Agents vs. Users: Visual Recommendation of Research Talks with Multiple Dimension of Relevance

KATRIEN VERBERT, KU Leuven
DENIS PARRA, Pontificia Universidad Católica de Chile
PETER BRUSILOVSKY, University of Pittsburgh

Several approaches have been researched to help people deal with abundance of information. An important feature pioneered by social tagging systems and later used in other kinds of social systems is the ability to explore different community relevance prospects by examining items bookmarked by a specific user or items associated by various users with a specific tag. A ranked list of recommended items offered by a specific recommender engine can be considered as another relevance prospect. The problem that we address is that existing personalized social systems do not allow their users to explore and combine multiple relevance prospects. Only one prospect can be explored at any given time – a list of recommended items, a list of items bookmarked by a specific user or a list of items marked with a specific tag. In this article, we explore the notion of combining multiple relevance prospects as a way to increase effectiveness and trust. We used a visual approach to recommend papers at a conference by explicitly presenting multiple dimensions of relevance. Suggestions offered by different recommendation techniques were embodied as recommender agents to put them on the same ground as users and tags. The results of two user studies performed at academic conferences allowed us to obtain interesting insights to enhance user interfaces of personalized social systems. More specifically, effectiveness and probability of item selection increase when users are able to explore and interrelate prospects of items relevance – i.e. items bookmarked by users, recommendations and tags. Nevertheless, a less technical audience may require guidance to understand the rationale of such intersections.

Categories and Subject Descriptors: H.5.2. Information interfaces and presentation (e.g., HCI): User interfaces. H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms: Human Factors; Design; Experimentation.

Additional Key Words and Phrases: Recommender systems, visualization, intelligent user interfaces.

ACM Reference Format:

1. INTRODUCTION

Interactive information visualization and recommendation techniques have both been researched as ways to help people deal with abundance of information. The main advantage of interactive visualization is that a multi-dimensional representation allows the user to more easily see multiple aspects of data while being in control when exploring information. The main advantage of the traditional recommendation approach is that offering a clear list of items ranked by perceived interest reduces cognitive overload associated with exploring a rich set of items.

In this article, we present our research on the combination of both approaches. We investigated how graphical representations and the ability to combine a personalized prospect offered by a recommender engine with other valuable prospects can improve user trust in the results offered by the black-box recommender engines and increase user ability to find interesting items.

Our work has been motivated by the presence of multiple relevance prospects in modern social tagging systems. An important feature pioneered by social tagging systems and later used in other kinds of social systems is the ability to explore different community relevance prospects by examining items bookmarked by a specific user or items associated by various users with a specific tag. Items bookmarked by a specific user offer a social relevance prospect: if this user is known and trustable or appears to be like-minded (bookmarked a number of items known as interesting) a collection of his or her bookmarks is perceived as an interesting and relevant set that is worth to explore for more useful items. Similarly, items marked by a
specific tag offer a content relevance prospect. Items related to a tag of interest or a tag that was used to mark many known interesting items are also perceived as potentially relevant and worth to explore. In a social tagging system extended with a personalized recommender engine [Parra and Brusilovsky 2009; Peng et al. 2010; Gemmell et al. 2011], top items recommended to a user offer a personalized relevance prospect. If the user believes that a recommender system does a good job in selecting most relevant items for her, these items are certainly worth exploring. Note that several recommender engines can operate in the same social tagging system with each of them offering a set of items that is especially relevant to the user from the prospect of this engine.

The problem that we address in this article is that existing personalized social systems do not allow their users to explore and combine multiple relevance prospects. Only one prospect can be explored at any given time – a list of items ranked by a specific recommender engine, a list of items bookmarked by a specific user or a list of items marked with a specific tag. We believe that exploring a single prospect is not sufficient since none of the prospects could be fully reliable and trustable by the users (that includes recommendations generated by black-box engines). In this context, the ability to combine prospects might offer a more reliable and attractive way to explore. For example, knowing that a specific item has been not only recommended by a recommender engine, but also bookmarked by two trustworthy users or that an item is recommended by two different recommender engines can remarkably increase user trust in the value of this item.

To solve the aforementioned problem and to offer users an approach to explore and combine multiple relevance prospects in personalized social tagging systems, we suggest a multi-prospect interactive visualization approach. The first key idea of our approach is to embody suggestions offered by various recommender engines within the same social tagging system as recommender agents that can be perceived as being analogous to human users. This makes personalized relevance prospects offered by recommender engines similar to the relevance prospects offered by users and tags in social systems. The second key idea is to offer a relevance-based interactive visualization interface that enables social tagging system users to explore multiple aspects of relevance in parallel, while offering special provisions for combining multiple relevance prospects (i.e., focus on items that are relevant from two or more prospects). We believe that using a multi-prospect visualization could help users in locating the most valuable items while also increasing transparency of the recommendation process and allowing the users to be in control of their exploration.

To assess the value of our ideas, we implemented the proposed approach in the context of a social conference system, Conference Navigator, that offers social bookmarking and recommender services for attendees of research conferences [Parra et al. 2012]. The resulting visual interface uses a cluster map visualization representation and was released as a component of Conference Navigator to the attendees of several research conferences. The conference context was used to run two user studies of our approach at two research conferences. In these studies we specifically focused on the expected values of multi-prospect interactive visualization such as its effect on findability of relevant items and trust of recommendations, while also collecting a broader set of data to better understand the bigger picture of its impact.

The first study was a controlled experiment in which users were asked to complete specific tasks with the visualization while the second was a field study in which users explored the visualization as a part of their regular work with the system. Results of these user studies indicate that (1) the ability to combine multiple sources of relevant information provides an effective way to find relevant items and (2) the specific cluster map visualization design offers a reasonable support for combining multiple relevance prospects, but it has its limitations.

This article is organized as follows: Section 2 presents related work in the area of relevance-based visualizations, recommender systems and user interfaces for such systems. Section 3 introduces our multi-prospect visualization approach and its implementation in TalkExplorer, a cluster map visualization of conference talks and their relations to users, tags, and recommendations. Sections 4 to 6 introduce our user studies and present the results of the controlled user study of TalkExplorer, followed by the results of a second observational user study. Finally, sections 7 and 8 discuss the results of the studies, lessons learnt and future research opportunities.

2. BACKGROUND AND RELATED WORK

2.1 Relevance-based visualizations

One of the important features of TalkExplorer presented later in the article is the application of so-called relevance-based visualization to visually recommend conference talks to their users. Relevance-based
visualization was originally developed in the field of information retrieval for visualization of search results. The idea of this visualization was to avoid limitations of the traditional one-dimension ranked list organization of results and instead to stress visually which results were relevant to different components of complex multi-term queries. We can distinguish three considerably different types of relevance-based visualization – direct, POI-based and set-based. Direct relevance visualization aims to separately indicate to which of the query terms each retrieval document is relevant and to what extent. The latter is usually visualized by color or greyscale density. Due to the simplicity of this approach, it has been independently reinvented several times and used in both research and commercial systems. Good examples of direct relevance visualization are HotMap [Hoebber and Yang 2006] and Pie Charts [Anderson et al. 2002]. The classic example of this approach is TileBars [Hearst 1995], which is actually more sophisticated than the majority of examples in this category since it is able to separately show term relevance not just for the whole documents, but for their fragments.

Unlike linearly organized direct relevance visualization, POI-based visualization attempts to build a truly spatial visualization of retrieved documents in relation to query terms. It treats different query terms as separate points of interests (POI) and places the retrieved documents in the visualization topology according to their relevance to each of these terms: the more a document is relevant for the term, the closer to the term POI it is displayed. This approach has been originally suggested in VIBE [Olsen et al. 1993] and later expanded in several similar systems such as VR-VIBE [Benford et al. 1995], INSPIRE’s Galaxy View [Wise et al. 1995], LyberWorld [Hemmje et al. 1994], WebSearchViz [Nguyen and Zhang 2006] and Radvis [Sharko et al. 2008]. A strong point of the POI-based approach is its interactive nature. The user is allowed to manipulate POIs causing movement of documents related to the manipulated POI. As a result, this approach supports both visualization and exploration of the document space.

Set-based visualization applies an alternative set-focused approach to spatially organize search results. For example, for a query that uses three terms, it will create seven set areas to show which results are relevant to each of the three terms, each of two pairs of these terms, and all three terms at the same time. The classic example of set-based relevance visualization is InfoCrystal [Spoerri 1993]. The Aduna visualization approach [Aduna 2014] used by TalkExplorer also belongs to this category, offering a more complex visualization paradigm and a better level of interactivity. A strong point of the set-based approach is a clear representation to which of the query terms each document is relevant along with grouping documents by this aspect. While the POI-based approach allows to guess to which terms specific documents are relevant, this guess should be checked by trying to move all potentially relevant POIs.

