Islands in the Stream: A Study of Item Recommendation within an Enterprise Social Stream

Ido Guy[*], Roy Levin[**], Tal Daniel[**], and Ella Bolshinsky[***]

[*]Yahoo Labs, Haifa, Israel idoguy@acm.org[1]

[**]IBM Research-Haifa, Israel {royl, taldan}@il.ibm.com

[***]Department of Computer Science, Technion, Haifa, Israel ellabo@cs.technion.ac.il

ABSTRACT

Social streams allow users to receive updates from their network by syndicating social media activity. These streams have become a popular way to share and consume information both on the web and in the enterprise. With so much activity going on, filtering and personalizing the stream for individual users is a key challenge. In this work, we study the recommendation of enterprise social stream items through a user survey with 510 participants, conducted within a globally distributed organization. In the survey, participants rated their level of interest and surprise for different items from the stream and could also indicate whether they were already familiar with the item. Thus, our evaluation goes beyond the common accuracy measure and examines aspects of serendipity and novelty. We also inspect how various features of the recommended item, its author, and reader, influence its ratings. Our results shed light on the key factors that make a stream item valuable to its reader within the enterprise.

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

Keywords: Beyond accuracy; enterprise; novelty; recommender systems; serendipity; social media; social streams; surprise.

1. INTRODUCTION

Social streams are becoming one of the most prevalent ways to consume information on the web. From the Twitter firehose to the Facebook news feed, social streams allow users to keep track of others' activity, typically within a social media website or a group of sites. Social streams have also recently emerged within enterprises [18,21,28], allowing employees to track updates from their colleagues and peers. For example, the enterprise activity stream within IBM was reported to include over 13,000 items per business day [22]. With this number of items on the rise, the need for effective filtering techniques becomes more acute, so individual employees can keep track of the portion of the stream most useful to them.

The task of effectively filtering a social stream is challenging, both within the enterprise and on the web. Twitter, the leading microblogging service, filters a user's stream of messages based on the people that user chooses to follow. This model is often insufficient, since there may be relevant messages from outside the user's list of "followees". On the other hand, some of the messages from the user's followees might not be relevant. Facebook, whose feed is based on the user's network of friends, recently stated that "*Our ranking isn't perfect, but in our tests, when we stop ranking and instead show posts in chronological order, the number of stories people read and the likes and comments they make decrease*" [14]. Yet, little is known about the algorithms Facebook applies to score news feed items.

Enterprise social streams pose a unique filtering challenge of their own. In a global organization, employees use the enterprise stream to keep track of relevant activity from colleagues, groups, projects, or topics they are involved with or are interested in. They also use it to share and promote ideas, get to know new people in the organization, and increase their awareness of themes and projects that take place across the organization [12,22]. The usefulness of an item to employees may also be affected by organizational characteristics, such as their business unit or work location.

In this work, we study recommendation of items within an enterprise social stream. We employ a comprehensive personalization model, which extends a previous work that compared the personalization of an enterprise stream using three types of user profiles, based on the user's related people, terms, and entities (blog posts, wiki pages, etc.), respectively [20]. In that work, filtering was based on a binary check of whether the item was related to a person (or term, or entity) in the user profile. As noted, this approach might be inadequate, since the user may be interested in more (or fewer) items than the profile can produce, which makes a finer-grained ranking of the stream's items necessary. Our extended model builds a profile that combines all three elements – people, terms, and entities – and assigns a personalized recommendation score to an item by issuing the profile objects as a query to a unified search index [3]. In addition, we experimented with two non-personalized popularity-based techniques, which generate items based on popular authors and popular entities, and compared their results with those of the personalized model.

Our evaluation is based on a user survey, in which 510 active users of an enterprise social media platform rated items from the platform's activity stream. In our survey, participants were asked to provide ratings not only of their interest in an item, but also of their surprise from it. We did not explicitly define interest or surprise, but let participants form their own interpretation. Surprise is commonly associated with the measure of *serendipity* in recommender systems (RS). There is an agreement within the RS community that accuracy on its own is insufficient for measuring user satisfaction; serendipity is one of the key

---

supplementary measures [19,24,30]. Nevertheless, most RS studies focus on recommendation accuracy as the sole evaluation measure and overlook serendipity [33], partly because it is hard to evaluate through A/B testing or offline experiments [24]. In some of the studies that do examine serendipity, it is measured as the pure portion of surprising or unexpected items [32,38], while in other studies serendipitous items are considered as items that are both surprising and accurate [1,19,24].

In the survey, we also asked participants if they were already familiar with the presented item. With this question, we wanted to assess *novelty*, which is another measure commonly mentioned as complimentary to interest [1,7,30,38]. In stream recommendation, however, novelty is likely to be less of an issue, due to the short relevancy period of items and the high appearance rate of new items, compared to traditional taste domains, such as movies, hotels, or music. As in previous studies, we identified the following link between novelty and serendipity: an item already known to the user cannot be surprising [1,26]. We thus asked for the surprise rating only when the item was not marked as already known. To the best of our knowledge, this is the first study to directly evaluate the recommendation of social stream items using beyond-accuracy measurements.

As part of our analysis, we examined the effect on ratings of different non-personalized characteristics of the recommended item, including its type and the activity rate of its author and entity. We also analyzed the ratings of recommended items based on the organizational characteristics – work location, business unit, and managerial status – of the item's reader as well as its author. The main contribution of this work is twofold: (1) we use a unique beyond-accuracy evaluation to analyze interest versus surprise ratings for personalized versus popular items; (2) our findings help understand what makes a valuable recommendation of a social stream item in the workplace.

In our evaluation, we found that most of the items that were rated surprising were also rated interesting, indicating that users mostly associate surprise with interest and interpret it in a positive way. While personalized items were found to have significantly higher interest ratings, popular items were found significantly more serendipitous and novel, with significantly higher portion of items rated as both interesting and surprising. Additionally, based on the set of inspected measurements, we report a list of factors that are likely to make an item more valuable to its reader within the organization. For example, items that originate from an active entity (e.g., an active blog), but a less active author, are likely to be more valuable; items that originate from the same country or business unit are likely to be more interesting, but less surprising; and items that originate from a manger in an internal (corporate) unit are likely to be more valuable, especially when read by an employee from a non-technical (sales or corporate) unit.

2. RELATED WORK

Much of the social stream research has been conducted over Twitter, the leading microblogging service, which is a homogenous stream of short messages ("tweets"). For example, several studies suggested extending the Twitter personalization model by taking into account topical context [5,13]) and others focused on personalized tweet ranking and recommendation [8,17]. In this work, we study a heterogeneous social stream that includes different types of items. The most prominent example for a heterogeneous stream is the Facebook news feed, which provides a summary of friends' activity, such as posting a photo, writing a message, sharing a link, or organizing an event. Paek et

al. [34] used machine learning to predict the importance of both friends and posts on the Facebook news feed. Their evaluation was based on a survey of 24 Facebook users. Cui et al. [9] proposed a matrix factorization approach for predicting item-level social influence in a stream. Their experiments were conducted over a Facebook-style Chinese social network site (SNS) based on user-post sharing interactions and showed their method increased the prediction's precision.

