest profiles of the community's members or a subset of these. We examined three such groups: a set of individuals randomly sampled from all the community's members (including owners), all owners of the community, and active members of the community (either owners or members). Finally, we examined hybridization of each of these member-based profiles with the content profile, leading to a total of seven profiles in our experiments—one content-based, three member-based, and three hybrid. This choice of profiles aimed to address the following research questions: (1) compare content-based and member-based methods and inspect which is more effective for recommendation to community owners; (2) examine which set of community members best represents the community when it comes to producing recommendations; and (3) examine whether hybridizing content-based and member-based methods improves the recommendations.

Our evaluation was based on a large survey of community owners, covering 851 owners and 796 communities. Each participating community in the survey was assigned to one of the seven profiles described above. The owners received ten recommendations based on that profile and one random recommendation, serving as a weak baseline. Recommendations included social media content, such as blog entries, shared files, and wiki pages that were not part of the community. Participants were asked to rate each recommendation with regards to both their own interest and the interest expected from their community. In our analysis, we compared the ratings of the recommendations generated using the different methods, while also inspecting the influence of different characteristics of the communities (type, size) and the type of recommended items on the ratings.

In our results, we found that owners ranked their self-interest in recommended items higher than the interest they expected the community to express in them; this indicates the complexity involved in recommending content for a community. Indeed, the ratings for both the owners' self-interest and that perceived for the community were considerably lower than previously reported interest ratings for personal recommendations. Rating comparisons indicated that all profile methods yielded significantly higher ratings than the random baseline. Moreover, they revealed that considering active members was the more effective member-based method for representing the community's interests and also outperformed the content-based method. Hybridization of active member-based and content-based profiles further improved recommendation quality. Additional analysis discovered that the effectiveness of the active member group stemmed from its high ratings for large communities with over 100 members. For smaller communities, we did not observe a difference between the active members and the regular members.

## 2. RELATED WORK

Our work ties to the area of group recommendation, which aims at recommending content items of interest to groups rather than to individuals. Group recommendations take into account the preferences of the entire group, considering both item preferences of the group members [21,29] and various other characteristics of the group such as homogeneity, relationships among the members, influence, and size [7,22,40]. Group recommendations have been

applied in various domains, such as movies or television programs [5,16,30,37,43], music [33], travel [4,22], recipes [7], and others.

Many papers have addressed the challenge of modeling a group's preferences and, in particular, how to aggregate the preferences of the individual members of the group. Aggregation strategies can be based on aggregating user preferences into one pseudo-profile that represents the group (e.g., [30,40,43]) or on aggregating the recommendations for individual members into one list (e.g., [1,4,5,37]). Senot et al. [40] tried to identify group factors, such as size, for automatically selecting the right aggregation strategy, but found that such factors do not have enough impact to justify an automatic approach. Berkovsky et al. [7] investigated different aggregation strategies for recommending recipes to groups and found that the best performance is obtained when individual user models are aggregated into a group model. Our member-based profiles are also based on profile-level aggregation of individual members. Gartell et al. [15] assigned different weights to members according to their influence on the group. We do not assign such weights, but experiment with three types of members, including owners, active members, and a subset of all members. As opposed to classic group recommendation, our direct target is not the group as a whole, but rather the owner, as a "proxy" to the group.

Another aspect of group recommendation is the generation of a preference or interest profile for an individual group member. Most approaches range from collaborative filtering [5,7,12,16,24] to content-based [3] and hybrid approaches [23]. Typically, hybrid approaches improve the overall performance [9]. Our approach is also hybrid and combines recommendations based on related people and related tags, using a model that was found successful for recommending social media content to individual users [18].

Our work focuses on recommendations to online communities in the enterprise. Ebrahim et al. [14] reviewed literature on virtual teams in the enterprise. Matthews et al. [31] and Muller et al. [35] examined online communities in the intranet and how community owners can enhance the value of their communities. Online communities can be implicit, identified by the system (e.g., [27,36]), or explicit through membership (e.g., [31]). Our work examines explicitly defined online communities in a large enterprise.

The target population of our recommendations is community owners. Xu et al. [42] stated that current community tools give very little direct support to community owners. Their work enabled owners to assess the performance of their communities by visually comparing them to other communities. Community Insights [31] provided actionable analytics that helped community owners foster healthy communities. Six owner needs were identified including: people, content, participation, sociability, leadership, and comparable communities—the first three being the most critical. One of the main actions proposed in that paper is the need to identify critical content for the community by owners.

Muller et al. [35] examined different types of enterprise communities and how they make use of social media tools. Specifically, they identified the following five community types: communities of practice, which are a group of people with a common interest or practice; teams that represent communities working on a shared goal for a particular client, project, or business function; technical support communities for a particular technology; idea labs communities in which members brainstorm around a set of questions or issues for a limited period of time; and recreation communities devoted to leisure activities unrelated

to work. In our survey, we asked the community owners to catalogue their community into one of these types, in order to identify whether recommendations could be useful for a particular type of community.

### 3. RECOMMENDER SYSTEM

#### 3.1 Research platform

This study was conducted over a deployment of IBM Connections [20] within IBM. IBM Connections is a social media platform, which enables employees to collaborate by using services such as blogs, wikis, file sharing, discussion forums, and status updates. Employees can create communities around topics or activities for collaboration and knowledge sharing [35]. Communities can be private, invitation-only (content is public, joining by invitation), or public. In this study, we focused on invitation-only and public communities, both visible to all employees. Each community includes all the social media services provided by IBM Connections. Therefore, a status update, a blog entry, a forum topic, a file, or a wiki page can all be shared in the context of a specific community, while all of that community’s members can participate. Communities define two user roles, *owners* and *members*. Members can post and view content, but may only edit their own content, whereas owners can also configure the community, edit any content, and manage the list of members. Communities range in size from a few members to tens of thousands of members. Some have very few owners, while others assign owner privileges to many members. A community owner may not necessarily be its formal leader [31]. Moreover, there is no indication in the system of who the formal leader of the community is. At the time of this research, the deployment of IBM Connections included nearly 200,000 communities, of which about 100,000 were public or invitation-only.

#### 3.2 Community Interest Profiles

Recommendations were generated based on an interest profile of the community. We examined seven interest profiles, which were either member-based, content-based, or hybrid. Member-based profiles (MBPs) represented the interest profile of the community members or a subset of these. The content-based profile (CBP) modeled the community’s interests according to its title and metadata. The hybrid approach tried to benefit from both worlds by combing the interests of the members with the content of the community.