The novelty of the approach suggested in our article is twofold. First, we are using a set-based relevance approach not just with keywords or tags where relevance approaches are usually applied, but with a diverse set or relevance-bearing entities (tags, users, recommendations). Second, our visualization is not just reactive to the selected query, but also adaptive. The set of entities for manipulation is selected for each user based on his or her past experience. For entities representing retrieval agents even the selection of documents is adapted to the user. The only other example of adaptive relevance-based visualization known to us is Adaptive VIBE [Ahn and Brusilovsky 2013] that combines reactive query terms with adaptively selected terms from the long-term user model. Our visualization, however, goes beyond Adaptive VIBE in the diversity of used entities.

2.2 Recommender Systems
Recommender systems help users find relevant items in an overloaded information space [McNee et al. 2006]. The usual approach to achieve this goal is to implement algorithms that automatically create lists of relevant items based on the past user’s preferences. These methods can be broadly categorized in three groups: (a) collaborative filtering, (b) content-based filtering, and (c) hybrid recommenders. Collaborative filtering methods recognize commonalities between users or between items on the basis of explicit or implicit relevance indications [Herlocker et al. 2004] such as ratings [Bennet and Lanning 2007] and tags [Parra and Brusilovsky 2009]. Implicit data used by recommender systems include actions like reading [Morita and Shinoda 1994] or watching TV series [Hu and Koren 2008]. A standard user-based collaborative filtering algorithm first identifies similar users based on their overlapping interactions or similar ratings of common items. It then makes recommendations based on preferences of these like-minded users, the nearest neighbors. On the other hand, a standard item-based recommendation algorithm analyzes similarities between items and then uses these similar items to identify the set of items to be recommended.
Collaborative filtering is the most widely implemented and most mature technology [Burke 2002], but it tends to suffer from data sparsity and from cold start issues, i.e. these systems cannot make effective recommendations for new users or for new items that have no explicit or implicit relevance indicators yet [Burke 2010].

An option to deal with the problems of collaborative filtering is using content-based filtering, which matches descriptions of items to descriptions of users [Pazzani and Billsus 2007]. This approach bases predictions on information about individual users and items, and ignores contributions from other users. Unlike collaborative filtering, a content-based recommender can suggest items that have not been rated or consumed, and it relies solely on the active user profile to create recommendations, decreasing the problem of data sparsity in collaborative filtering. However, content-based recommenders have problems of over specialization – by default they do not optimize for serendipity or novelty among the items recommended – and they still require some preference feedback from the active user in order to generate a user profile and produce recommendations [Lops et al. 2011]. The third category of these systems, hybrid recommender systems, combines recommendation techniques to gain better performance with fewer drawbacks [Burke 2002]. Adomavicious et al [2005] summarize four different ways to combine the collaborative and content-based approaches into a hybrid recommender: combining separated recommenders, adding content-based characteristics to collaborative methods, and vice versa, and developing a unifying approach. Hybrid approaches have shown empirically a better performance than only collaborative and content-based approaches, but these methods focus on improving the performance on prediction accuracy and do not consider other measures of success in recommender systems.

One emergent research direction to tackle these challenges, which departs from the initial focus on improving algorithmic accuracy, is incorporating the user experience into the evaluation of the recommender performance. In a recent article, Konstan and Riedl [2012] surveyed existing research on recommender systems from a user-centric perspective, showing the importance of diversity [Ziegler et al. 2005], the impact of transparency on the user trust [Sinha and Swearingen 2002] and the role of explainability [Tintarev 2007] on improving the user experience. Moreover, Konstan and Riedl [2012] show in their review scarce research on interactive recommendation dialogues and they acknowledge that further work needs to be done to understand the effect of explanations and transparency as dimensions that give some control back to the user. This expectation has a good foundation, since previous research has shown that small increases in prediction accuracy doesn’t reflect a better user experience on a recommender [Sinha and Swearingen 2001], or that recommender interfaces can lead to better user retention without increasing prediction quality [McNee et al. 2003]. In this article, we follow this line of research and contribute by implementing and evaluating an interactive visual recommendation interface that supports enhanced user control and transparency. We are particularly interested to enable user interaction with recommender systems as a basis to support exploration and controllability and to study its effect on the user experience, measured with several metrics and dimensions.

2.3 Visualizing Recommender Interfaces

The traditional way to present recommendations is a ranked list of items sorted by some relevance score. However, providing alternative visual representations and adding interface interaction can enhance user satisfaction with the system due to improved system explainability [Tintarev and Masthoff 2011] and user controllability [Knijnenburg et al. 2012, Bostandjiev et al. 2012, Parra et al. 2014]. Most existing work in the area of visualizing recommendations focuses on interaction with collaborative filtering recommender systems. PeerChooser [O’Donovan et al. 2008] is a visual interactive recommender that uses a graph-based representation to show relations between users and recommended items of a collaborative filtering recommender system. Similarly, SmallWorlds [Gretarsson et al. 2010] allows exploration of relationships between recommended items and similar friends, in multiple layers of similarity. These systems enable users to explore such relationships as a basis to provide transparency and to support the user to find new relevant items. Pharos [Zhao et al. 2010] is a social map-based recommender system that visualizes a summary of social network activity of different communities. The system uses topic modeling [Blei et al. 2003] to provide new users with an overview of the site in order to alleviate the cold start problem of collaborative filtering recommenders.

Some systems focus specifically on tags that are used by social recommenders. SFViz (Social Friends Visualization) [Gou et al. 2011] visualizes social connections among users and user interests as a basis to
increase awareness in a social network and to help people find potential friends with similar interests. This system uses a Radial Space-Filling (RSF) technique [Chuah 1998] to visualize a tag tree and a circle layout with edge bundling to show a social network.

FaceTag [Tonkin et al. 2008] is a tool that helps users see the relationship between user-generated tags and recommended facets in a classification scheme. Tagsplanations [Vig et al. 2009] presents both tag relevance (the degree to which a tag describes an item) and tag preference (the user’s sentiment toward a tag) as a basis to explain recommendations. Kammerer et al. [2009] designed a tag-based search browser to recommend relevant tags for further search. Research on this stream only focuses on information and meta-information concerning items, and ignores the users who contributed such information and relationships among those users [Gou et al. 2011].

Another area that has been revived recently is the recommendation of visualization. The initial work of Gotz et al. (2008) to recommend visualizations has been extended by the framework proposal of Mutlu et al. (2015). These works show an interesting area since visual analytics has been growing interest in several communities recently, and being able to recommend users the best visualization to unveil patterns under certain decision-making tasks could have a strong impact to create an area of personalized visual analytics. The present article develops a different task, which is using a specific type of visualization, multimodal graphs, rather than recommending visualizations to the user to explore the data.

Other recent areas of related work focus on personalized and interactive graph navigation that leads to recommendations through users’ exploration. For instance, Chau et al. [2012] introduced the Apolo system, whose primary focus is to allow users to make sense of large network datasets, but also guides them to find relevant articles through an incremental guided exploration. This system updates the personalized recommendations by means of the Belief Propagation algorithm that leverages user preference inferred from her interaction with the system. Another example is the work of Crnovrsanin et al. [2011] that focuses on adaptive network visualization. Their system adjusts in real-time the graph layout and node visibility to accommodate recommendations and to reduce the visual clutter of nodes that users should explore. ForceSPIRE [Endert et al. 2012] and StarSPIRE [Bradel et al. 2014] are promising examples that adapt spatial visualizations based on expert knowledge. The systems incorporate so-called semantic interactions (document movement, text highlighting, search and annotation) to update the underlying model of the visualization. The iCluster system [Drucker et al. 2011] uses a similar mixed initiative approach to improve clustering of documents. These approaches are more generally applicable to visual analytics research than our work that has a focus on improving interfaces of recommender systems. A more recent system that also deals with displaying graph information with recommendations, but with a focus on big data graphs over an array of screens is iGraph. The visualization algorithm relies on multiple GPUs and CPUs in order to scale the visualization in a display wall of up to 50 million pixels. But unlike these systems that deal with visualizations with only one type of node, our work represents an interactive multimodal network, i.e., a network with different types of nodes that represent diverse dimensions of relevance (tags, users, agents).