Another prominent example of a heterogeneous social stream is the LinkedIn stream, which enables users to receive updates from their professional network. Hong et al. [25] proposed a probabilistic latent factor model that combined information retrieval and collaborative filtering techniques to rank updates in the LinkedIn stream and evaluated it based on clickthrough data. In a recent study, Agrawal et al. [2] described the machine-learning system used for ranking the LinkedIn's stream items. They included online bucket evaluation based on click-through data and found that a personalized model often achieves a large performance enhancement. Berkovsky et al. [4] developed a predictive model for items in a heterogeneous stream of an eHealth portal; the model was based on user-to-user and user-to-action scores. Evaluation used click-through data and showed that their personalization method achieved better accuracy than non-personalized baselines. The focus of all of these studies was on improving the accuracy of the items in the stream.

The literature on enterprise social streams is rather sparse. Freyne et al. [18] suggested a method for narrowing the stream of an enterprise SNS based on person and action relevance inferred from users' browsing behavior; they provided an initial evaluation based on clickthrough data. Daly et al. [10] proposed a user experience with multiple sharable user profiles, called "lenses", to support better filtering of the enterprise stream. Lunze et al. [28] ran a small experiment with 9 users over the Communote enterprise social stream and found that analyzing the item's text is essential for identifying important items. Guy et al. [20] compared the use of people, terms, and entities in a user's profile for personalizing an enterprise activity stream. They showed that building the user's profile based on data from the stream itself is effective for the personalization task. Our own personalization method builds on that model and further generalizes it to combine people, terms, and entities in one profile. Despite being held within an enterprise, none of these studies explored the effect of enterprise-specific user characteristics, such as business unit, work location, or managerial status, on personalization quality.

In our survey, aside from rating the interest level in an item, participants were asked to indicate if they already knew it and how surprised they were by it. McNee et al. [30] mentioned that evaluating recommender systems by accuracy alone is insufficient and suggested other measures, including novelty and serendipity, to complement RS evaluation. Novelty refers to the quality of a recommended item being unknown to the user [1,7,30]. Zhang et al. [37] defined it as the ability to introduce users to items they have not previously experienced in real life. Often times, novelty is enhanced by increasing the diversity among recommendations, with respect to different features [38].

Serendipity is the quality most related to an item being surprising. There have been several attempts to define serendipity. McNee et al. [30] defined it as the experience of getting an unexpected and fortuitous item recommendation. Desrosiers and Karypis [11] defined it as the extension of novelty by helping users find an interesting item they might not have otherwise discovered. Herlocker et al. [24] defined serendipity as the extent to which the

items are both attractive and surprising to users. Zhang et al. [37] stated that serendipity represents the "unusualness" or "surprise" of recommendations and noted that in taste domains, a serendipitous recommendation challenges users to expand their tastes, in addition to potentially increasing user satisfaction.

Serendipity is not only hard to define, but also hard to evaluate, due to its subjective characteristics [19,24,26]. Murakami et al. [32] proposed a measure of unexpectedness based on the distance from a basic prediction method's results (for a whole recommendation list). Ge et al. [19] built on this unexpectedness measure and suggested intersecting it with a usefulness measure to evaluate serendipity. Iaquinta et al. [26] enhanced serendipity by promoting items that have both a strong positive and a strong negative prediction scores. Onuma et al. [33] proposed a graph-based approach, which gives high scores to nodes that are well connected both to the user's preferred items and to unrelated items. Zhang et al. [37] suggested a measure of "un-serendipity" based on the average similarity between items in the user's history and a new recommendation. These studies were all conducted in taste domains, such as television shows and music. In this work, we bring the serendipity notion to social stream recommendation and measure it directly through user feedback.

3. EXPERIMENTAL SETUP
In this section, we describe our research settings, including the platform used for our experiments, the recommendation algorithms, and the user survey we conducted.

3.1 Research Platform
For our research, we used a deployment of IBM Connections (IC) [27] within a large global enterprise, in which social media has been widely used for several years. IC is a social media application suite for the enterprise. It consists of different types of social media applications that allow employees to share and interact behind an organization's firewall: an enterprise SNS that enables employees to tag and connect to each other; a blogging application that facilitates the creation of blogs; a social bookmarking application that allows employees to store, share, and tag intranet and internet pages; a file sharing system; a forum application for creating and replying to forum topics; a microblogging service that allows posting messages of up to 500 characters; a collaborative task management service that allows employees to create, assign, and mark tasks as complete; and a wiki system that allows co-editing of pages.

IC publishes an activity stream that includes all public actions taking place within its applications [22]. Each item in the stream includes a textual description with the activity, author, entity(ies) involved, and a short excerpt of the text. The author and entities are linked to their unique IC identifier. For example, an item can tell that "Alice Oh liked the blog post '10 most useful Eclipse tips' in the Eclipse Development blog." Table 1 describes the different types of items in the stream, including the involved entity and possible activities. Each of these items may be performed as part of a community [35] or as a "standalone" activity by the individual user. Previous research has found that the vast majority of the enterprise stream's items focus on workplace-related activity, such as discussing projects, products, ideas, potential customers, organizational tools and processes, and similar topics [12,22,28]. The stream is also composed of network activities, which include connecting to and tagging another person on the enterprise SNS. We did not include these items in our recommendations since they were previously found of particularly low interest [20].

Table 1. Items included in the IC's activity stream

Entity	Activity
Blog post	create, edit, comment, like
Bookmark	create, edit
File	create, edit, share, comment, like
Forum topic	create, edit, reply, like
Microblog message	create, edit, share, comment, like
Task	create, edit, assign, complete
Wiki page	create, edit, comment, like

IC includes various mechanisms to update users about relevant activity in the stream. Users can follow other individuals and entities to get email notifications whenever a related item occurs. IC also sends weekly and/or daily digests summarizing activity related to friends and followed individuals or entities.

3.2 Recommendation Algorithms
Our experiments included recommendations of both personalized and popular items. In this section, we describe the algorithms used for both types of recommendation.

3.2.1 Personalized Items
Our user profile is based on data originating from the stream itself, which was shown to be advantageous in a previous study [20]. That study separately explored the use of people, terms, and entities (e.g., blog posts, wiki pages; referred to as 'places' in that work) for recommending stream items and found that all three are effective. Entities produced the most accurate recommendations (79%), followed by people (58.1%), and then terms (44.8%). On the other hand, terms were shown to produce items more frequently (528 items per term per month) than people (104.3) and entities (only 12.2). Based on these results, we extended the model and built a profile that jointly contains people, terms, and entities, with entities boosted by a factor of 5 over people and people boosted by a factor of 5 over terms. We experimented with three profile sizes: for each user u, a profile $Pr_s(u)=\{T_s(u),P_s(u),E_s(u)\}$, $s \in \{5,10,15\}$ was created. The user profile included a set of s related terms $T_s(u)$, s related people $P_s(u)$, and s related entities $E_s(u)$, all with their relationship score to the user u[1]. The value of s for each user was selected in a round-robin order.

For the profile, we considered a stream that included all items in the year that preceded our experiment. We define the *user's stream* as the set of all items authored by that user. We next describe how we generated the profile for a given user.