##### 3.2.1 Member-based Profiles

We examined three types of MBPs that were generated based on different subsets of community members. The *Members* profile was based on 50 random members (including owners) of the community, or on all members of the community, if it had 50 members or less. The *Owners* profile considered all the owners of the community. The *Actives* profile was based on the set of members (including owners) who contributed at least once to the community along its lifetime. Contributions included, for example, authoring a blog entry, editing a wiki page, posting a status update, sharing a file, or writing a forum reply.

Our method for generating the MBPs builds on the method used for recommending social media items to individuals [18]. Individual profiles consisted of *profile elements* that included related people, denoted by  $P$ , and related tags, denoted by  $T$ . In our experiments, we set  $|P|=|T|=30$ , as in [18]. For the community profile, we considered the individual profile of each of the members and aggregated them into a single community profile,

which itself included 30 people ( $P_c$ ) and 30 tags ( $T_c$ ). The aggregation of multiple individual profiles into one community profile was based on the number of members whose profiles contained each profile element and the relative position (rank) of the profile element in each of these profiles. For tags, we also considered stemming and inverse document frequency. Below, we describe the method in more detail.

Let  $M$  denote the set of members an aggregated profile was based on. For each member  $m \in M$ , we computed an individual profile, denoted  $prof(m)$ , which included the top  $|P|$  related people and top  $|T|$  related tags, ranked by their relationship strength to that member. Related people were calculated and ranked based on familiarity relationships reflected in social media, such as explicit “friending”, wiki page co-editing, file sharing, and others, as well as similarity relationships, such as bookmarking of the same pages, usage of the same tags, membership in the same communities, and others. Related tags included tags used by the member to annotate different entities as well as tags that were assigned to her by others within an enterprise people tagging application. Full details of the individual profile calculation can be found in [18].

Given these individual profiles, the list of  $|P_c|$  people to be included in the aggregated profile for  $M$  was determined according to the following scoring formula:

$$score(p, m) = \begin{cases} 2|P| - rank_{prof(m)}(p) & p \in prof(m) \\ 0 & p \notin prof(m) \end{cases}$$

$$score(p, M) = \begin{cases} \sum_{m \in M} score(p, m) & count(p, M) \geq 2 \\ 0 & count(p, M) < 2 \end{cases}$$

where  $count(p, M)$  denotes the number of members in  $M$  that included person  $p$  in their individual profile, and  $rank_{prof(m)}(p)$  denotes the rank of person  $p$  out of all  $|P|$  people included in the individual profile of a member  $m \in M$ . The rank of the top person in the profile would be 0, the second person would get a rank of 1, and so on. Note that the measure  $2|P| - rank_{prof(m)}(p)$ , assigns the top person with a score that is almost double the score of the bottom person in the profile –  $2|P| - 0 = 60$  versus  $2|P| - (|P| - 1) = |P| + 1 = 31$ , respectively, in our case. This is done to limit the influence of the rank within an individual profile up to a factor of two. For example, a person who appears at the bottom of two members’ profiles would get a higher score than a person who appears at the top of only one member’s profile, with 62 versus 60. Ultimately, we summed these scores across all members in  $M$ , considering people who appeared in at least two member profiles, to make sure they had at least two different member “votes”, and selected the top  $|P_c|$  according to their score. Thus, the more members in  $M$  a person is related to and the stronger the relationship to them is, the higher the chances of that person to be included in the aggregated community profile. Finally, we normalized all scores by the score of the top person in the profile.

The list of  $|T_c|$  tags in the aggregated profile was calculated in a similar manner, with two adaptations addressing the need for stemming and for penalizing popular tags, which tend to be very broad and less meaningful. We first applied stemming [28] in order to merge similar forms of tags, such as “travel”, “traveler”, and “traveling”. The score of a stemmed tag  $t$  for the aggregated profile of  $M$  was calculated according to the following formula:

$$score(t,m) = \begin{cases} 2|T| - \min_{t' \in stem(t,m)} (rank_{prof(m)}(t')) & t \in prof(m) \\ 0 & t \notin prof(m) \end{cases}$$

$$score(t,M) = \begin{cases} idf(t) \cdot \sum_{m \in M} score(t,m) & count(t,M) \geq 2 \\ 0 & count(t,M) < 2 \end{cases}$$

where  $stem(t,m)$  denotes the set of tags in the profile of a member  $m$  that convert into tag  $t$  after stemming. Analogously to the people case,  $count(t,M)$  denotes the number of members whose profiles include  $t$  and  $rank_{prof(m)}(t')$  denotes the rank of a non-stemmed tag  $t'$  out of the  $|T|$  tags included in the profile of a member  $m$ . Finally, the inverse document frequency of a stemmed tag  $t$ ,  $idf(t) = \ln(N/N_t)$ , is computed as the logarithm of the ratio between the total number of documents in the system ( $N$ ) and the number of documents tagged with at least one tag that converts into  $t$  after stemming ( $N_t$ ). Similar to the vector-space  $idf$  score for terms [28], the  $idf$  score for tags penalizes popular tags, which are related to many documents. The total score of the stemmed tag was calculated by summing the scores over all members, for tags that appeared in the profiles of at least two members. The top  $|T_c|$  tags with highest scores were then selected for the aggregated profile, with their scores normalized by the highest value. Intuitively, a tag would have higher chances of being included in the aggregated profile if it is related to more members in  $M$ , if the relationship to each of these members is stronger, and if the tag is generally less common.

### 3.2.2 Content-based Profiles

The CBP considered the community's title, summary, and tags. We used the  $KL+TB$  measure [11] to identify the most significant terms in the extracted content. This method was previously found effective for term extraction from concise social media content [11]. The method uses the Kullback-Leibler ( $KL$ ) measure, which is a non-symmetric distance measure between two given distributions. In our case, we sought out terms, in their stemmed form, which maximize the  $KL$  divergence between the language model of the community's content and the language model of the entire community collection's content. On top of the  $KL$  statistical score, we applied a tag-boost ( $TB$ ), which promotes keywords that are likely to appear as tags, based on a given well-tagged folksonomy. For this purpose, we used the folksonomy generated by the IBM Connections' bookmarking application [34].

Ultimately, a community's content profile included all terms that had a  $KL+TB$  score that was at least 30% of the maximum  $KL+TB$  score of a term in that community. We experimented with various other thresholds, but found 30% to yield the best trade-off between the overall number of extracted terms and their quality.