Some systems have implemented interactive visualizations of hybrid recommender systems to allow controllable and explainable personalization of different recommender methods. TasteWeights [Bostandjiev et al. 2012] was introduced as a system that allows users to control the importance of friends and peers in social systems to obtain recommendations. Similar to our work, TasteWeights introduces the concept of an interface for hybrid recommender systems. The system elicits preference data and relevance feedback from users at run-time and uses these data to adapt recommendations to the current needs of the user. To our knowledge, this is one of the first systems that enables interaction with a hybrid recommender system and that can be adjusted by end users to control the output of the systems. In our work, we extend this concept of visualizing and combining the output of multiple recommenders as a basis to support exploration and controllability. While our work also enables end users to explore intersections of multiple recommenders (and hence in this respect we build on concepts that have been introduced by TasteWeights), the major difference and innovation of our work is that we allow end users to combine multiple relevance prospects in order to increase the perceived relevance and meaning of items. Our visualization embodies suggestions offered by various recommender systems as recommender agents. Items bookmarked by a specific user offer a social relevance prospect: if this user is known or appears to be like-minded a collection of her bookmarks is perceived as an interesting and relevant set that is worth to explore. Similarly, items marked by a specific tag offer a content relevance prospect. In a very recent work inspired by TasteWeights and by our initial work with TalkExplorer, Parra et al [2014] created SetFusion, a visual recommender that allows the user to fuse
different prospects of relevance using sliders and an interactive Venn diagram that complement a ranked list of recommended items. The sliders allow the users to control the importance of each context of relevance, while the Venn diagram lets the users filter the list of items recommended, and at the same time make sense of how certain items are recommended by one or more contexts. To the best of our knowledge, this combination of multiple relevance prospects (agents, tags and real users) has not yet been explored as a means to support exploration and controllability, and to increase trust and acceptance of recommendations. Secondly, to our knowledge, our work is the first attempt to evaluate whether enabling end users to combine multiple relevance prospects increases effectiveness and probability of item selection.

3. MULTI-PROSPECT VISUALIZATION
The problem that we address is that existing personalized social systems do not allow their users to explore and combine multiple relevance prospects. Only one prospect – such as a list of recommendations, items bookmarked by a specific user or items with a particular tag – can be explored at a given time. We believe that the ability to combine prospects might offer a more reliable and attractive way to explore items. For example, knowing that a specific item has been not only recommended by a recommender engine, but also bookmarked by another trustable user can increase user trust in the value of this item.

To assess the value of our ideas, we implemented the proposed approach in the context of a social conference system, Conference Navigator, that offers social bookmarking and recommender services for attendees of research conferences [Parra et al. 2012]. We first present Conference Navigator 3 and social features that it provides. Then, we present the objectives of visualizing multiple relevance prospects of this system and details on the design and development of our visual interface – named TalkExplorer.

3.1 Conference Navigator
Conference Navigator 3 (CN3) is a social personalized system that aims to support users in several activities related to academic conferences [Parra et al. 2012]. At the time of writing, 30 conferences have been supported by CN3. Among different features, Conference Navigator provides a conference schedule, a list of the conference talks and details of each talk (illustrated in Fig. 1). It also provides information about people related to the conference, such as the list of attendees and the list of authors. Users can add papers to create a personal schedule, they can add tags to each talk, and they can also connect with other CN3 users by following them (unidirectional relationship) or connecting with them (bidirectional relationship). Social information collected by CN3 is extensively used to help users find interesting papers. For example, in the page called “Top Items”, CN3 summarizes the most popular papers, the most active people, their institutions, and also the most popular tags associated to the papers.

When visiting talk information, as shown in Fig. 1, users can also see who scheduled each talk during the conference and which tags were assigned to this talk. This social information is also used to provide links to similar papers (“People who scheduled this presentation, also scheduled:”) mimicking the well-known Amazon.com’s suggestions [Linden et al. 2003]. Similarly, when visiting a user page, other users can see which talks this user is planning to attend, given that personal privacy settings provide access to this information. Finally, talks marked with a specific tag by the community of users can be explored.

In addition to social information access, CN3 offers personal recommendation of talks and papers. CN3 supports two kinds of recommendations: content-based and tag-based. The content-based engine constructs the user interest profile as a vector of terms with weights based on TF-IDF [Baeza-Yates and Ribeiro-Neto 2011] using the content of the papers that the user has scheduled. Then, it recommends papers that match the profile of interests. The tag-based recommender engine makes use of the tags that users associate to conference talks. By matching user tags (tags applied by the active user) with item tags (tags assigned to different talks by the community of users) using the Okapi BM25 algorithm [Manning et al. 2008], the engine identifies relevant talks and suggests them to the active user. These recommendations are presented as a ranked list – as illustrated in Fig. 2. As a baseline for the studies reported in this paper, we measured how often users interacted with these recommendations pages and how often they followed a recommendation. Thus, the baseline for our studies is the success rate of recommendations when they are presented as a ranked list. The data is reported in Section 5 (controlled user study) and Section 6 (field study).
3.2 Visualizing Recommendations, Tags and Users

In the original CN3, ranked lists produced by the content-based and tag-based recommenders are presented in separate pages and can be used by users to find new talks. In addition, users can explore bookmarks and
tags of other users as a basis to find new items. In this article, we explore the value of a multi-prospect interactive visualization approach that enables end users to explore and combine these multiple relevance prospects. The suggested approach was implemented in TalkExplorer, an interactive visualization developed on top of CN3. The TalkExplorer implementation was driven by two key ideas. The first idea is to embody suggestions offered by various recommender engines as recommender agents that can be perceived as being analogous to human users. This makes personalized relevance prospects offered by recommender engines similar to the relevance prospects offered by users and tags in social systems. The second key idea is to offer a relevance-based interactive visualization interface that enables users to combine multiple aspects of relevance – i.e. to filter out those items that are relevant from two or more prospects, such as items that are recommended by an agent and also bookmarked by a trustable user. In the context of recommender system interfaces, we were also interested in offering new ways for end users to interact with recommender engines and items that are suggested to them. The resulting interface serves to both explain the provenance of recommendations in a transparent way and to support exploration and control by end users. More importantly, users can browse and interrelate bookmarks of other users, tags and suggestions of recommender agents as a basis to find relevant items. We selected the set-based Aduna cluster map visualization library [Aduna 2014] to test our ideas. This software library visualizes sets of categorized objects and their interrelationships. The library has been used in related research to explore the interaction of users, resources and tags in social tagging systems [Klerkx and Duval 2009]. We extended the approach to represent recommendations embodied as agents. We discuss other set-based visualization techniques that can be used to evaluate our approach in the Section 7.

While our approach allows multiple recommender systems to be embodied and presented as agents, we started with two different engines, a content-based and a tag-based recommender, that were visible to the users as content-based and tag-based recommender agents. Each agent is associated with the set of top items recommended by the corresponding engine. In parallel with agents, the most relevant users (including the active user herself) are shown in the list of prospects. Each user is associated with a set of talks that he or she bookmarked. As shown in Fig. 3 (left panel), fifteen users from the neighborhood of the active user are shown and users can explore items that these users have bookmarked as a basis to find new items. The selection of these users is based on uni- and bidirectional connections between users - i.e. users that the active user follows (user name preceded by [f] in Fig. 3) and connects to (user name preceded by [c]), respectively. In addition, the neighborhood includes users that have similar interests based on common bookmarks (preceded by [s]). The multi-prospect visualization interface allows the interrelationship between all these prospects to be explored. With a simple selection of two or more relevance prospects, users could explore interrelationships between two agents, two or more users as well as interrelationships between agents and users (i.e. which other users have bookmarked talks that are recommended to them by one or more agents). Finally, relationships with tags can be explored in the same way. In the multi-prospect interface, each tag is associated with items that were tagged with it. As a result, the interface allows exploring interrelationship between tags, tags and agents, tags and users and more complex combinations using the same regular approach.

TalkExplorer includes three principal components, as shown in Fig. 3. On the left side, the entity selection panel allows users to select tags, users and recommender agents that are added and displayed in the canvas area. This canvas area, at the center of the screen, shows a clustermap visualization - i.e., different clusters of tags linked by connected components. The labeled circles in this canvas area represent either real users, recommender agents or tags. Yellow circles represent individual talks, and the bubbles that involve them represent clusters of talks. In Fig. 3, two users with their bookmarked talks are shown (P Brusilovsky and L Aroyo), as well as suggestions of the tag-based and content-based recommender agent. The clustermap visualization enables users to explore relationships between items that were suggested to them by these recommender agents and bookmarks of users on the screen. For instance, users can see which other users have bookmarked a talk that is suggested by a recommender agent by exploring the intersection of the agent and a specific user. In the example presented in Fig. 3, the active user (P Brusilovsky) can explore which of his bookmarked talks are also bookmarked by user L Aroyo (label 1), which additional talks are bookmarked by L Aroyo but not recommended by any agent (label 2) and which talks are recommended to him by both recommender agents and are also bookmarked by L Aroyo (label 3). Finally, the rightmost panel shows the detailed list of talks, either all talks presented in the canvas area, or a subset of them related to the selected
entity. If users click on a cluster (for example, the cluster showing talks that were bookmarked by L Aroyo and a specific agent) the list of these talks and their details are presented.