Related entities were extracted by considering all the entities in which the user was active, i.e., all entities that appear on the user's stream. These may include blog posts the user authored, commented on, or liked; wiki pages s/he created and edited; files s/he edited and shared; and so forth. These entities were scored by the number of items in the user's stream that related to them.

---

[1] In practice, we did not expect the "ideal" profile to include an identical number of terms, people, and entities. But as we had no prior knowledge about the desired ratio, we opted to experiment with three configurations with equal number of each, to fairly inspect the effect of the overall profile size. Our goal was to examine the profile size as one of many factors we experimented with, rather than find the optimal profile configuration, which is left beyond the scope of this paper.

Thus, if a user performed more activity over an entity, its relationship score to that user would grow.

Related people were extracted by considering both direct and indirect relations. Direct relations considered other people who appear on items from the user's stream (e.g., connecting on the enterprise SNS or tagging one another). Indirect relations were inferred by considering users who have common entities with the user, i.e., entities that appear both on the user's stream and their stream. For example, people who commented on the same blog post as the user or people who liked a file the user created. The overall relationship score with a person was determined by considering all direct and indirect relations between the user and that person (see full details in [20]).

Related terms were extracted by detecting the most representative keywords in the user's stream. To this end, we used the KL+TB method [6], which has been shown effective for term extraction in social streams [21]. The method uses the Kullback-Leibler divergence (KL), which is a non-symmetric distance measure between two given distributions. In our case, we sought out terms that maximize the KL divergence between the language model of the user's stream and the language model of the entire stream. Intuitively, a term would receive a higher KL score if it appeared more often on the user's stream and less frequently on the rest of the stream. On top of the KL measure, a tag boost (TB) was applied, promoting keywords that are likely to appear as tags when appearing in the content. This "likelihood" is determined based on a well-tagged folksonomy, in our case based on the IC's bookmarking application. The weighted list of a user's related terms was generated by applying KL+TB on that user's stream, after filtering out people's names and reserved keywords (such as 'blog', or 'like') and stemming.

Given a user profile $Pr_s(u)=\{T_s(u),P_s(u),E_s(u)\}$, we generated the personalized items by issuing an OR query containing all the profile objects (people, terms, entities) to a social search system [21]. The social search system, which is built on top of Lucene [29], indexes all the items in the stream, as detailed in Table 1. It takes advantage of Lucene's real-time indexing capabilities to keep the index fresh with the most recent items, up to a one-minute delay between an item's publishing time and its inclusion in the index [21]. The system uses a unified approach, which maps the relationships among the stream's items, related people, related entities, and related terms, in a way that makes all four (items, people, terms, entities) both searchable and retrievable [3]. For the task of producing recommendations, the query to the social search system included a combination of people, terms, and entities, while the results were stream items that matched the query, ordered by their recommendation score. The recommendation score of an item i to user u was calculated according to the following formula:

$$RSc(u,i) = e^{-\alpha\tau(i)} \cdot [\beta \sum_{e \in E(u)} w(u,e) \cdot w(e,i) + \gamma \sum_{p \in P(u)} w(u,p) \cdot w(p,i)$$
$$+ (1-\beta-\gamma) \sum_{t \in T(u)} w(u,t) \cdot w(t,i)]$$

where $\tau(i)$ is the number of days passed since the occurrence of i; α is a decay factor (set in our experiments to *0.05*); β and γ are parameters that control the relative weight among entities, people, and terms. According to our boosting mentioned previously, we set $\beta = {}^{25}/_{31}$ and $\gamma = {}^{5}/_{31}$; $w(u,e)$, $w(u,p)$, and $w(u,t)$ are the scores of an entity e, person p, or term t, given as part of $Pr_s(u)$ and reflecting their relationship strength to the user u, as explained before; and $w(e,i)$, $w(p,i)$, and $w(t,i)$ denote the relevance score of

item i to e, p, or t, respectively, as determined by the social search system.

After retrieving the top 100 items with their recommendation scores, the list was traversed from top to bottom and two types of items were filtered out: (1) items that the user authored, and (2) items that belong to a thread of which another item has already been recommended. To this end, we define a *thread* of items as a set of items in the stream that includes all activities that relate to the same entity (see Table 1 for a list of entities). For example, a thread can include all items (creation, comments, likes) that relate to a certain microblog message. This way, we only recommended the top-scored item of a thread and avoided recommending multiple items from the same thread.

3.2.2 Popular Items
We experimented with two different methods for generating popular items. The first, denoted *pop-auth*, was based on popular authors and the second, *pop-ent*, was based on popular entities. We selected popular authors based on the number of people who tagged them within the IC enterprise SNS [16]. Popular entities were identified based on the number of distinct users who performed any type of activity over them during the month that preceded the survey. We identified the 50 most popular authors and 50 most popular entities. For each, we retrieved the most recent related item that occurred during the month preceding the survey (if such existed, in the case of popular authors). This produced a list of (at most) 50 items, of which we selected the popularity-based recommendations at random for each user.

3.3 User Survey
Our evaluation was based on a user survey, in which employees were asked to rate (up to) 15 items originating from the IC activity stream. Of these 15 items, 11 were generated based on the personalization algorithm and 4 were based on popularity: 2 *pop-auth* and 2 *pop-ent*. After generating all 15 recommendations, we randomized their presentation order. In the (rare) cases of an identical item among the three groups (personalized, *pop-auth*, *pop-ent*), such item was presented once and analyzed as part of each of the groups it belonged to.

In the survey, participants were asked to rate each item with regard to their interest in it, their surprise from it, and whether they were already familiar with it. As in previous studies, we asserted that an item marked as already known cannot be surprising and therefore only asked for the surprise rating if the item was not marked as known [1,26]. The different questions allowed us to evaluate the recommendations by aspects that go beyond the common accuracy metric [19].

Figure 1 illustrates an item in our survey. The upper part shows the item itself, including a photo of the author and an icon representing the originating IC application. The item's text includes a description and an excerpt from the content, when relevant. Each underlined element in the item's description is a link to its corresponding IC page. Below the text is an indication of the item's freshness, e.g., "3 days ago".

The lower part asks for the user's feedback: the interest level (not interesting, interesting, very interesting), whether the item is already known (yes/no by a checkbox, which is unselected by default) and the surprise level (not surprising, surprising, very surprising). If the user selected the "already know" box, the surprise rating would gray out.

The survey participants were active users of IC, for whom we could extract at least 15 related terms, 15 related people, and 15
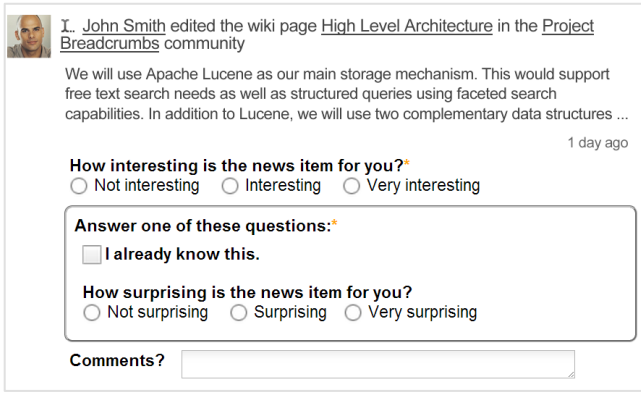
Figure 1. Item recommendation as presented in the survey.

related entities. We identified 1715 such users and sent them an invitation to participate in the survey by email. We note that this sample does not represent the entire organization's employee population, but rather active enterprise social media users, who are the target population for our recommendations. We received a response from 510 users who fully completed the survey (29.7%), rating a total of over 7600 items. More demographic information about our participants is provided in Section 4.4.