### 3.2.3 Hybrid Profiles

We hybridized each of the three MBPs with the CBP by considering both the people and tags included in the MBP and the terms included in the CBP. Accordingly, the *MembersContent*, *OwnersContent*, and *ActivesContent* profiles were defined, consisting of people and tags from the MBP and content-terms from the CBP. We further describe how recommendations were generated for the hybrid profiles in the next section.

## 3.3 Item Recommendation

Given a community profile, we generated recommendations by issuing a query containing the profile elements to a social search

system [38], similar to the way it was done for an individual profile in [18]. The social search system, which is built on top of Lucene [32], indexes social media documents of different types, including blog entries, wiki pages, shared files, forum threads, activities, and bookmarks (see [35] for more details on each of these types). The system maps the relationships among these documents, related terms and tags, and related people, in a way that makes all types of entities both searchable and retrievable [38]. For the task of producing recommendations, the query to the social search system included a combination of people, tags, and terms, while the results were documents that matched the query, ordered by their relevance score. Below we describe in more detail the queries and the calculation of the relevance score.

For the non-hybrid profiles, we retrieved the top 100 documents by issuing an OR query to the social search system. This query included all the profile elements as its arguments, each boosted with its corresponding score, calculated as explained in the previous section. For a profile that included people  $p_1 \dots p_u$  with scores  $s(p_1) \dots s(p_u)$  and tags  $t_1 \dots t_v$  with scores  $s(t_1) \dots s(t_v)$ , we issued the following query:

$$q = (p_1 \wedge s(p_1) \vee \dots \vee p_u \wedge s(p_u)) \vee (t_1 \wedge s(t_1) \vee \dots \vee t_v \wedge s(t_v))$$

The symbol  $\wedge$  denotes the boosting factor.

For a hybrid profile, consisting of an MBP with people  $p_1 \dots p_u$  scored by  $s(p_1) \dots s(p_u)$  and tags  $t_1 \dots t_v$  scored by  $s(t_1) \dots s(t_v)$ , and a CBP with content-terms  $c_1 \dots c_r$  scored by  $s(c_1) \dots s(c_r)$ , recommendations were created by issuing the following query to the social search system:

$$q = ((p_1 \wedge s(p_1) \vee \dots \vee p_u \wedge s(p_u)) \vee (t_1 \wedge s(t_1) \vee \dots \vee t_v \wedge s(t_v))) \wedge (c_1 \wedge s(c_1) \vee \dots \vee c_r \wedge s(c_r))$$

The query retrieved the top 100 documents that were relevant to at least one person or tag from the MBP and one content-term from the CBP. This way we made sure that the returned documents matched both parts of the hybrid profile.

Upon receiving a query  $q$ , the relevance score of a document  $d$  in the social search system was calculated as follows:

$$RS(d,q) = e^{-\alpha\tau(d)} \cdot [\beta \sum_{i=1}^u s_q(p_i) \cdot w(d,p_i) + \gamma \sum_{j=1}^v s_q(t_j) \cdot w(d,t_j) + (1 - \beta - \gamma) \sum_{k=1}^r s_q(c_k) \cdot w(d,c_k)]$$

Notice that the third and final element of the summation is only relevant for hybrid profiles, otherwise it was disregarded. In the equation,  $\tau(d)$  denotes the time in days since the creation date of  $d$ ;  $\alpha$  is a time-decay factor, used to promote fresher documents (set in our experiments to 0.025, as in [18]);  $\beta$  and  $\gamma$  are parameters that control the relative weight among people, tags, and content-terms. In our experiments, we set both to 1/3, giving equal importance to all ingredients;  $S_q(p_i)$ ,  $S_q(t_j)$  and  $S_q(c_k)$  are the scores of the respective profile elements, given as part of the query  $q$ ; and  $w(d,p_i)$ ,  $w(d,t_j)$ , and  $w(d,c_k)$  denote the relevance score of the document to the person, tag, or content-term, as calculated by the social search system (see more details in [18,38]).

Ultimately, we selected the top 10 items for recommendation in our survey after applying the following two steps over the 100 retrieved documents: (1) *filtering*: documents that were already published in the community were filtered out; (2) *diversifying*: in order to promote diversity across document types (blog entry,

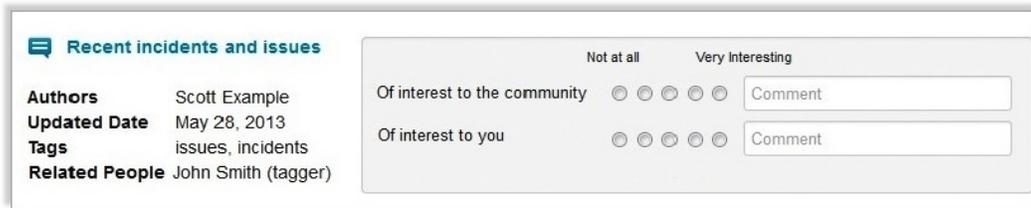


Figure 1. Sample recommendation.

wiki page, etc.), we used the type as the first sorting criterion and the relevance score only as a secondary criterion. Therefore, we first took the top document of each type, if such existed among the top 100 documents, and ordered these by their relevance score. We then took the second of each type, if one existed, and ordered this group by the relevance score, and so forth until we reached 10 items (documents) in total. Finally, we randomized the order of all 10 recommendations.

## 4. Experimental Setup

### 4.1 Survey Description

Our evaluation was based on a large user survey, where community owners were asked to rate recommended items. On the first page of the survey, owners were presented with four general questions about their community. The first question asked them to select the community type according to the categories described in [35]: Community of Practice (COP), Team, Technical Support, Recreation, Idea Lab, and Other, as detailed in Section 2. In the next two questions, participants indicated whether they felt engaged in the community and whether the community was engaged as a whole (Very Engaged, Engaged, Minimally Engaged, Not at all). The last question referred to the activity level the owner would like to establish in their community (Daily, Weekly, Monthly, Sporadic).

In the second and main phase of the survey, each owner was presented with 11 recommended items. 10 recommended items were selected according to one of the seven profiles, as described in Section 3, while an extra item was randomly selected from the social search index to serve as a lower-bound baseline. The position of the random item within the list of 11 items was randomly drawn for each owner.

Figure 1 shows an example of a recommendation as presented in our survey. Each recommended item included an icon that represented its type, its title with a link to the original entry in IBM Connections, the names of the authors, the last-update date, and up to 5 related tags and 5 related people, if existed. Related tags included tags that had been directly assigned to the item in IBM Connections. Related people were individuals apart from the authors who had performed actions on the item, such as commenting, editing, or sharing. For each item, the owner was asked to provide a rating on a 5-point Likert scale, ranging from “Not at all” to “Very Interesting”, regarding the community’s interest in the item and the owner’s self interest in it.