![Cluster map representation of bookmarks of the active user (P Brusilovsky) and two other users (T Ley and K Verbert)](image)

Users can explore the information space by using the cluster map visualization in a number of ways. First of all, users can explore their own bookmarks, explore bookmarks of other users, see which tags are used to annotate talks, and explore these tags to find relevant talks. Users start with a visualization of their bookmarks, along with bookmarks of two other users with whom they shares most interests. Fig. 4, shows for example the bookmarks of the active user (P Brusilovsky) and their relationships to the bookmarks of two other users (K Verbert and T Ley). The set-based cluster visualization on this figure shows bookmarks that were made by one user only, each pair of users, and all three users. Fig. 5 shows that users can explore these bookmarks individually, by users, and by clusters. To examine individual talks, users can hover over bookmarks represented by yellow circles to see the details of particular bookmarks (Fig. 5.a). Alternatively, they can click on a user in the visualization, as shown in Fig. 5.b. By doing so, details of all talks bookmarked by the users (that could be spread over several clusters) are shown in the right side panel (Fig. 5.c). For each talk the panel shows the title, authors, tags of the talks, and users who bookmarked the talk. Users can access the full details of the talk by clicking on the “open in browser” link, that will open the corresponding page in CN3. Users can also bookmark the talk by clicking on the bookmark link. The talk will then be added to their schedule in CN3.

![Cluster map representation of bookmarks of the active user (P Brusilovsky) and two other users (T Ley and K Verbert)](image)
Fig 5. Exploring bookmarks: (a) users can hover over talks in the cluster map, represented by yellow circles; Alternatively, users can also be selected in the cluster map (b) to explore details of bookmarks of this user in the results panel (c). The bookmarks of this selected user are highlighted in yellow. Other bookmarks are represented in grey.

While only the two most similar users are shown at the start, users can examine any subset of related users shown on the left side panel of TalkExplorer. To add (or remove) a user to the visualization, users have to select (or unselect) the check box next to the name of the user. When a new user is selected, the bookmarks of this user are added to the visualization (Fig. 6.a) and immediately split into clusters representing talks bookmarked by this user only, and talk overlaps with users already present in the visualization (Fig 6.b). The clustermap ability to show the overlaps of talks bookmarked by two or more users as separate clusters supports more advanced exploration. For example, Fig. 6 presents the active user (P Brusilovsky) and two similar users with whom he shares bookmarks (T Cochrane, A Karaseva). In search for relevant talks, users can select a cluster of talks bookmarked by T Cochrane but not yet bookmarked by themselves (Fig. 6.b) or a cluster of talks bookmarked by both T Cochrane and A Karaseva (Fig. 6.c). The fact that two similar users bookmarked these talks hints that these talks might be worth to examine. The number next to user names in the left side panel (Fig. 6.a) and in the cluster map visualization (Fig. 6.b) represents the number of total bookmarks of these users. When a (sub)set of these bookmarks is selected in the cluster map, for instance the set of talks that have been bookmarked by both T Cochrane and A Karaseva (Fig. 6.c), the number of bookmarks in this subset is represented next to the user name too. In Fig. 6.c, this subset contains 11 items out of 28 total bookmarks of T Cochrane (11/28) and 23 bookmarks of A Karaseva (11/23).

3.2.1 (a)

(b)

3.2.2 (c)

Fig 6. Exploring bookmarks of users by selecting sets of bookmarks of related users in the cluster map (b) or by exploring bookmarks in the intersection of two users (c). Additional users can be added with the navigation panel on the left side (a).
To further support users in finding relevant talks, TalkExplorer embodies different recommender engines as agents that “bookmarked” top talks recommended by the corresponding engine. To interact with agents, users can use the panel on the left side and add a specific agent to the visualization just like real users (Fig. 7.a). This design puts recommenders on the same ground as users, expanding the user exploration approach introduced above to work with recommendations of these agents. For example, users could explore which recommendations have been suggested by multiple recommender agents, such as a tag-based recommender agent and a content-based recommender agent (Fig. 7.b). Users can also explore which recommendations have been bookmarked by other users (Fig. 7.c). The key idea is again that these overlaps can be used as additional relevance indicators: talks that are selected by multiple agents or users can potentially have higher relevance.

In a similar way, tags can be added to the visualization in parallel with users and agents. The left side panel contains all tags that have been added to talks (Fig. 8.a). Tags are by default ranked by name, but they can be rearranged by frequency. By selecting the check box next to a specific tag, this tag is added to the visualization as another relevance prospect and the talks with this tag are added to the visualization in several clusters according to their relationships with other relevance prospects (tags, users, agents) shown in the canvas. For example, Fig. 8.b shows which talks that are recommended by the tag-based agent are also labeled by the tag community. Users can then explore talks with this tag in their relationships with bookmarks of users and recommendations of agents (Fig. 8.c).
Exploring tags. Tags can be added to the visualization by selecting them in the left side pane. Users can also explore which recommendations have been annotated with a specific tag (b) to filter out the potentially more relevant recommendations, and explore interrelations with users (c).

TalkExplorer is implemented as a Java applet and uses the Aduna clustermap visualization library [Aduna 2014]. We started from the implementation of Klerkx and Duval [2009] and extended it to visualize the interactions of users, tags, and recommender agents in terms of conference talks that they have in common – as illustrated in Fig. 3. To do that in Aduna’s set-based format, we had to convert a ranked list produced by each recommender engine into a set top-10 agent “bookmarks”. While the transition for a ranked list to a set implied some loss of information, by placing multiple relevance prospects (social, content, personalized recommendations) on the same level, this design enabled multi-prospect visualization allowing users to examine various prospects and their overlaps. While extending the original approach, we kept some of its good features in the TalkExplorer design. For example, we chose to let users start with a visualization of their own bookmarks just as it was done in [Klerkx and Duval 2009], which, in turn, followed the philosophy “start with what you know, then grow” [Heer et al., 2005]. After the first (controlled) user study (see Section 5) a few modifications were implemented to address problems that came out of the think-aloud data. Most importantly, we added functionality to bookmark talks from within the visualization. In the original version, users had to navigate to the talk page of CN3 to bookmark it. In the new version that was used in the second (field) study (see Section 6), a bookmark link is added for each talk in the right side panel (see Fig. 5.c). We also realized that it is hard for users to distinguish their bookmarked talks when their name is not selected for the visualization and implemented a clear feedback for a situation where users attempted to bookmark an earlier bookmarked talk.

4. ASSESSING THE VALUE OF MULTI-PROSPECT VISUALIZATION

As mentioned above, TalkExplorer was released to the attendees of several conferences as a component of Conference Navigator. This provision helped us to run two realistic user studies to assess the value of our multi-prospect visualization. In these studies we focused on several kinds of impact provided by TalkExplorer such as increased relevancy and meaning of recommendations, effectiveness to find relevant items, and trust. We specifically examined the value of integrating more than one relevance prospect that was the key feature of the interface. Among the questions we attempted to answer were the following: could we help users in the task of talk selection by showing multiple relevancies? Could it increase trust in recommendation while refining choice for people and tags? To what extent overlapping relevancies could help users to find relevant talks? Which of multiple relevance prospects and their overlaps are most valuable in finding good talks?

The first study (N=21) was designed as a controlled experiment with users at two conferences. Users were asked to perform three tasks (exploring users in their neighborhood, exploring agents and exploring tags).
We used screen recording, captured think-aloud data, and collected user feedback. The controlled nature of this study helped us to gain important insights into the relative effectiveness of each of these entities and provided evidence that multiple relevance prospects are useful. The details of this study and the obtained results are presented in Section 5. At the same time, this controlled user study suffered from some limitations:

- Participants were asked to complete three tasks: exploring users, exploring agents and exploring tags. For each of these entities (users, agents, tags), we asked to try to find a relevant talk to attend at the conference by exploring the entities.
- The order of the tasks was fixed. First they explored the users, then the agents and then the tags.