4. EXPERIMENTAL RESULTS

Overall in our survey, 59.1% of the items were rated as either interesting or very interesting (15.4% were rated very interesting). Figure 2 shows the interest ratings of personalized versus popular items. Personalized items were significantly more interesting, with 65.1% of the items vs. 42.5% for popular items (p<.001)[2].

25.8% of the items were marked "I already know this". This relatively high portion is likely due to the existing mechanisms for updating IC users, including email notifications and periodic digests according to user preferences, as explained in the previous section. For personalized items, 32.6% were marked as already known, compared to only 7.1% for popular items (p<.001). This fits the intuition that popular items are more exploratory than items that were tailored for the users, and are thus more likely to be novel.

Comparing the ratings for the two types of popular items, *pop-auth* and *pop-ent*, reveals that they were very similar to each other. For interest ratings, 43.8% of the *pop-auth* items were rated [very] interesting (i.e., interesting or very interesting) vs. 41.3% of the *pop-ent* items (p>.05). For surprise ratings, 31.3% of the *pop-auth* items were rated [very] surprising vs. 30.8% of the *pop-ent* items (p>.05). Already-know (AK) portions were also similar:
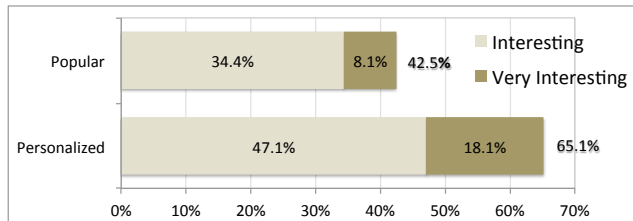

Figure 2. Interest ratings for personalized vs. popular items.

---

[2] Our tests for statistical significance were performed using a two-tailed unpaired t-test when comparing two groups and a one-way ANOVA with Games-Howell post-hoc analysis when comparing three or more groups.
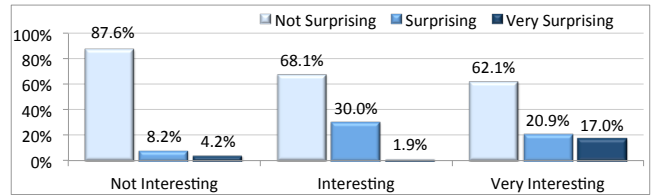

Figure 3. Surprise ratings by interest ratings.

7.1% for *pop-auth* vs. 7.2% for *pop-ent* (p>.05). We therefore jointly refer to both types of popular items from this point onward.

4.1 Surprise Ratings

Overall, 24.8% of the items were rated surprising, of which 5.2% were rated very surprising. Figure 3 presents the distribution of surprise ratings according to the item's interest ratings. It can be seen that less than 13% of the non-interesting items were marked [very] surprising. In contrast, over 30% of the interesting items and almost 40% of the very interesting items were found [very] surprising. The differences among the three interest groups were significant, $F(2,7650)=243.42$, $p<.001$. The portion of very surprising items is especially high for very interesting items at 17%. Overall, it seems that most participants interpret "surprising" as a positive surprise that is joined with interest in the item itself. In total, **79.6%** of the items rated surprising were also rated interesting.

Table 2 (upper part) shows the surprise ratings for personalized versus popular items. Popular items had a significantly higher portion of items rated surprising (p<.001). The lower part of Table 2 shows the surprise ratings for personalized versus popular, focusing only on items that were rated [very] interesting. It can be seen that given that an item is interesting, it has a significantly higher chance of being surprising if it is a popular item (over 50%) rather than a personalized item (less than 30%, p<.001).

As we have seen, most items that were rated surprising were also rated interesting. Ultimately, it is desirable to recommend an item that is both interesting and surprising to the user, in order to achieve a "good surprise" [19,24]. As we have also witnessed, personalized items had a higher portion of interesting items, while popular items had a higher portion of surprising items. But which of them had a higher portion of items rated as **both** interesting and surprising? We found that popular items had a significantly higher portion than personalized items at **22%** vs. **18.9%** (p<.01).

Table 3 summarizes the results mentioned throughout this section comparing personalized and popular items[3]. It can be seen that while personalized items have a clear advantage in terms of

Table 2. Surprise ratings for personalized vs. popular items

	Not Surp.	Surp.	Very Surp.
Personalized	77.5%	18.3%	4.2%
Popular	69%	23.3%	7.7%
Personalized & [Very] Interesting	70.9%	24.2%	4.9%
Popular & [Very] Interesting	48.2%	41.9%	9.9%

Table 3. Summary comparison of personalized vs. popular items

	Int	V. Int	AK	Surp	V. Surp	Int & Surp
Personalized	65.1%^	18.1%^	32.6%^	22.5%	4.2%	18.9%
Popular	42.5%	8.1%	7.1%	31%^	7.7%*	22%*

---

[3] t-test significant differences: +p<.05, *p<.01, ^p<.001

accuracy, in all other beyond-accuracy aspects – AK, surprising, very surprising, and interesting+surprising rates – popular items have the advantage (all differences are statistically significant).

4.2 Personalization Characteristics

In the following analysis, we examine the effect of two factors on the ratings of personalized items: the size of the profile, as determined by s, and the recommendation score, calculated as explained in Section 3. We refer to *interest (surprise) rates* as the portions of items rated [very] interesting (surprising) and *AK rates* as the portion of items marked as already known.

4.2.1 Profile Size

Figure 4 shows the rating results for the three types of profile size s we used in our experiments[4]. As explained in Section 3, a profile size s indicates that the profile includes s related people, s related terms, and s related entities. While we speculated that a smaller profile size would yield higher accuracy, results indicate that the larger the profile, the higher the ratings. Interest rates for $s=5$ were significantly lower than for $s=10$ and $s=15$, $F_{(2,5610)}=9.853$, $p<.001$. It could be that a smaller profile produces a smaller amount of accurate items. It is also likely that a larger profile produces more diverse items, while items for the smaller profile may sometimes be perceived as similar to those that previously appeared and therefore yield less interest. For example, with a smaller profile it is more likely that many items originate from the same author. The mid-size profile ($s=10$) yielded significantly higher surprise rates than the two others ($F_{(2,5610)}=15.034$, $p<.001$) and also lower AK rates ($F_{(2,5610)}=5.433$, $p<.01$). Overall, we see that a profile of $s=15$ yielded the highest interest rates, while a profile of $s=10$ yielded the highest surprise rates. From this point onward, the analysis for personalized items is performed across all three profile sizes.