We sent the survey to community owners via email. The message included a general description of the survey, with a link to the online community and to the web page of the survey. The subject included the title of the community, as owners could get multiple messages for different communities they owned. Owners had the option of submitting only partial feedback.

### 4.2 Survey Participants

We sent out the survey to owners of communities that had a certain level of activity during the period of two months preceding the survey. As there were many inactive communities in the system, we limited ourselves to those with at least some activity by their members. We assumed that owners of these communities would have more interest in discovering new content for their communities and in increasing the activity level, whereas owners of inactive communities would be less interested, especially as the platform does not provide a good way to notify members of new content. Only active contributions, such as creating a blog entry, commenting on a forum thread, adding a bookmark, and similar actions were taken into account. Just viewing content was not accounted for since the system does not publish this data. The final set included communities with at least six activities during the two-month period, at least five members, and at least two owners.

For each of the resulting communities, we sent the survey to at most three owners. We also made sure an owner would get at most three surveys for different communities. When choosing the owners, we first chose randomly out of those who had been active in the community in the past. If there were fewer such owners, we chose randomly out of the remaining owners. The rationale behind this process was that active owners would be more likely to participate in the survey. Each community in the sample was randomly assigned to one of the seven profiles described before.

## 5. RESULTS

### 5.1 Participation

We received 907 responses to our 7,592 survey invitations (12%). These responses cover a total of 851 distinct owners of 796 different communities. Overall, 12.7% of the communities were covered by 2 owners and 0.63% were covered by 3 owners; 5.17% of the owners responded for 2 communities and 0.71% responded for 3 communities. These numbers indicate that most communities had one owner responding for them and most owners responded for one community. In the rest of this section, we will refer separately to each response from an owner for a specific community.

Table 1 presents general statistics of the 796 communities that were covered in our survey, including size (total number of members, owners included); number of owners; number of active members, as defined in the previous section; activity, measured as the number of contributions in the two months preceding the survey; and age (number of days since community creation). As can be seen, our participating communities had highly diverse characteristics.

For these communities, the average portion of owners who were also active in the community was 43.8% (stdev: 28.1%, median: 36.4%), while the average portion of active members who were

**Table 1. General characteristics of the 796 communities**

|         | Size    | Owners | Actives | Activity | Age   |
|---------|---------|--------|---------|----------|-------|
| Average | 478.7   | 7.42   | 4.91    | 23.2     | 568.6 |
| Stdev   | 1,841.8 | 7.1    | 9.41    | 33.59    | 484.6 |
| Median  | 76      | 5      | 3       | 13       | 433.5 |
| Min     | 5       | 2      | 1       | 6        | 32    |
| Max     | 35,090  | 57     | 179     | 381      | 2077  |

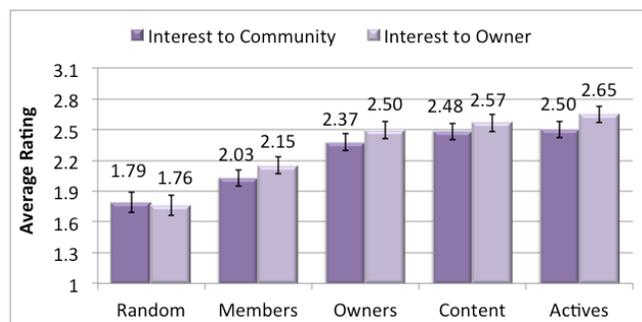
owners was 55.8% (stdev: 33.1%, median: 50%). This indicates that while there is some decent overlap between the owners and active members in a community, they are also quite different. Our experiment aimed, among other things, to explore whether and how this difference affects the ability of these groups to produce recommendations for the community.

## 5.2 Profile Comparison

As a single measure for interest in a group of recommended items, we opted to use the average rating of the items in the group. This captures the differences across the entire 1-to-5 Likert scale as opposed to, for example, taking just the proportion of items that were rated 4 and 5. Our results also indicate the 95% confidence intervals (CIs) for the rating averages. In the following, we refer to the rating of the owners expressing their own interest in a recommended item as “owner rating” and the interest they expressed for their community as “community rating”.

Generally comparing the owner ratings with community ratings indicates that the two are highly correlated (Pearson coefficient of 0.85). For 85.1% of the rated items, their owner and community ratings were equal, for 8.4% owner ratings were higher, and for 6.5% community ratings were higher. The overall owner rating average was significantly higher than the community rating average at 2.56 compared to 2.44 (one-tailed unpaired t-test,  $p < .001$ ). In spite of the fact that our recommendations were tailored for the community, the owner’s self-interest was higher in general than the expected community interest, emphasizing the challenge (perceived by owners) in recommending content to a whole community. We note that owner interest ratings are still substantially lower than ratings found for recommendations that were specifically targeted for individuals [18].

Figure 2 shows the rating average for the four non-hybrid (“pure”) profiles and the random baseline, both in terms of interest to the community and interest to the owner, as rated by the participating owners. It can be seen that with the exception of the random baseline, owner interest ratings are consistently higher than community interest ratings. Differences between interest to the owner and interest to the community were found significant for all



**Figure 2. Average ratings of interest to communities and to owners for the four pure profiles and the random baseline.**

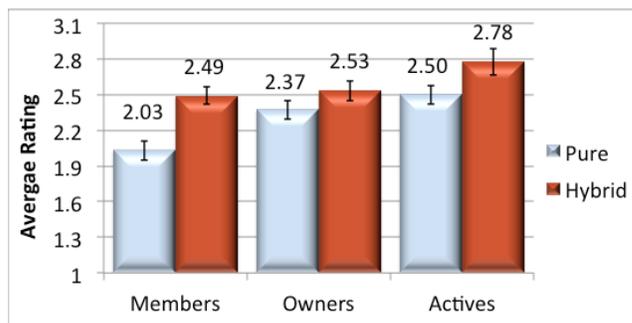
three member-based profiles (one-tailed unpaired t-test,  $p < .05$ ), while insignificant for the *Content* profile (one-tailed unpaired t-test,  $p = .08$ ) and for the *Random* items (one-tailed unpaired t-test,  $p = .32$ ).