This controlled nature makes it hard to make assessments about the use and the value of TalkExplorer in natural user-driven settings. In the second study (N=18), we deployed TalkExplorer at two other conferences and asked users to explore the visualization without giving them any specific tasks. The users were free to interact with the visualization to address their own needs or curiosity. All interactions with the visualization were logged. We have measured the number of times that they explored users, agents and tags and how often these entities and their combinations were successful to find a relevant talk. With this second field study, we addressed some of the limitations of the first study where the set of tasks was fixed and gained insight into the usefulness of the visualization in an open setting. The details of this field study are presented in Section 6. A summary of the key data related to four conferences used in our studies is presented in Table 1.

### Table 1. CN3 data user studies

<table>
<thead>
<tr>
<th>4.1.1 Conference</th>
<th>4.1.2 Number of users</th>
<th>4.1.3 Number of talks</th>
<th>4.1.4 Number of tags</th>
<th>4.1.5 Number of bookmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.6 Controlled user study</td>
<td>4.1.8 HT 2012</td>
<td>4.1.9 6</td>
<td>4.1.10 2</td>
<td>4.1.11 4</td>
</tr>
<tr>
<td>4.1.12 UMAP 2012</td>
<td>4.1.13 1</td>
<td>4.1.14 1</td>
<td>4.1.15 6</td>
<td>4.1.16 9</td>
</tr>
<tr>
<td>4.1.17 Field study</td>
<td>4.1.19 E C-TEL 2012</td>
<td>4.1.20 1</td>
<td>4.1.21 1</td>
<td>4.1.22 3</td>
</tr>
<tr>
<td>4.1.18 LAK 2013</td>
<td>4.1.24 1</td>
<td>4.1.25 5</td>
<td>4.1.26 5</td>
<td>4.1.27 1</td>
</tr>
</tbody>
</table>

5. **CONTROLLED USER STUDY**

In the first user study, we have deployed TalkExplorer at two conferences where CN3 was used as the main conference assistance system: the ACM Hypertext 2012 conference in June 2012 (HT 2012) and the User Modeling, Adaptation, and Personalization conference in July 2012 (UMAP 2012). Both evaluations were performed with attendees of respective conferences using real conference data (i.e., using actual talks and schedules and bookmarks, tags and ratings of the conference participants). Users were asked to explore conference talks using the visualization provided by TalkExplorer. As explained in Section 3, the
visualization provided access to the content-based and tag-based recommender agents and allowed to explore talks bookmarked by users or tagged with user-employed tags.

5.1 Participants
In total, fourteen HT 2012 attendees and seven UMAP 2012 attendees participated in a controlled study. To assess their domain knowledge, we inquired about the number of conferences in the corresponding series that participants attended, as well as their knowledge of recommendation and visualization techniques, respectively. The latter were rated on a five point Likert scale. On average, HT 2012 participants attended 1.5 Hypertext conferences in the past (std. deviation 0.732) and UMAP 2012 participants attended 2 UMAP conferences (std. deviation 1.5). At both conferences, participants’ average familiarity with recommendation techniques was high – 3.7 at HT 2012 (std. dev. 0.8) and also 3.7 at UMAP 2012 (std. deviation 0.95). They also have a high level of familiarity with visualization techniques - 4.285 at HT 2012 (std. deviation 0.7) and 4.2 at UMAP 2012 (std. deviation 0.76).

5.2 Experimental Setup and Data Collection
To address cold-start issues, participants were first asked to bookmark and tag at least five talks of interest. This data was used to generate recommendations and to select users to show in the visualization. After that, we provided a short explanation of the visualization interface. We explained the concept of agents that represent recommendations of different recommendation algorithms and the cluster map visualization that enables to interrelate bookmarks of users, recommendations of agents and tags of talks. We also explained the think aloud procedure and briefly introduced three tasks that the users were asked to complete next. The tasks had to be completed in the following order:
1. In the first task, they were asked to find a new relevant talk by exploring talks that users in their neighborhood bookmarked (Task 1 – T1).
2. In the second task, subjects had to find a new relevant talk by exploring the content-based and tag-based recommender agents (Task 2 – T2).
3. In the third task, they were asked to find a new relevant talk by exploring the tags (Task 3 – T3).

Data was collected in two ways. The think aloud protocol was used during the task stage of the study to increase the bandwidth of feedback from participants. We recorded the screen and voice of participants using Camtasia Studio [2014]. At the end of the study, participants were asked to fill out a survey inquiring about their needs at a conference and the usefulness of the visualization to address these needs.

5.3 Measures
In this section and the remainder of this article, we refer to selectable users, agents and tags as entities in the visualization. Papers or talks associated with these entities are referred to as items. We refer to intersections of entities when multiple entities were selected at the same time and their common items were explored.

To assess the value of interactive multi-prospect visualization offered by TalkExplorer, we have analyzed the way in which users explore and use the visualization. These interactions are counted as follows:
1. Explorations: when a user selects or hovers over an entity in the visualization, this interaction is counted as an exploration. For instance, when a user selects an agent and explores items that are recommended by this agent, this interaction is counted as an exploration of the agent. Likewise, if the user hovers over or selects an item in the intersection of two agents, as in Fig. 7.b, the select or hover interaction is counted as an exploration of the intersection of two agents.
2. Uses: when the user decides to add the item to his schedule – for instance by clicking on the bookmark link (see Fig. 5.c) – this interaction is counted as a use of the intersection of two agents.
We used these *explore* and *use* counts to measure the *effectiveness* and *yield* of different combinations of entities. These measures are used to gain insight in the relative success rate of different combinations of entities to find relevant items.

*Effectiveness* measures how frequently a specific combination type produced a display that was used to bookmark at least one interesting item. It is calculated as the number of cases where the exploration of this combination type resulted in a bookmark, divided by the total number of times this combination type was explored. For instance, the set of items of a related user was explored 75 times by all participants. Twenty-three of these sets were used to bookmark a new item. Thus, the effectiveness of exploring the set of items of a specific user is 23/75=31%. The number of item sets explored and the item sets used to bookmark a relevant talk, as well as the *effectiveness*, are presented in Fig. 10.

In addition, we counted the number of items in the sets where the selection was made to check *yield* (productivity) of different kinds of sets. The yield of a specific combination type was measured by summing up the total number of selections made from each combination type, divided by the total number of items that were shown in the combinations where the selection was made. In other words, yield measures a chance of a random item shown in a specific combination type to be useful (i.e. bookmarked by the user). Yield results are presented in Fig. 11.

In Fig. 9 presents an example where the active user, E Duval, used the intersection of two other users as a basis to find a relevant item. In this example, E Duval used the set of eight items in the intersection of two other users (P Brusilovsky and D Parra) to find an item. The yield indicates the number of selections made from a specific set of entities divided by the sum of the number of items in this set (eight in the example presented in Fig. 9).

For the intersection of two users, twice an item was selected out of two items and twice an item was selected out of a set containing one item. We have yield or probability of selection of (1+1+1+1)/(2+2+1+1)=0.66.

We also use a similar acceptance measure to compare the performance of TalkExplorer with the performance of the traditional ranked list recommendation approach in the context of the same conferences. These data were extracted from interactions with CN3 recommendation pages of HT 2012 and UMAP 2012, as illustrated in Fig. 2. We measured how often users interacted with ranked list recommendation pages and how often they added a recommendation to their schedule as a result of this interaction (Table 2). Considering that accepted recommendations are those recommendations that were followed by users – i.e. added to their personal schedule – we defined *acceptance* as the ratio of the number of successful recommendations to the total number of explorations of ranked list recommendations. The acceptance data for HT2012 and UMAP 2012 provided in Table 2 were used as an additional baseline in assessing the value of TalkExplorer.
Table 2. Baseline data controlled user study

<table>
<thead>
<tr>
<th></th>
<th>Number of explorations of recommendations</th>
<th>Number of recommendations</th>
<th>Acceptance rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.1</td>
<td>5.3.2</td>
<td>5.3.3</td>
<td>0.25</td>
</tr>
<tr>
<td>HT 2012</td>
<td>101</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>5.3.4</td>
<td>5.3.5</td>
<td>5.3.6</td>
<td>0.29</td>
</tr>
<tr>
<td>UMAP 2012</td>
<td>178</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>5.3.7</td>
<td>5.3.8</td>
<td>5.3.9</td>
<td>0.27</td>
</tr>
<tr>
<td>Overall</td>
<td>279</td>
<td>77</td>
<td></td>
</tr>
</tbody>
</table>

5.4 Results

Task 1. In the first task, users were asked to find a relevant talk by exploring bookmarks of users in their neighborhood. Results are presented in Fig. 10 and Fig. 11. The set of items of one specific user with whom the active user is related was explored 75 times by all participants. Twenty-three of these sets were used to bookmark a new item. The number of items in these 23 sets is 276. Thus, the effectiveness of 23/75 = 31% (first top bars in Fig. 10) and the yield is 23/276 = 8% (first bar in Fig. 11).