4.2.2 Recommendation Score

Figure 5 shows the AK, interest, and surprise rates of personalized items as a factor of the recommendation score, RSc, as described in Section 3 (based on 8 equally-sized bins). As expected, a higher RSc leads to higher interest rates, up to 76.7% at the top bin. The more substantial rise starts right after the median point (fourth bin). Also, starting at that point, the AK rates noticeably increase with the RSc, up to 49.2% for the top bin. In parallel, the surprise rates start decreasing. Overall, a high recommendation score produces items with a higher likelihood of being interesting, but also already known and less surprising.

4.3 Recommended Item Characteristics

In this sub-section, we examine the effect of various features of the recommended item on its ratings. The analysis usually presents the results separately for personalized and popular items, since their distributions across the feature values were different.
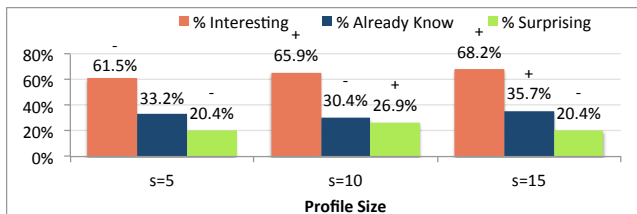

Figure 4. Ratings by profile size (s).

[4] ANOVA Results: values marked by '+' are significantly higher than values marked by '-'; in addition, values marked by '>' are significantly higher than values marked by '<'.
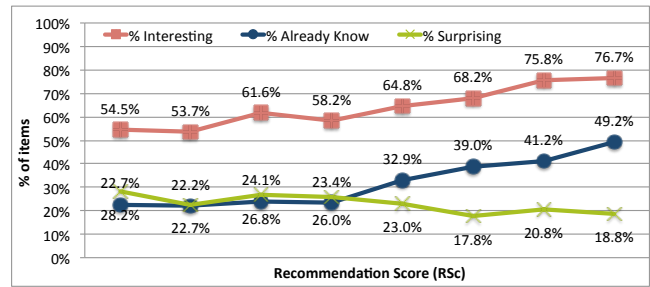
Figure 5. Personalized item ratings by recommendation score.

4.3.1 Application Source

As Table 1 indicates, items in the stream can originate from seven different applications. Figure 6 (upper part) shows the interest ($F_{(6,5610)}=35.635$, $p<.001$) and surprise ($F_{(6,5610)}=3.561$, $p<.01$) rates for personalized items for each of these applications. In brackets is the occurrence of the source, i.e., the portion of items that belonged to it out of all personalized items in our survey. Wikis and blogs were the most common, accounting for 36.1% and 28% of the items, respectively. However, while blogs yielded items with the highest interest rates among all sources (76.9%), wikis had the lowest interest rates (54%). Microblogs and forums also had high interest rates, while tasks had low interest rates. For surprise rates, bookmarks and microblogs, followed by blogs, had the highest surprise rates, while file items were the least surprising. Overall, blogs and microblogs produced the best combination of high interest and high surprise rates, while wikis and tasks had both low interest and low surprise rates. Blogs and microblogs present a more personal type of updates, while wikis and tasks are "dryer" and typically consist of many incremental edits.

The lower part of Figure 6 shows the same statistics for popular items. The occurrence of the sources is quite different than for personalized items. In general, it can be seen that popular items tend to be of the sources that also receive higher ratings. Therefore, the portion of blogs (almost 50% of all items) and microblogs substantially increases and that of wiki substantially decreases, as compared to personalized items. As for interest ($F_{(6,2040)}=10.394$, $p<.001$) and surprise ($F_{(6,2040)}=2.385$, $p<.05$) rates, the results are rather similar to the case of personalized items, with blogs, microblogs, and bookmarks having the highest interest and surprise rates, while tasks and wikis yielding low interest and surprise.

As mentioned in Section 3, an item can be performed in the context of a community [35]. For 5 of the 7 applications, their content type can be associated with a community. These include
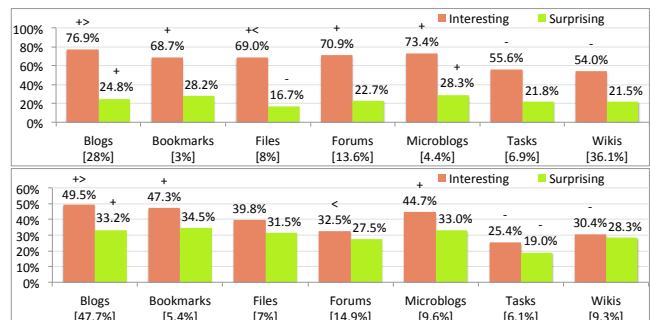

Figure 6. Ratings by application source for (top:) personalized and (bottom:) popular items.

Table 4. Ratings based on association with a community

Community Related?	Personalized				Popular			
	%	**Int**	**AK**	**Surp**	**%**	**Int**	**AK**	**Surp**
No	23	61.7	33.6	20.6	26.2	40.7	5.8	33
Yes	77	66.8*	32.8	23.1	73.8	44.3+	7.4	31.1

Table 6. Ratings of personalized items by entity and author's activity frequency

	Entity Activity				Author Activity			
	%	**Int**	**AK**	**Surp**	**%**	**Int**	**AK**	**Surp**
Top	50	69.2^	36.7^	22.4	50	64.1	32.2	21.7
Bottom	50	60.5	28.1	22.8	50	66.2	33.1	23.4

blogs, bookmarks, files, forums, and wikis. For these sources, we compared the ratings received for items associated with a community to items that were not related to a community. Overall, 77% of the personalized items and 73.8% of the popular items of these 5 types were associated with a community. Table 4 presents this comparison's results. It can be seen, for both personalized and popular items, that those associated with a community were significantly more interesting, while similarly surprising. Belonging to a community may scope the item in a clearer way and therefore make it more interesting. An item performed as part of a community may also have an initial audience who is more likely to notice it and is more committed to give feedback, helping it become more interesting.

4.3.2 Activity Type
Table 5 presents the ratings per each of the four most common activity types (across all applications): create, edit, comment, and like. Together these accounted for 94.3% of the personalized items and 99.4% of the popular items in our survey. As can be seen in the occurrence columns (marked by '%'), edits were most commonly recommended for personalized items, but much less commonly as popular items, for which liking activities were the most common.

Inspecting the interest rates, it is evident that edit activities, which incrementally change a document's content and include many wiki-page edits, are the least interesting for both personalized and popular items. For personalized items, liking and commenting were the most interesting with over 70%. Curiously, creation activities came only third, behind the two feedback activities. AK rates for personalized items were highest for commenting and liking, and lowest for editing. For popular items, AK rates were low in general, but lowest for comments and edits. Finally, inspecting the surprise rates reveals they were highest for create activities. Likes also received high surprise rates, while comments and edits were less surprising. These findings are similar for both personalized and popular items. Overall, among the four activity types, liking yielded the highest interest, while creation yielded the most surprise. On the other hand, editing triggered both low interest and low surprise.