Of the three member-based profiles, the *Actives* profile produces the most highly rated results and is the only MBP to outperform the *Content* profile. The *Members* profile produces the lowest results of all profiles and its ratings are only moderately higher than for *Random* items. A one-way ANOVA indicates that community ratings across the five groups were significantly different,  $F(4,5129) = 23.73$ ,  $p < 0.001$ . Tukey post-hoc comparisons indicate that the average rating for the *Random* items was significantly lower than for all other groups, except for the *Members* group. Average rating for the *Members* profile was significantly lower than for the *Owners*, *Content*, and *Actives* groups, while differences among the last three groups were insignificant. For owner ratings, a one-way ANOVA with Tukey post-hoc comparisons indicates that average rating for *Random* items was significantly lower than for all other four groups, while average rating for the *Members* profile was significantly lower than for *Owners*, *Content*, and *Actives*,  $F(4,5068) = 26.52$ ,  $p < .0001$ .

From this point on, unless stated otherwise, our results focus on ratings in terms of interest to the community and not the owner.

Overall, out of the 724 participants who rated all 10 non-random recommendations, 69.1% rated at least one as interesting to the community (either 4 or 5 on the Likert scale), 52.5% rated at least two, 39.5% at least three, and 20.2% rated at least five out of 10 as interesting for the community. These numbers indicate that while the accuracy of recommendations to owners for their community may not be as high as recommendations for individuals, a batch of recommended content is quite likely to include a few “good” items. In the rest of the section, we will try to understand what factors may further increase the portion of interesting recommendations.

Figure 3 shows the average ratings for the three pure member-based profiles compared to the hybrid profiles, which combine each member-based profile with the *Content* profile. Hybridization is shown to improve each of the pure profiles, both compared to the pure member-based and the pure *Content* profile (the latter’s average was 2.48). Hybridization with content is especially effective when applied to the *Members* and *Actives* profiles and less effective for *Owners*, perhaps indicating that owners produce more similar items to the ones produced based on the content. The differences between pure and hybrid profiles across all three types were found to be significant (one-tailed unpaired t-test,  $p < .001$ ).



**Figure 3. Average ratings for pure vs. hybrid profiles.**

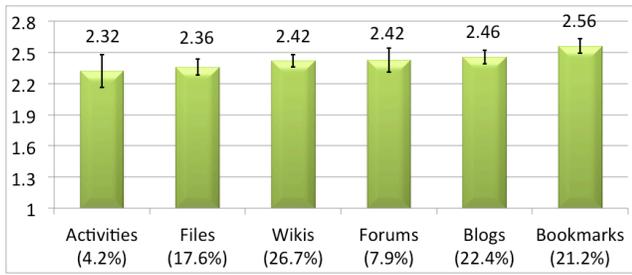


Figure 4. Average ratings by item type.

### 5.3 Item Types

Figure 4 depicts the average rating results according to the type of the recommended item. In parenthesis is the occurrence of the type, i.e., its percentage out of all recommended items. Wikis, blogs, and bookmarks account for over 70% of all recommendations, while forums and activities are the least frequent. In terms of average ratings, there is not a big difference among the types, with bookmarks being the type with highest average rating of 2.56, while activities have the lowest at 2.32. Bookmarks are the only type that can also point to external content and may thus have the potential to yield more interest [34]. A one-way ANOVA with Tukey post-hoc analysis indicates that, except for the difference of bookmarks with files and wikis, all other rating differences among item types were insignificant,  $F(5,7595)=3.87, p<.005$ . For owner ratings, we observed very similar results, with bookmarks being the most highly rated type and files (rather than activities) being the lowest ranked type.

All in all, the results show that all types of recommended social media content produce rather similar results and there is no clearly superior type. This implies that mixing recommended item types makes sense for this recommendation task, similar to individual recommendations [18]. The two lowest types, files and activities, are not among the commonly recommended ones. Bookmarks have a slightly higher interest rate and it may be desirable to boost them a bit further.

### 5.4 Number of Members and Owners

As mentioned before, the number of members in our inspected communities was highly diverse. We therefore set out to explore the ratings for *small communities* of size equal to or smaller than 100 (57% of the communities) compared to *large communities* with over 100 members (43%). The results were very different for these two types of communities. The total average rating was significantly higher for large communities compared to small communities – 2.56 vs. 2.36 (one-tailed unpaired t-test,  $p<.0001$ ). While one might have thought it would be easier to interest owners of smaller communities in recommendations, since they

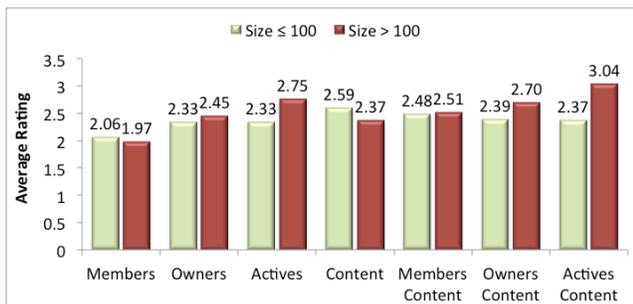


Figure 5. Average ratings for small vs. large communities.

are more cohesive, apparently owners of larger communities perceived the recommendations as more interesting. One possible explanation is that smaller communities are more focused and less open to recommendations. Inspecting the rating distribution of small and large communities across each of the seven profiles, as depicted on Figure 5, sheds more light on this finding.

For small communities, the *Content* profile yields the highest interest ratings of all seven profile types. The *Actives* profile is not very effective for small communities and yields exactly the same average rating as *Owners*. Moreover, the most effective hybrid profile appears to be *MembersContent*. For large communities, the *Actives* profile clearly achieves the highest results among all pure profiles, while its hybridization *ActivesContent* reaches an average higher than 3. Both pure and hybrid *Owners* profiles are also substantially more effective than for small communities. On the other hand, the *Members* profile and its hybridization, as well as the *Content* profile, produce lower interest ratings for large communities. A one-way ANOVA with Tukey post-hoc analysis for large communities across all seven profiles indicates that average rating for the *Actives* profile was significantly higher than all three other pure profiles and average rating for the *ActivesContent* profile was significantly higher than for all other six profiles,  $F(6,3386)=21.09, p<.0001$ . On the other hand, for small communities, the average rating for the *Content* profile was found significantly higher than for all other pure profiles,  $F(6,4581)=8.87, p<.0001$ . For both small and large communities, average rating for the *Members* profile was significantly lower than all other profiles.

Overall, these results indicate that for small communities a content-based recommender is more effective. It is likely that the title, description, and tags for smaller communities are focused on narrower topics and thus yield more accurate recommendations. For MBPs, there is no clear benefit in using owners or actives over regular members in small communities. But when it comes to large communities the picture changes: content-based recommendation becomes less effective (metadata is more likely to represent broader themes), profiles based on regular members become noisy, and profiles that are based on a smaller group of either owners, or, to a significantly larger extent, active members, provide the best means for recommendation. Hybridization with content significantly contributes to further improving the results for large communities. We therefore observe that the general usefulness of the *Actives* and *ActivesContent* profiles stems from their high performance for large communities.