Fifteen users explored intersections of two related users focusing on talks that they have not yet bookmarked (as illustrated in Fig. 9, left). This kind of set was used to bookmark a talk four times (effectiveness = 27%) and the sum of items in the used sets was six (yield = 66%).

Talks in the intersection of three or four other users were explored twelve and six times and used five and three times, respectively (effectiveness of 42% and 50%). The number of items in the selection set was fourteen and seven, respectively (yield of 37% and 43%). As we can see from this data, the general trend is clear: the sets that allow users to explore the overlap of several prospects are both more effective and have higher yield. Moreover, there is a general tendency for effectiveness and yield to increase when more entities are used in the selection process. Small fluctuations within the general trend can be explained by the small sample.

Task 2. In the second task, users were asked to find a relevant talk by exploring the output of recommender agents (a content-based and a tag-based agent). Results are presented in Fig. 10 and 11, the middle set with six possible actions.

One out of nine users found a relevant talk by exploring suggestions by the content-based agent that were not related to any other entities on the screen (effectiveness=11%, yield=11%). Five out of 15 users found a relevant talk by exploring suggestions of the tag-based agent. Three out of nine users found relevant items by exploring the intersection of agents (i.e. talks that were suggested to them by both the content-based and the tag-based agent). Four out of eight users found relevant items by exploring the intersection of the agents with another entity. Fig. 9 (right) presents such an example.

The set of items of the content-based agent in combination with another user was explored twice, but not used to find a relevant item. The tag-based agent in combination with one or more entities was explored 10 times and was successful in 50% of the cases. The results presented in Fig. 11 indicate the same trend: a higher number of entities (prospects) involved in the intersection increases the effectiveness and the yield of the resulting set.
Fig. 10. Summary of actions explored, used to bookmark a paper, and effectiveness of those actions.

**Task 3.** In the third task, we asked users to find interesting talks by exploring tags that were added by users to talks. Results are presented in the bottom set of three actions in Fig. 10 and Fig. 11.

As the data shows, using a single tag prospect (i.e., exploring items related to one selected tag) results in the lowest effectiveness registered in the study – as only three users were able to find a relevant item (effectiveness 6%). The sum of the number of items in the set when a selection was made was 13 (yield 3/13=23%).

Combining a tag prospect with one or more additional entities was more effective. Nineteen users explored the combination of a tag with another entity and eight users used this intersection to bookmark an item (effectiveness=42%, yield=40%). A tag in relation to two other entities was effective in 57% of the explored cases. Some users indicated that they particularly liked this functionality – as this allows them to retrieve specifically items of their topic of interest from users they know or who have a high reputation in the field.
Fig. 11. Summary of yield results

5.5 Interpretation

Effectiveness and yield results are summarized in Fig. 12. Overall, these results indicate that effectiveness of an explored set increases once more entities are integrated. Similar trends are observed when we look at yield. These results indicate that the probability of selecting an item from a set increases if more entities are overlapped.

<table>
<thead>
<tr>
<th>5.5.1</th>
<th>Sign. effectiveness</th>
<th>Sign. yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.5.2</td>
<td>multiple versus one entity</td>
<td>5.5.3</td>
</tr>
<tr>
<td>5.5.4</td>
<td>user versus (user + tag</td>
<td>5.5.5</td>
</tr>
</tbody>
</table>
Fig. 12. Summary of effectiveness and yield results: (a) average effectiveness and yield results of single users, agents, tags and their combinations and (b) average effectiveness and yield results of exploration of a single entity and multiple entities. Effectiveness and yield of exploring multiple entities are significantly higher than effectiveness and yield of exploring a single entity (c).

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.5.10</td>
<td>5.5.11</td>
<td>5.5.12</td>
</tr>
<tr>
<td></td>
<td>Sign.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>effectiveness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.5.13</td>
<td>5.5.14</td>
</tr>
<tr>
<td></td>
<td>0.9717</td>
<td>0.1793</td>
</tr>
<tr>
<td></td>
<td>5.5.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0304</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.5.16</td>
<td>5.5.17</td>
</tr>
<tr>
<td></td>
<td>0.0304</td>
<td>0.1793</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 13. Comparison ranked list (baseline) versus visual exploration in TalkExplorer: (a) average effectiveness of a ranked list versus visual exploration of agents, agents in combination with other entities and multiple entities in general in TalkExplorer; (b) significance of difference.

In Fig. 12.b, we can see that the effectiveness and yield increases from 22% and 9% when a single entity is used to 57% and 50% when four entities are used, respectively. In Fig. 12.a, we can see that yield increases from 8% when the set of items of one user is used, 10% when the set of items of one recommender
agent is used and 23% when the set of items of one tag is used to 50% when the intersection of a user and another entity used, 75% when the intersection of an agent with another entity is used and 54% when a tag is used in combination with another entity.

For pairwise comparisons, we applied a 2-sample test for equality of proportions with continuity correction using R. As presented in Fig. 12.c, there is a significant difference in effectiveness when a tag is used in combination with another entity ($\chi^2(1) = 15.4011, df = 1, p < .001$). Effectiveness differences between users and agents in combinations with other entities are not significant. In general, effectiveness and probability of item selection is significantly higher when multiple entities are used ($\chi^2(1) = 8.5429, df = 1, p < .003$). These results indicate that enabling end users to explore interrelationships between prospects (sets of items in the overlap of entities) increases the probability of finding a relevant item. We discuss these results in Section 5.8.

Finally, to place the TalkExplorer performance in the context of traditional recommender systems, we measured whether the visual exploration offered by TalkExplorer improves acceptance of recommendations in comparison to a traditional ranked list of items – as defined in section 5.3. In this comparison, the acceptance rate generated by the ranked list interface at the same conferences (Table 2) is used as a baseline. As presented in Table 2 and Fig. 13.a, users of the HT 2012 and UMAP 2012 explored recommendation pages of CN3 279 times, bookmarking 77 items as a result of this exploration. The average success rate of the ranked list representation is 0.28 in this user study and is similar to the effectiveness of a single agent in TalkExplorer (effectiveness of 0.25 for a single agent – see Fig. 13.a). Effectiveness of visual exploration of an agent in combination with another entity increases to 0.41, but this difference is not statistically significant in this user study. As presented in Fig. 13b, exploration of multiple entities in general, with or without involving a recommendation agent, increases to 0.40 and is statistically significant ($\chi^2(1) = 4.65, df = 1, p = .03$).

5.6 Questionnaire results

To collect additional feedback, we used a questionnaire to inquire about needs and issues that users have when attending a conference and the extent to which our visualization addresses those needs and issues. This questionnaire was used to collect some preliminary additional feedback and assembled only a few questions we considered important at this stage.

Fig. 14.a presents the results of our questionnaire on a five point Likert scale. The first column of bar charts presents answers to questions that inquire about the importance of issues and needs at a conference. The second column presents to which extent TalkExplorer addresses these needs and issues. These results indicate that conference attendees perceive ‘finding relevant talks at a conference’ as an important need (median 5) and that TalkExplorer addresses this need in a good way (median 4). ‘Being aware which talks my friends are attending’ is also perceived as important (first column, median 4) and most participants indicate that this need is addressed by TalkExplorer (second column, median 4). ‘Being aware which talks people I follow are attending’ is considered less important and is also not so well addressed by our tool (median 3). ‘Finding other users who have similar interests’ is important for many participants (median 4) and is addressed well by TalkExplorer (median 4). ‘Gaining insight into why talks are recommended’ is important for many, but not all, participants. This need is addressed well by TalkExplorer according to most, but again not all, participants (median 4).
In addition, we inquired about perceived usefulness of the visualization to support transparency and exploration. These results are presented in Fig. 14.b and indicate that participants perceive the visualization as useful because it gives more insight than a plain list of talks (median 4). In addition, most participants liked our idea of adding agents in parallel to real users as a means to find interesting talks (median 4) – among others to compare relevancy of recommendations (median 5).

5.7 Think-aloud data analysis

An analysis of think-aloud data revealed some usability issues with the visualization and some additional insights into the usefulness of combining multiple relevance prospects. This data indicates that it is not clear why 15 users are shown in the neighborhood of a user. Participants asked why some users are shown by default and how these users are selected. Explaining the rationale of the selection of these users, such as the similarity match based on common bookmarks, can potentially resolve this issue. In addition, it would be interesting to explore how users can navigate through interest data of all users – while still distinguishing potentially more relevant bookmarks from users with common interests. People recommendation as opposed to recommendation of papers and talks was another suggestion from a participant.