4.3.3 Entity and Author's Activity
In this sub-section, we examine the activity frequency of the item's entity and author and their effect on the item's ratings. We focus on personalized items, since popular items have an inherent bias towards active entities and authors. For the analysis, we inspected the number of items produced by the author or entity in the six months preceding our survey. Table 6 compares the ratings for the top half, which includes the more active entities, with the bottom half, which includes the less active entities (median was 7

Table 5. Ratings by activity type

Activity Type	Personalized				Popular			
	%	**Int**	**AK**	**Surp**	**%**	**Int**	**AK**	**Surp**
Create	17.8	68.9+<	31.7	26.6+	21.2	44.7+	9.2+	32.7
Edit	44.9	56.3-	27.6-	22.2-	15.6	33.2-	5.3	25.1
Comment	10.9	72.5+	44.3+>	18.7-	19.7	40.7	4.7-	30.5
Like	20.7	76.9+>	37.6+<	22.9	42.9	45.8+	7.8	32.4

activities) and analogously for authors (median was 20). It can be seen that items originating from entities that were more active were significantly more interesting, with significantly higher AK rates. Despite the higher AK rates, the surprise rates were similar to items from less active entities. Intuitively, active entities are likely to represent popular or trendy posts, pages, or topics, whose items yield more interest.

For authors, interestingly, the situation is different. Less active authors produce slightly higher interest (p=.07). The AK rates are similar for more active and less active authors, while the surprise rates are also slightly higher for the less active authors (p=.05). Inspecting the two extreme deciles based on author activity reveals a stronger trend: 65.1% interest for the bottom decile (authored 2 items or less) vs. 60.8 for the top decile (272 items or more, p<.05), and 24.6% vs. 19.1% for surprise rates (p<.05). One could have thought that the more active authors are more experienced in producing interesting items and commonly arouse a lot of interest, which also encourages them to keep active. Our results, however, indicate that originating from a less active user increases an item's likelihood of being interesting and surprising. It could be that very active authors often produce repetitive or noisy items, while infrequent authors may arouse more curiosity since it is less common to see them on the stream.

4.4 Reader-Author Relationship
In this section, we examine the effect of different organizational characteristics of an item's reader and author on its ratings.

4.4.1 Work Location
We examine work location in two granularities: country and office address. Our survey participants (readers) originated from 37 countries, while the authors of items presented in the survey originated from 54 countries. The upper part of Table 7 compares the ratings of items whose authors were from the same country and items from different countries. In general, almost half of the personalized items (49.2%) originated from the same country, compared to only 17.8% of the popular items. This indicates that our personalization method is biased towards the same country, even though it does not directly consider it. Inspecting the rating results, we see that interest rates of items from the same country were significantly higher for both personalized and popular items. For personalized items, AK rates were significantly higher for items from the same country, leading to significantly lower surprise rates . For popular items, these differences did not exist.

Survey participants came from 186 different office addresses and items' authors came from 414 different addresses. 15.7% of the personalized items and only 1.6% of the popular items had the same office address for both the reader and the author. Due to the very low number for popular items, we only conducted the comparison of similar versus different office address for personalized items. The lower part of Table 7 presents these results. Interest and AK rates were significantly higher for items originating from the same office address. In spite of the large difference in AK rates, surprise rates were insignificantly higher for items originating from a different office address. One explanation for this can be the reader's expectation of knowing

Table 7. Ratings by reader and author work location

		%	Int	AK	Surp
Country - Personalized	Same	49.2	69.2^	37.8^	20.2
	Different	50.8	59.3	26.9	24.5*
Country - Popular	Same	17.8	45.7+	7	31.4
	Different	82.2	40.9	6.8	31.4
Office Address - Personalized	Same	15.7	72.7^	43.6^	21
	Different	84.3	63.7	30.6	22.8

"everything going on in their office". Overall, items originating from authors in the same office location were more interesting, but also more likely to be already known.

4.4.2 Manager vs. Employee

In our survey, 16.3% of the participants and 18.6% of the items' authors were managers (the general portion of managers within the organization is about 13% [23]). Table 8 shows the ratings for each of the four employee-manager reader-author combinations. For personalized items, interest rates were lowest when both reader and author were employees, and significantly higher when both were managers. For popular items, the highest interest rates were for an employee reading a manager's item. In general, for personalized items, when the author was a manager, interest rates were higher compared to an employee author (regardless of the reader) at 72.7% vs. 63.5% (p<.001). For popular items, this difference was similar at 48.5% vs. 40.8% (p<.01). It is plausible that managers' greater involvement in business decisions and their often-broader business perspective make their items more likely to be interesting.

Inspecting the AK rates, for personalized items, they were significantly higher when the reader was a manager as compared to an employee reader at 39.3% vs. 31.3% (p<.001). They were highest when both author and reader were managers, at 43.4%. For popular items, there was also a slight difference in AK rates for a manager reader compared to an employee reader, at 8.1% vs. 6.9% (p>.05); in addition, there was a significant difference when the author was a manager as compared to an employee author at 10.1% vs. 6.2% (p<.01). Overall, managers marked more items as already known, probably as they are more connected in the organization. For popular items, an item authored by a manager was more likely to be known by others within the organization.

It can be clearly seen that managers were less surprised than employees. For personalized items, the surprise rates for a manager reader were 17.3%, compared to 23.6% for an employee reader (p<.001), and for popular items, they were 26.2% compared to 31.9%, respectively (p<.05). It could be that managers have more years of tenure within the organization and are also better connected, so they are less likely to be surprised. Overall, the analysis in this section reveals that managers are likely to produce more interesting items, while they also tend to be familiar with and less surprised by the items they read.

Table 8. Ratings by managerial role of author and reader

Manager	Personalized				Popular			
	%	Int	AK	Surp	%	Int	AK	Surp
None	70.6	62.9^-<	31.5^-	23.1+	66.3	41.2-	6	31.3
Reader	11.7	67.3<	37.7+	17.4-	12.3	38.7-	7.5	26.1-
Author	13.1	69.5+<	30.5-	26+>	17.5	50+	10.1	34.4+
Both	4.6	81.8+>	43.4+	17.1<	3.9	41.8	10.1	26.6

4.4.3 Business Unit

The studied organization consists of four main business units (divisions): Sales, Services, R&D (including Software, Systems, and Research), and Corporate (CIO's office, HR, Finance, Legal, etc.). Our analysis mostly focuses on the personalized items, due to data sparsity for popular items. Overall, 19.8% of our participants were from Sales, 29.8% from Services, 25.6% from R&D, and 24.8% from Corporate. The distribution of authors for personalized items in our experiment was very similar, but for popular items, R&D authors accounted for only 10.3% of all items, while Corporate and Sales authors had a higher proportion.

Overall, 63.9% of all personalized items originated from the same division as the reader's. As we observed for location, our personalization method favors similar people without directly taking the similarity attributes into account. In contrast, only 24.9% of the popular items were from the same division, indicating no bias. For personalized items, those from the same division were significantly more interesting than those from a different division (69.8% vs. 54%, p<.001) and had significantly higher already-know rates (37.8% vs. 22.7%, p<.001). For popular items, these differences were insignificant at 44.4% vs. 40.6% for interest rates (p=.167) and 8.9% vs. 6.2% for AK rates (p=.075). For personalized items, surprise rates within the same division were significantly lower than across different divisions at 21.1% vs. 25% (p<.01). For popular items, surprise rates were almost identical at 31.2% versus 31.5%, respectively (p>.05).