Figure 6 compares average ratings based on the number of owners in a community. The size and number of owners in a community had only a slight positive correlation (Pearson's coefficient of 0.18). The total average rating is significantly higher for communities with more than 5 owners than communities with 5

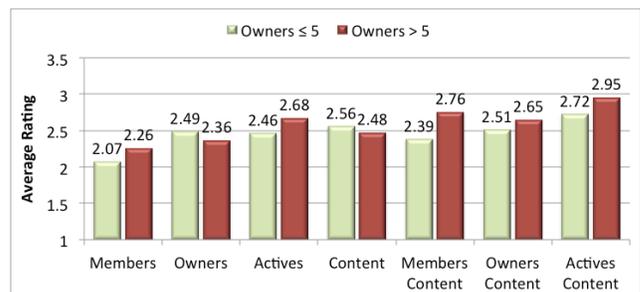


Figure 6. Average ratings based on number of owners.

**Table 2. Average ratings by community type**

|                        | COP                    | Team           | Tech Support           | Recreation      | Other           |
|------------------------|------------------------|----------------|------------------------|-----------------|-----------------|
| % of all               | 44.6%                  | 35.7%          | 5.5%                   | 1.5%            | 12.7%           |
| Median Size            | 111                    | 41             | 51.5                   | 131             | 134.5           |
| Avg rating<br>(95% CI) | <b>2.57</b><br>(±0.05) | 2.4<br>(±0.05) | <b>2.56</b><br>(±0.15) | 1.82<br>(±0.21) | 2.21<br>(±0.09) |

owners or less – 2.59 vs. 2.43 (one-tailed unpaired t-test,  $p < .0001$ ). Yet, the *Owners* profile produces better rating results for communities with fewer owners. When more owners are defined, it is likely that some of them are not really taking an active part in the community or its related activity and thus form a less effective profile for recommendation. This is where the *Actives* profile becomes more effective for recommendation. Indeed, differences between *Actives* and *Owners* for communities with more than five owners were found to be statistically significant, as part of a one-way ANOVA with Tukey post-hoc comparisons,  $F(6,3767)=11.896, p < .0001$ .

### 5.5 Community Type and Engagement

In this sub-section, we examine the effects of answers to the general questions in the first phase of the survey on the rating results. The first question asked owners to map the community by its type. Table 2 shows the average rating across all five types (including “Other” and excluding “Idea Lab”, which was not chosen by any owner). For each type, the table also indicates the type’s percentage out of the total set of participating communities and its median community size. *COP* and *Tech Support* received the highest ratings, followed by *Team*. A one-way ANOVA indicates that ratings across the five types were significantly different,  $F(4,7792)=19.51, p < 0.001$ . Tukey post-hoc comparisons indicate that differences among all types were significant, except between *COP* and *Tech Support*. The fact that between the two most popular types, *COP* and *Team*, *COP* received better ratings, suggests that recommendations are more effective for communities around a topic of interest rather than for more focused project groups. This result is also in line with our findings regarding community size, as teams tend to be smaller than *COPs*. *Technical Support* communities, in spite of being small, receive high ratings. *Recreation* communities receive very low ratings as they focus on topics external to the workplace, making it harder to produce effective recommendations for them based on work-related content. Communities marked as *Other* also receive low ratings.

Table 3 presents the distribution of answers for the three other general questions. Owners generally indicated they were either engaged or very engaged in the communities we asked them about. Less than 17% were only minimally engaged, or not engaged at all in a few rare cases. This distribution of answers is not surprising, since we chose to invite owners who were active in their communities and also since less engaged owners were less likely to respond. Thus, most of our survey was based on engaged owners. In contrast, when asked about the engagement of their community, over 50% indicated it was only minimally engaged, and in rare cases not engaged at all. Only 10.3% thought their community was very engaged. These answers re-iterate the challenge owners see in establishing community engagement. When asked about the activity level they would like to establish, most owners chose weekly, but some also chose daily, monthly, or sporadic.

**Table 3. Distribution of and average ratings by answers to engagement questions**

|                        |        | Not at all | Minimally | Engaged | Very  |
|------------------------|--------|------------|-----------|---------|-------|
| Owner Engagement       | Dist   | 1.6%       | 15.2%     | 44.8%   | 38.4% |
|                        | Rating | 2.33       | 2.25      | 2.43    | 2.55  |
|                        | 95% CI | ±0.25      | ±0.08     | ±0.04   | ±0.05 |
| Community Engagement   | Dist   | 3.8%       | 48.3%     | 37.6%   | 10.3% |
|                        | Rating | 2.06       | 2.34      | 2.51    | 2.88  |
|                        | 95% CI | ±0.15      | ±0.04     | ±0.05   | ±0.1  |
| Desired Activity Level |        | Sporadic   | Monthly   | Weekly  | Daily |
|                        | Dist   | 8.4%       | 14.2%     | 56.5%   | 20.9% |
|                        | Rating | 1.94       | 2.23      | 2.48    | 2.71  |
|                        | 95% CI | ±0.1       | ±0.08     | ±0.04   | ±0.07 |

Average ratings for the recommendations clearly increased as the engagement level of the owner increased and even more sharply as the engagement level of the community increased. For owner engagement, a one-way ANOVA with post-hoc Tukey comparisons indicates that all differences among engagement levels, except those involving the “Not at All” group, were significant,  $F(3,7926)=13.55, p < .0001$ . For community engagement, all differences are found to be significant,  $F(3,7926)=40.48, p < .0001$ . The findings about owner engagement indicate that more engaged owners are more easily interested in recommendations for their community and expect them to interest the community. The findings about community engagement are particularly important: owners who feel their community is minimally or not engaged are also more “skeptical” about the community’s interest in recommendations, while owners who feel their community is very engaged expect more recommendations to interest the community. Therefore, recommendations may be more effective in maintaining high engagement within communities that are already engaged, rather than establishing high engagement in minimally-engaged communities. This can be thought of as a variation of the cold-start problem [39], where jump-starting engagement in a community is particularly hard from a recommendation perspective. It could be that other incentives should be used alongside recommendation to establish engagement.

Finally, it can be seen that average ratings increased as the owner’s desired activity level in the community increased. A one-way ANOVA with post-hoc Tukey comparisons indicates that average ratings across all four categories were significantly different,  $F(3,7936)=56.52, p < .0001$ .