Moreover, when users removed themselves from the visualization and explored bookmarks of other users only, it was not clear which of these items were already in their schedule. We resolved this issue for the UMAP 2012 evaluation by indicating in the right side panel whether a talk is already bookmarked (see Fig. 3). Using different colors in the canvas could be a second solution.

Finally, participants remarked that they particularly liked combining tag prospects with agents or user prospects – as this functionality enables them to identify talks of a specific topic of interest from a set of recommended items or items from a user. These remarks illustrate the perceived usefulness of combining multiple prospects and are also reflected in the data presented in Fig. 12. These data indicate that effectiveness of individual tags is significantly lower than a tag in combination with another entity.

5.8 Summary and Discussion

Results of our controlled user study indicate that effectiveness of an explored set increases when multiple entities are interrelated. As summarized in Fig. 12.b, effectiveness increases from 22% when a single entity is used to 57% when 4 entities are used. The difference between exploring a single versus multiple entities is statistically significant. A similar effect can be observed when we look at yield. In Fig. 12.a, we can see that yield increases from 9% when the set of items of one entity is used to 50% when the intersection of multiple entities is used. This yield is significantly higher when multiple entities are used (first row Fig. 12.c). These
results indicate that enabling end users to explore interrelationships between prospects (sets of items in the overlap of entities) increases the probability of finding a relevant item.

The results also allow us to make several interesting observations. First, it is interesting to note that the least effective kind of set is the set of items related to exactly one tag (6% effectiveness). Incidentally, this is the only option to use tags for item exploration offered in many tag-based systems. As shown by our data, the systems that do not allow exploring items related to combinations of several tags are not doing a good service to their users.

In contrast, exploring a prospect of a single related user in relation to the active user is a relatively effective approach – almost 1/3 of explored combinations produced bookmarks. It shows that the prospects of human users are more valuable (and trustable) than prospects offered by tags.

Overall, these results illustrate the added value of enabling users to combine social and personalization features of recommenders as a basis to increase the effectiveness of suggested items.

Subjective results that are presented in Section 5.6 indicate that the ability to combine visually multiple relevance prospects (i.e., bookmarks of users, suggestions of recommender agents and talks marked by specific tags) is perceived as useful. Results from our questionnaire are generally positive and indicate that participants value our visualization as a way to explore multiple relevance dimensions. In addition, they indicate that such a visualization gives more insight than a typical ranked list of recommendations.

While these results illustrate the usefulness of visualizing and combining recommendations, tags and users, there are several limitations to this study that should be articulated and addressed in follow up studies. First, we asked users to explicitly explore users in their neighborhood, recommender agents and tags in three separate tasks. While results of these tasks give some interesting insights in the usefulness of these entities and the way users interacted with additional entities during these tasks, we cannot draw strong conclusions about the relative effectiveness of tags, users and agents in this way. The order of the tasks may have had an influence on the effectiveness of these entities. Moreover, we explicitly asked to explore these entities. Second, participants at both UMAP 2012 and HT 2012 are knowledgeable on topics of recommendation and information access in general and so they might not be representative of the general conference audience. In our second study presented below we attempted to address some of these limitations.

6. FIELD STUDY
The second user study was conducted at the European Conference on Technology Enhanced Learning (EC-TEL 2012) in September 2012 and the ACM Learning Analytics and Knowledge 2013 conference in April 2013 (LAK 2013). The user study was performed with attendees of these conferences using real conference data - i.e., using actual talks and bookmarks, tags and ratings of the conference participants.

As explained above, we wanted to address some of the limitations of the first user study in this field user study (three separate tasks in a fixed order). While the use of real conference data and real conference participants who were truly interested to find relevant talks made this study reasonably realistic, its controlled nature made it hard to make assessments about the use and the value of TalkExplorer in natural user-driven settings. To address this problem, users in the field study were free to explore the visualization in any way they like, while we observed their work, measured the number of times that they explored users, agents and tags and how often these entities and their combinations were successful to find a relevant talk. This experimental setup allows exploring “natural” interaction patterns that can yield interesting insight into workflows of users and different strategies that these users employ to find items. We use the analysis method proposed by Dou et al. [2012] to understand the reasoning of participants when using TalkExplorer to find relevant items. The overall objective is to understand how a user performs a successful discovery, as a means to understand how the interface can be used in a successful way and to adapt the visualization for future use.
6.1 Participants
In total, 18 users participated in the experiment (ten at LAK 2013 and eight at ECTEL 2012). We inquired about the number of conferences participants have attended, as well as their knowledge of recommendation and visualization techniques, respectively. The latter were rated on a five point Likert scale. On average, participants attended 2.5 conferences in the past (std. deviation 1.15). Most of the participants have again knowledge about or expertise with visualization techniques (average 4.23, std. deviation 0.79). In contrast with participants of the first study (see Section 5.1), familiarity with recommendation techniques was less high (average 3.15, std. dev. 1.23). In general the audience of both the LAK 2013 and ECTEL 2012 conferences is less technical than the audience of UMAP 2012 and HT 2012, as many participants have a background in educational sciences – although with a focus on the use of technology in this field.

6.2 Experimental Setup and Data Collection
The task of this second study was simple and straightforward: explore the visualization and try to find papers that are of interest to you. No requirements were imposed on the type of entities that should be explored or their order. As with the first study, to address cold-start issues, participants were first asked to bookmark and tag at least five papers of interest. This data was used to generate recommendations and to select users to show in the visualization. In contrast to the first study, users did not receive an explanation of the visualization. This design choice was made to make user interaction with TalkExplorer unprompted and more realistic. The user had a chance to learn about the system in a “natural” way since the URL of the visualization was distributed using Twitter during the conferences. For the same reason, no think-aloud was performed.

Data was collected in two ways. Every interaction with the visualization was captured, including adding and removing users, agents, tags to the visualization, exploring items bookmarked by a users, exploring items recommended by the agents, exploring tags and exploring items in intersections of entities. In addition, we captured bookmarks of items and the entities that were used to find these items. Afterwards, participants filled out the same survey that was used in the first user study.

6.3 Data Encoding Schemes and Measures
We adopted three encoding schemes introduced in Brown et al. [2013] as a basis to identify strategies of different participants and their relative effectiveness and efficiency. The encodings are:
1. state-based, which captures which entities were displayed and explored in each interaction step. This encoding is used in the first user study as well to analyze which combination of entities was most effective to find relevant items.
2. event-based, which captures all user interactions with TalkExplorer. These events record the different actions of participants, such as adding or removing an entity, and their timestamps.
3. sequence-based, which encodes sequences of events. Such a sequence can for instance indicate that a participant first added an agent, then added a user, and then explored the intersection of the agent and the user.

These states, events and sequences are used to analyze interaction patterns of participants. As in work of Dou et al. [2012], we make a distinction between strategies of participants and methods that they employed to implement the strategy. More specifically, the following components are adopted:
• strategy is used to describe the means that the user employed to try to find a relevant item.
• methods are the steps that the user adopted to implement the strategy. A method is composed of two components: an operation (add, remove, explore) and a state (user, tag, agent or their combinations) to which this operation is applied.

For example, a strategy of a user can be to first explore items suggested by an agent and then to interrelate this set with bookmarks of a specific user as a means to filter out the potentially more relevant
items. This strategy is denoted as $agents \rightarrow users$ in this section. In order to implement this strategy, the user can employ the following methods:

- **add agent:** this method will add items suggested by an agent to the visualization.
- **add user:** this method will add bookmarks of a specific user to the visualization.
- **explore agent + user:** this method enables the user to inspect items in the intersection of the agent and the user, i.e. items that are recommended by the agent and also bookmarked by the user.

To assess the value of interactive multi-prospect visualization offered by TalkExplorer, we have analyzed again the way in which users explore and use the different combinations of entities (states) – as explained in Section 5.3. We have measured the *effectiveness* and *yield* of different states to gain insight in the relative success rate of different combinations of entities to find relevant items. Similar to the first user study, a comparison to baseline data is included, as introduced in Section 5.3. Table 3 presents the baseline data for this field study. Acceptance rates of recommendations are lower than in the controlled user study (Table 2). A possible reason is that participants of this field study are less familiar with recommendation techniques, as indicated in Section 6.1.

As participants were able to freely explore the visualization in this user study, we conducted two additional analyses that give insight into workflows and strategies of different users. In addition to effectiveness and yield of different states, we measured and analyzed the performance of participants and the effectiveness and efficiency of different strategies that they employed. The event-based encoding scheme is used to gain insight into the performance of participants. The sequence-based encoding scheme is used to gain insight into the effectiveness and efficiency of different strategies.