Table 9 shows the ratings for each reader-author pair across the different divisions, for personalized items. ANOVA-based significance marks are shown only for the Total-Reader and Total-Author comparisons. It can be seen that for interest rates, Sales were interested in Corporate and R&D, in addition to their own items, but much less interested in Services; Services were also interested in Corporate; R&D were also interested in Corporate and Sales; and Corporate were most interested in Sales, in addition to their own items. Overall, Corporate authors yielded the most interesting items, while Services attracted less interest (see Total-Author column). It is reasonable that Corporate employees write more about internal programs and processes that are of broad interest to the entire employee population. Corporate and Sales were generally more interested in items than Services and R&D (Total-Reader row). This can be explained by the fact that both of these divisions require deeper knowledge of the

Table 9. Ratings by reader and author's business unit

Reader \ Author	Sales	Services	R&D	Corporate	Total-Author
% Interest					
Sales	74.1	46.3	55.6	62.1	64.9+<
Services	42.7	64.8	39.1	50.5	57.8-
R&D	65.4	53.7	67.8	54.1	64.2+<
Corporate	69.6	63.3	61.1	75.5	72.2+>
Total-reader	66.2+	61.5-	61.7-	68+	
% Already Know					
Sales	40.2	16.3	17.4	35.6	31.5
Services	21.9	38.3	18.6	23.3	32.5
R&D	29.6	20.7	34.3	18	30.2-
Corporate	28.4	22	28.3	39.1	35.3+
Total-reader	33.7+	32.8	29.3-	34+	
% Surprise					
Sales	20.7	23.8	21.3	20.7	21.3-
Services	21.3	21.4	18.6	24.8	21.6-
R&D	22.8	29.9	19.4	28.7	22.2
Corporate	30.4	31.2	31.9	22.5	25.1+
Total-reader	22.2	23.4+	20.7-	23.4+	

organization to successfully carry out their tasks, compared to the more technical divisions.

For AK rates, items from Corporate employees were slightly more known, while items from R&D people were the least known. More noticeably, R&D readers marked fewer items as already known, perhaps indicating they are less aware of what is going on across the organization. Corporate authors yielded the highest portion of surprise rates, even though their items also had higher already-know rates. For readers, R&D employees were the least surprised by items they read.

5. CONCLUSIONS AND FUTURE WORK

5.1 Result Summary and Discussion

Our results indicate a trade-off between accuracy, reflected in interest ratings, to serendipity and novelty, reflected in surprise and already-know ratings. In terms of accuracy, personalized items achieved 65% interest rates across all recommended items, while popular items only reached 42.5%. The accuracy for personalized items went beyond 75% under certain conditions, such as for items with a very high recommendation score, items that originate from the blog application, or involve a 'liking' activity.

In contrast, personalized items were less effective than popular items in terms of novelty and serendipity. This was reflected in significantly higher already-know rates and significantly lower surprise rates, ultimately leading to a significantly lower portion of personalized items that were both interesting and surprising. These results suggest that popular items pose their own value with respect to novelty and serendipity, at the expense of accuracy. This stands in contrast to previous literature in traditional RS domains, where popular items were suggested as a baseline for non-serendipitous and non-novel recommendations [7,38]. The reason for this difference is the unique character of popular items in the social stream domain, due to their short life span: while movies or books, for example, usually remain popular for years, a popular stream item normally lasts only a few days, after which other trendy items take its place. Stream filtering applications should therefore consider combining popularity-based recommendations within a personalized stream. This could be done, for example, by a popularity boost or by interleaving pure popularity-based items in the recommended stream. Users can also be involved in this process, by indicating their desired level of surprise, as has been previously suggested for other RS [36].

We experimented with two different types of popular items, generated based on popular entities and popular authors. The two methods yielded very similar rating results for interest, already-know, and surprise ratings, granting more validity to our findings about their superiority over personalized items in terms of serendipity. The RS literature proposes various methods to enhance the serendipity of recommendations [26,32,33]. Future research should examine the adaptation of such methods to the social stream domain.

We experimented with three types of profile size. Our results show that the two larger profiles produced better combinations of interest and surprise rates. The largest profile produced the highest interest rates, while the mid-size profile produced the highest surprise rates. We believe that the diversity supported by a larger profile size contributes to its overall recommendation quality. Future work should further examine different size configurations, with different number of people, terms, and entities.

Our study examined various organizational aspects that influence an item's ratings. We found that items from the same work location and business unit are rated more interesting, but less surprising. This indicates that homophily ("love of the same") [31] plays an important role in stream personalization: users tend to be more interested in activity from people who are similar to them, but such activity is also less likely to surprise them. Our analysis also indicated that managers author more interesting items and tend to be more familiar with and less surprised by the items they read. Further analysis revealed that items originating from an internal (Corporate) unit produce higher interest and surprise. Additionally, employees who belong to a technical (R&D) unit are more indifferent to items they read, reflected in lower interest, surprise, and already-know ratings.

Overall, we examined different factors that may influence the value of an item to its reader within the enterprise, in terms of accuracy, novelty, and serendipity. We discovered that an item is more likely to be valuable when it has the following characteristics:

- Originates from the blog, microblog, or bookmark applications, which typically include more appealing content.
- Is carried out in the context of a community, which may provide more focus and scope.
- Is about a create, comment, or 'like' activity, rather than an 'edit' activity, which is typically of smaller value.
- Involves an active entity, but a less active author. Active entities may represent trendy posts, while infrequent authors may sometimes arouse more curiosity when they finally take action.
- Its author is a manger from an internal (Corporate) unit, who is more likely to discuss a topic of relevance to broader parts of the company.
- Its reader belongs to a Corporate or Sales unit (rather than a technical unit), who may have a stronger need to understand the organizational environment.

Understanding the characteristics of valuable items in the enterprise stream can help enhance the adoption and use of enterprise social media in general and enterprise social streams in particular within organizations. This is becoming a key challenge as social media continues to gain popularity and as younger populations, more accustomed to consuming information through social streams and news feeds, are joining the workforce.

5.2 Addressing Limitations

The experiments in this work were based on a user survey and do not include analysis of user behavior in a live system. A/B testing is often used by the industry to analyze usage, such as click-through, on a large scale. In social streams, however, interest is not necessarily reflected by click-through data, as users often read a valuable item and continue scrolling through the stream, without any click. Moreover, novelty and serendipity are particularly difficult to evaluate by analyzing user behavior [24]. The survey enabled us to directly ask for user feedback regarding these qualities. Facebook has recently conducted a user survey to gain an in-depth understanding of what factors give posts on the news feed higher quality [15]. As far as we are aware, the results have not been published. Our survey was relatively short, to allow broad participation with feedback on many items. Future research may use additional questions to expand the understanding of user satisfaction, for instance by directly asking how much an item is "useful" or "valuable".