## 6. DISCUSSION AND FUTURE WORK

Our results indicate that hybrid profiles that combine community’s member-based and content-based data are generally more effective for the task of recommending social media content to community owners. This result is in line with most literature on recommender systems, which have shown time and again that hybrid approaches improve accuracy [9]. Active members, i.e., members who made a contribution to the community along its history, emerge as the most effective group for producing interesting recommendations. They outperformed the group of formal owners, in spite of the fact that recommendations were evaluated by owners. Further analysis reveals that the effectiveness of the *Actives* group lies in large communities, where content on its own is too broad and many of the regular

members are just lurking or completely ignore the community [35]. The hybrid *ActivesContent* profile achieved a particularly high average rating when used for large communities. On the other hand, for small communities, we did not observe a particular benefit in considering active members or owners over regular members. Regular members may also be useful for recommendation to newly created communities, where active members are not yet established.

We found that owners and active members only partly overlap. Apparently, active members who are not formal owners play a more central role in representing the community's interests than non-active owners. Such owners do not fulfill the role of an actual community leader. Indeed, when sending the study's invitations we received quite a few responses from owners who pointed us to the "real" owner as they had changed role or were simply not acting as owners despite being given the privileges. This hinted that the formal owners might not be the ideal group for producing recommendations. In this paper, we experimented with one simple definition of active members and showed it was more effective for recommendation than the formal owners. Future work should examine other ways of identifying active members, such as considering only active owners or also taking into account viewing frequency in the community, in addition to contributions.

We experimented with a single content-based method that is based on the community's metadata – title, summary, and tags. The *Content* profile performed best for small communities, probably since they are more focused on specific topics, but was not very effective on its own for large communities. It was still effective, however, when hybridized with member-based profiles. We opted not to consider the full content items already in the community as we believed they would be noisy [2]. The metadata we experimented with was rather rich for most communities: the median number of tags was 4, the median summary length was 38 words, and the median title length was 4. Only 16.3% of the 796 participating communities did not have a summary, only 20.5% did not have tags, and all had a title.

Previous work on social media recommendation to individuals showed that recommending mixed types of items can be productive [18]. In this work, we also experimented with recommending mixed social media content. Our results indicate that ratings across all content types are rather similar, suggesting that there is no one prevailing type upon which recommendations should focus. Bookmarks had slightly higher ratings than other types, while shared files and activities had slightly lower ratings. Future work can factor the item type into the recommendation score based on these results. Moreover, boosting by item type can be adapted to the preferences of a specific community, for instance, by considering the common types of its existing content.

This work provides a baseline for a new type of recommendation task – recommending content to owners for sharing with their community. General rating scores received in our experiments are lower compared to the task of recommending similar social media items to individuals [18], with a total average rating below the neutral score of 3. Still, ratings were significantly higher compared to a random baseline. Furthermore, ratings became higher for specific conditions, such as large communities, engaged owners and communities, and certain profile types. Future work should compare the results of owner ratings to ratings by all community members, as in traditional group recommendation. Further group recommendation techniques could be used as baselines in addition to the random one.

In our survey, we asked owners to rate both their own interest and the expected interest of the community per item. We found that the average rating for interest to the community was lower than the rating for the owner's own interest, even though recommendations were tailored for the community as a whole. This gives some indication of the complexity in this recommendation task. The owner needs to be individually interested in a recommendation and then believe it would interest the community as a whole in order to give it a high rating. The latter is particularly challenging since recommendations need to adapt to the interests and preferences of different members. Moreover, contexts and purposes of communities vary and recommendation needs to account for these too. This became evident in our analysis by community type, which showed diverse ratings for different types. Also, we found that many owners perceive their community as minimally engaged and may thus be selective when they evaluate the potential interest of the community in new content. Future work should consider recommendation to other types of members. For example, it would be interesting to examine whether active members are more receptive to sharing content with their community than owners.

As part of our recommendations, we showed limited evidence that included the tags and people related to the recommended items. More focused evidence, for example, highlighting the specific relevance of terms and people to members and metadata in the community, can help convey the community's potential interest in an item to the owner. Previous work has shown the instant and long-term value of explanations for personal recommendations [18,19]. In future work, we plan to explore the value of explanations in this work's recommendation context.

The evaluation in this paper is solely based on owners' ratings in terms of interest to the community. We did not follow up to examine whether owners took real action and actually shared the recommended content with their community. Future research should inspect how owners act on recommendations by allowing them to share recommended items with their community and tracking this behavior. Such research can also examine the longer-term impact of shared recommended content on the community's health and engagement.

The results of this work are influenced by the specific characteristics of the studied organization and its use of social media behind the firewall. We hope to see further studies on the topic in the future, but note that the principal notions discussed, such as members, owners, active members, and community's metadata, are broadly relevant to online communities and can thus be valid for other organizations. Moreover, due to the generality of these concepts, our findings may also be relevant for online communities on the web, such as LinkedIn groups [6]. Future research should examine a similar type of recommendation outside an organization's firewall, as online communities continue to proliferate.

## 7. REFERENCES

- [1] Amer-Yahia, S., Roy, S.B., Chawlat, A., Das, G., and Yu, C. 2009. Group recommendation: semantics and efficiency. *Proc. VLDB Endow.* 2, 1 (August 2009), 754-765.
- [2] Amitay, E., Carmel, D., Har'El, N., Ofek-Koifman, S., Soffer, A., Yogev, S., and Golbandi, N. 2009. Social search and discovery using a unified approach. *Proc. HT '09*, 199-208.
- [3] Anand, D. 2013. Group movie recommendations via content based feature preferences. *International Journal of Scientific & Engineering Research* 4, 2 (Feb. 2013),

- [4] Ardissono, L., Goy, A., Petrone, G., Segnan, M. and Torasso, P. 2001. Tailoring the recommendation of tourist information to heterogeneous user groups. In *Hypermedia: Openness, Structural Awareness, and Adaptivity*, 280-295.
- [5] Baltrunas, L., Makcinkas, T., and Ricci, F. 2010. Group recommendations with rank aggregation and collaborative filtering. *Proc. RecSys '10*, 119-126.
- [6] Barzilay, O., Hazzan, O., and Yehudai, A. 2011. Using social media to study the diversity of example usage among professional developers. *Proc. SIGSOFT '11*, 472-475
- [7] Berkovsky, S. and Freyne, J. 2010. Group-based recipe recommendations: analysis of data aggregation strategies. *Proc. RecSys '10*, 111-118.
- [8] Bourhis, A., Dubé, L., and Jacob, R. 2005. The success of virtual communities of practice: the leadership factor. *Electronic J. of Knowl. Mgt.* 3, 1 (2005), 23-34.
- [9] Burke, R. 2002. Hybrid recommender systems: survey and experiments. *UMUAI*, 12(4), 331-370.
- [10] Cantador, I., Castells, P. 2012. Group recommender systems: new perspectives in the social web. In *Recommender Systems for the Social Web*, 139-157.
- [11] Carmel, D., Uziel, E., Guy, I., Mass, Y., and Roitman, H. 2012. Folksonomy-based term extraction for word cloud generation. *ACM TIST*, 3 (4), 60.
- [12] Chen, Y., Cheng, L., and Chuang, C. 2008. A group recommendation system with consideration of interactions among group members. *Expert Syst. Appl.* 34, 3 (April 2008), 2082-2090.
- [13] DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., and Muller, M. 2008. Motivations for social networking at work. *Proc. CSCW '08*, 711-720.
- [14] Ebrahim, A. N., Ahmed, S., and Taha, Z. 2009. Virtual R&D teams in small and medium enterprises: A literature review. *Scientific Research and Essays* 4 (13), 1575-1590.
- [15] Gartrell, M., Xing, X., Lv, Q., Beach, A., Han, R., Mishra, S., and Seada, K. 2010. Enhancing group recommendation by incorporating social relationship interactions. *Proc. GROUP '10*, 97-106.
- [16] Gorla, J., Lathia, N., Robertson, S., and Wang, J. 2013. Probabilistic group recommendation via information matching. *Proc. WWW '13*, 495-504.
- [17] Guy, I., Ronen, I., and Raviv, A. 2011. Personalized activity streams: sifting through the river of news. *Proc. RecSys '11*, 181-188.
- [18] Guy, I., Zwerdling, N., Ronen, R., Carmel, D., and Uziel, E. 2010. Social media recommendation based on people and tags. *Proc. SIGIR '10*, 194-201.
- [19] Herlocker, J. L., Konstan, J. A., and Riedl, J. 2000. Explaining collaborative filtering recommendations. *Proc. CSCW '00*, 241-250.
- [20] IBM Connections – social software for business. <http://www-03.ibm.com/software/products/us/en/conn/>
- [21] Jameson, A. and Smyth, B. 2007. Recommendation to groups. In *The Adaptive Web*, 596-627.
- [22] Jameson, A. 2004. More than the sum of its members: challenges for group recommender systems. *Proc. AVI '04*, 48-54.
- [23] Jung-Hyun, L.E.E. 2004. User preference mining through hybrid collaborative filtering and content-based filtering in recommendation system. *IEICE T Inf Syst* 87 (12), 2781-2790.
- [24] Kim, A.J. 2000. Community building on the web: secret strategies for successful online communities. *Peachpit press*.
- [25] Kraut, R.E., Resnick, P., Kiesler, S., Ren, Y., Chen, Y., Burke, M., Kittur, N., Riedl, J., and Konstan, J. 2012. Building successful online communities: evidence-based social design. The MIT Press.
- [26] Leimeister, J.M., Sidiras, P., & Krcmar, H., Success factors of virtual communities from the perspective of members and operators: an empirical study. *Proc. HICSS '04*.
- [27] Li, D., Lv, Q., Shang, L., and Gu, N. 2011. YANA: an efficient privacy-preserving recommender system for online social communities. *Proc. CIKM '11*, 2269-2272.
- [28] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 1). Cambridge: Cambridge University Press.
- [29] Masthoff, J. 2011. Group recommender systems: combining individual models. In *Recommender Systems Handbook*, 677-702.
- [30] Masthoff, J. 2004. Group modeling: selecting a sequence of television items to suit a group of viewers. *UMUAI* 14, 1 (Feb. 2004), 37-85.
- [31] Matthews, T., Whittaker, S., Badenes, H., Smith, B.A., Muller, M., Ehrlich, K., Zhou, M.X., and Lau, T. 2013. Community insights: helping community leaders enhance the value of enterprise online communities. *Proc. CHI '13*, 513-522.
- [32] McCandless, M., Hatcher, E., and Gospodneti, O. 2010. Lucene in action, 2nd edition. *Manning Publications Co.*
- [33] McCarthy, J.F. and Anagnost, T.D. 1998. MusicFX: an arbiter of group preferences for computer supported collaborative workouts. *Proc. CSCW '98*, 363-372.
- [34] Millen, D.R., Feinberg, J., and Kerr, B. 2006. Dogear: social bookmarking in the enterprise. *Proc. CHI '06*, 111-120.
- [35] Muller, M., Ehrlich, K., Matthews, T., Perer, A., Ronen, I., and Guy, I. 2011. Diversity among enterprise online communities: collaborating, teaming, and innovating through social media. *Proc. CHI '12*, 2815-2824.
- [36] Nepal, S., Paris, C., Pour, P.A., Freyne, J., and Bista, S.K. 2013. Interaction based content recommendation in online communities. *Proc. UMAP '13*, 14-24.
- [37] O'Connor, M., Cosley, D., Konstan, J.A., and Riedl, J. 2001. PolyLens: a recommender system for groups of users. *Proc. ECSCW '01*, 199-218.
- [38] Ronen, I., Shahar, E., Ur, S., Uziel, E., Yogev, S., Zwerdling, N., Carmel, D., Guy, I., Har'el, N., and Ofek-Koifman, S. 2009. Social networks and discovery in the enterprise (SaND). *Proc. SIGIR '09*, 836.
- [39] Schein, A.I., Popescul, A., Ungar, L.H., and Pennock, D.M. 2002. Methods and metrics for cold-start recommendations. *Proc. SIGIR '02*, 253-260.
- [40] Senot, C., Kostadinov, D., Bouzid, M., Picault, J., and Aghasaryan, A. 2011. Evaluation of group profiling strategies. *Proc. IJCAI '11*, 2728-2733.
- [41] Wenger, E., McDermott, R., and Snyder, W.M. 2002. Cultivating communities of practice: a guide to managing knowledge. *Harvard Business Press*.
- [42] Xu, A., Chen, J., Matthews, T., Muller, M., and Badenes, H. 2013. CommunityCompare: visually comparing communities for online community leaders in the enterprise. *Proc. CHI '13*, 523-532.
- [43] Yu, Z., Zhou, X., Hao, Y., and Gu, J. 2006. TV program recommendation for multiple viewers based on user profile merging. *UMUAI* 16, 1 (March 2006), 63-82.