We used a personalization technique that extends a previous method used for stream personalization in the enterprise and includes all three elements in the user profile: related people (as

currently done in leading social media sites such as Twitter), related terms (as suggested in previous studies [5,13]), and related entities (found productive in [20]). Additionally, we segmented the results based on various factors that may affect the personalization performance, including the item's personalization score, application source, activity type, and profile size. Other personalization techniques can be applied, which may perform differently than our own method. Yet, we believe that the comprehensive user model and the segmentation analysis help make our results broadly relevant for stream recommendation in the enterprise.

The results of our study are influenced by the characteristics of the studied organization and its use of enterprise social media. We hope to see further research on the topic in the future, but note that the basic concepts discussed (related people/terms/entities, popular authors/entities, work location and managerial status, sales, technical, and internal business units) are generally relevant to enterprise social streams and can therefore be valid for other organizations. Moreover, we experimented with applications that represent common social media both within the enterprise and outside the firewall. We therefore believe that some of our findings may also apply for social streams on the web.

6. REFERENCES

[1] Adamopoulos, P. & Tuzhilin, A. 2011. On unexpectedness in recommender systems: or how to expect the unexpected. *RecSys '11 Workshop on Novelty and Diversity in RS*, 11-18.

[2] Agarwal, D. et al. 2014. Activity ranking in LinkedIn feed. *Proc. KDD '14*, 1603-1612.

[3] Amitay, E., Carmel, D., Har'El, N., Ofek-Koifman, S., Soffer, A., Yogev, S., & Golbandi, N. 2009. Social search and discovery using a unified approach. *Proc. HT '09*, 199-208.

[4] Berkovsky, S., Freyne, J., & Smith, G. 2012. Personalized network updates: increasing social interactions and contributions in social networks. *Proc. UMAP '12*, 1-13.

[5] Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E.H. 2010. Eddi: interactive topic-based browsing of social status streams. *Proc. UIST '10*, 303-312.

[6] Carmel, D., Uziel, E., Guy, I., Mass, Y., & Roitman, H. 2012. Folksonomy-based term extraction for word cloud generation. *ACM TIST*, *3*(4), 60.

[7] Celma, Ò. & Herrera, P. 2008. A new approach to evaluating novel recommendations. *Proc. RecSys '08*, 179-186.

[8] Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., & Yu, Y. 2012. Collaborative personalized tweet recommendation. *Proc. SIGIR '12*, 661-670.

[9] Cui, P., Wang, F., Liu, S., Ou, M., Yang, S., & Sun, L. 2011. Who should share what?: item-level social influence prediction for users and posts ranking. *Proc. SIGIR '11*, 185-194.

[10] Daly, E.M., Muller, M.J., Gou, L., & Millen, D.R. 2011. Social lens: personalization around user defined collections for filtering enterprise message streams. *Proc. ICWSM '11*.

[11] Desrosiers, C. & Karypis, G. 2011. A comprehensive survey of neighborhood-based recommendation methods. *Recommender systems handbook*, 107-144.

[12] DiMicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., & Muller, M. 2008. Motivations for social networking at work. Proc. CSCW '08, 711-720.

[13] Esparza, S.G., O'Mahony, M.P., & Smyth, B. 2013. Catstream: categorising tweets for user profiling and stream filtering. *Proc. IUI '13*, 25-36.

[14] Facebook for Business – A Window into News Feed: https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed

[15] Facebook for Business – Showing More High Quality Content: https://www.facebook.com/business/news/News-Feed-FYI-Showing-More-High-Quality-Content

[16] Farrell, S., Lau, T., Nusser, S., Wilcox, E., & Muller, M. 2007. Socially augmenting employee profiles with people-tagging. *Proc. UIST '07*, 91-100.

[17] Feng, W. & Wang, J. 2013. Retweet or not?: personalized tweet re-ranking. *Proc. WSDM '13*, 577-586.

[18] Freyne, J., Berkovsky, S., Daly, E.M., & Geyer, W. 2010. Social networking feeds: recommending items of interest. *Proc. RecSys '10*, 277-280.

[19] Ge, M., Delgado-Battenfeld, C., & Jannach, D. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. *Proc. RecSys '10*, 257-260.

[20] Guy, I., Ronen, I., & Raviv, A. 2011. Personalized activity streams: sifting through the river of news. *Proc. RecSys '11*, 181-188.

[21] Guy, I., Steier, T., Barnea, M., Ronen, I., & Daniel, T. 2012. Swimming against the Streamz: search and analytics over the enterprise activity stream. *Proc. CIKM '12*, 1587-1591.

[22] Guy, I., Steier, T., Barnea, M., Ronen, I., & Daniel, T. 2013. Finger on the pulse: the value of the activity stream in the enterprise. *Proc. Interact '13*, 411-428.

[23] Guy, I., Ur, S., Ronen, I., Weber, S., & Oral, T. Best faces forward: a large-scale people search in the enterprise. *Proc. CHI '12*, 1775-1784.

[24] Herlocker, J.L., Konstan, J.A., Terveen, L.G., & Riedl, J.T. 2004. Evaluating collaborative filtering recommender systems. *ACM TOIS*, *22*(1), 5-53.

[25] Hong, L., Bekkerman, R., Adler, J., & Davison, B. D. 2012. Learning to rank social update streams. *Proc. SIGIR '12*, 651-660.

[26] Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., & Molino, P. 2008. Introducing serendipity in a content-based recommender system. *Proc. HIS '08*, 168-173.

[27] IBM Connections – Enterprise Social Software: http://www.ibm.com/software/products/conn

[28] Lunze, T., Katz, P., Röhrborn, D., & Schill, A. 2013. Stream-based recommendation for enterprise social media streams. *Proc. BIS '13*, 175-186.

[29] McCandless, M., Hatcher, E., & Gospodneti, O. 2010. Lucene in action, 2nd edition. *Manning Publications Co.*

[30] McNee, S.M., Riedl, J., & Konstan, J.A. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. Proc. *CHI EA '06,* 1097-1101.

[31] McPherson, M., Smith-Lovin, L., & Cook, J.M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

[32] Murakami, T., Mori, K., & Orihara, R. 2007. Metrics for evaluating the serendipity of recommendation lists. *Proc. JSAI '07*, 40-46.

[33] Onuma, K., Tong, H., & Faloutsos, C. 2009. TANGENT: a novel, 'surprise me', recommendation algorithm. *Proc. KDD '09*, 657-666.

[34] Paek, T., Gamon, M., Counts, S., Chickering, D. M., & Dhesi, A. 2010. Predicting the importance of newsfeed posts and social network friends. *Proc. AAAI '10*, 1419-1424.

[35] Ronen, I., Guy, I., Kravi, E., & Barnea, M. Recommending social media items to community owners. *Proc. SIGIR '14*, 243-252.

[36] Tintarev, N. & Masthoff, J. 2007. A survey of explanations in recommender systems. *3rd WPRSIUI Workshop, Proc. ICED '07*, 801-810.

[37] Zhang, Y.C., Séaghdha, D.Ó., Quercia, D., & Jambor, T. 2012. Auralist: introducing serendipity into music recommendation. *Proc. WSDM '12*, 13-22.

[38] Ziegler, C.N., McNee, S.M., Konstan, J.A., & Lausen, G. 2005. Improving recommendation lists through topic diversification. *Proc. WWW '05*, 22-32.