![Table 3. Baseline data for the field study](image)

<table>
<thead>
<tr>
<th>6.3.1</th>
<th>6.3.2 Number of successful recommendations</th>
<th>Acceptance rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3.3 EC-TEL 2012</td>
<td>6.3.4 352</td>
<td>6.3.5 24</td>
</tr>
<tr>
<td>6.3.6 LAK 2013</td>
<td>6.3.7 178</td>
<td>6.3.8 29</td>
</tr>
<tr>
<td>6.3.9 Overall</td>
<td>6.3.10 530</td>
<td>6.3.11 53</td>
</tr>
</tbody>
</table>

### 6.4 State-based analysis

Our first analysis corresponds to the analysis of the controlled user study. We captured and analyzed the different entities and their combinations (states) that users explored and used, and measured their effectiveness and yield as defined in Section 5.3. Fig. 15 shows the number of times that users explored and used different entities and their combinations. A user with common interests was explored 63 times and used to bookmark an item 11 times. The number of items in these 11 sets is 100 (see Fig. 15). Thus, the effectiveness is $11/63 \approx 17\%$ and the yield is $11/100 = 11\%$ (first bar Fig. 16). The intersection of two users was explored 19 times. The intersection of three users was explored seven times. These sets were used four and three times, respectively (effectiveness=21% and 43% respectively).

Agents were explored almost as frequently as users. The content-based agent was explored 13 times, the tag-based agent 70 times. The agents were used to bookmark a relevant paper two and nine times respectively, resulting in an effectiveness of $2/13=15\%$ for the content-based agent and $9/70=13\%$ for the tag-based agent. Intersections of agents with another entity were explored nine times in total and include intersection of both agents, intersection of an agent with a tag and intersection of an agent with a user. The intersection of both agents was as effective as in the first user study (33%) – although this intersection was explored less often (nine times in the first user study and three times in the second user study). The
intersection of an agent and a user was explored five times – out of which three explorations were successful. The number of items in these three sets was ten, resulting in an effectiveness of 60% and a yield of 33%. The intersection of an agent and a tag was only explored once.

Tags are also explored often (88 times). These tags were used eight times to find a relevant item (effectiveness=9%). This result is very similar to the result of the previous user study, where tags were also the least effective. As we can see from this data, the general trend is clear: tags, users and agents were explored equally often by users when they are free to explore the visualization. Effectiveness of a single entity (an agent, a tag or a user) is again the lowest: 17% for a single user, 15% and 13% for a single agent and 9% for a single tag. A summary of these results is presented in Fig. 17. Effectiveness increases when multiple entities are interrelated: to 34% when a user is interrelated with another entity. (Fig. 17.a), 60% when an agent is interrelated with another entity and 100% when a tag is interrelated with another entity – but this last number is based on a too small data sample. Average effectiveness is 13% when a single entity is used, 32% when two entities are used and 50% when three entities are used (see Fig. 17.b). As presented in Fig. 17.c, the difference in effectiveness of a single entity versus multiple entities is again statistically significant ($\chi^2(1) = 10.96$, $df = 1$, $p < .001$).

Whereas the general trend can be confirmed, the number of times that these intersections were explored is lower than in the first user study: items in the intersection of two entities were explored 28 times in the second user study (versus 53 times in the first user study) and items in the intersection of three entities were explored eight times (versus 37 in the first user study). Items in the intersection of four entities were not explored. The data of both user studies are summarized and compared in Fig. 18. Fig. 18.b shows the number of explorations in both studies. We discuss these results in Section 6.7.

Finally, we measured whether visual exploration improves acceptance of recommendations when they are presented in a traditional ranked list of items – as presented in Fig. 2. The number of times that users accept recommendations when they are presented in such a ranked list is used as a baseline. As presented in Fig. 19, the success rate of such a representation is 0.1 in this user study and is similar to the effectiveness of a single agent in our visual representation (effectiveness of 0.15 for the content based agent and 0.13 for the tag-based agent). Effectiveness of visual exploration of an agent in combination with another entity increases to 0.6 and is significantly more successful than exploration of a ranked list of recommendations ($\chi^2(1) = 20.3364$, $df = 1$, $p < .001$). Also exploration of multiple entities in general, with or without involving a recommendation agent, is significantly more effective to find a relevant talk ($\chi^2(1) = 10.5542$, $df = 1$, $p = .001159$).
Fig. 15. Summary of second user study: actions explored and used to bookmark a paper, and effectiveness of those actions (used/explored)

Fig. 16. Summary of yield results
Fig. 17. Summary of effectiveness and yield results: (a) average effectiveness and yield results of single users, agents, tags and their combinations and (b) average effectiveness and yield results of exploration of a single entity and multiple entities. Effectiveness of exploring multiple entities is significantly higher than effectiveness of exploring a single entity (c).

Fig. 18. Comparative results of study 1 (controlled user study) and study 2 (field study): summarizing average effectiveness (a) and the total number of explorations (b)
Fig. 19. Comparison ranked list (baseline) versus visual exploration in TalkExplorer: (a) average effectiveness of a ranked list versus visual exploration of agents, agents in combination with other entities and multiple entities in general in TalkExplorer; (b) significance of difference.

6.5 Event-based analysis

The event-based encoding scheme captures clicks of participants and their timestamps. In a second analysis, we compared the events of participants who were successful in finding relevant items with those of participants who were not successful. The two groups of participants are labeled achievers and non-achievers in this section. We examined the number of entities that were examined by both user groups and the number of different methods that participants employed. These methods include add, remove and explore methods operating on different states – i.e. users, tags and agents and their combinations. We also examined the time that participants spent working with the visualization.

Fig. 20. Achievers and non-achievers explored a similar number of entities (mean 15 and 16.4, respectively), but achievers used significantly more methods and states to examine these entities from different perspectives, e.g. by combining users with agents or users with tags.
The analysis indicates that both achievers and non-achievers explored a similar number of entities, as indicated in Fig. 20. Non-achievers explored on average 15 entities. Achievers explored on average 16.4 entities. Achievers did use significantly more methods to inspect these entities from different perspectives, including add, remove and explore operations on different entity types (users, agents, tags) and their combinations. On average, achievers used nine different methods, whereas non-achievers only tried five different methods. If we consider specifically the different states on which these methods operate (for instance users, users+agents, users+tags), achievers used on average 4.5 different states, whereas non-achievers used only 2.2 different states. The data are normally distributed. A t-test indicates that the achievers used significantly more methods ($t = 2.7135$, $df = 16$, $p = .015$) operating on different states ($t = 2.94$, $df = 16$, $p = .0097$).

We further analyzed time spent by all participants. Most participants needed the most time to bookmark the first item, between 4:00 and 15:00 minutes. Participants needed between 2:30 minutes and 00:21 minutes for the second bookmark. On average, participants needed 06:00 minutes for the first bookmark and 01:16 minutes for bookmarks 2, 3, 4 and 5. A pairwise t-test indicates that this difference is significant ($t = 7.0261$, $df = 7$, $p < 0.001$) and suggests that users become more proficient when they gain experience. These results suggest a rather steep learning curve.

### 6.6 Sequence-based analysis

In a third analysis, we analyzed sequences of events to identify the different strategies of participants. We then compared the effectiveness and efficiency of these strategies:

- **effectiveness** measures the success rates of different strategies and is counted as the number of times the strategy was used successfully (i.e. the participant found a relevant item) divided by the total number of times that the strategy was used.
- **efficiency** measures how fast the strategy resulted in successful bookmarks and is counted as the total time spent by participants using this strategy divided by the number of bookmarks they made.

Some participants used the visualization more than once and used different strategies in different sessions. In total, 27 sessions were recorded and analyzed. Fig. 21 summarizes different strategies that were used in these sessions. Fig 21.a lists the different strategies and their relative effectiveness. Five strategies start with an explore user method (Fig 21.a, bars 1-5). Strategy 2 (users → tags) and strategy 3 (users → agents) explore tags and agents respectively in a second step. Strategy 4 and strategy 5 explore three different types of entities: users → tags → agents and users → agents → tags, respectively. Strategy 6 starts with adding agents to the visualization and uses additional methods to explore and add users (agents → users). Strategies 7 and 8 start with adding tags and both use additional methods to explore and add both users and agents.

As presented in Fig. 21.a, four participants explored only users and were not able to find relevant items (first bar Fig. 21.a). Two participants explored first users and then added tags to the visualization (users → tags, second bar Fig. 21.a). None of these sessions were successful. One participant added an agent in a next step (users → tags → agent, fourth bar Fig. 21.a), but also this strategy was not successful. Nine participants explored first the users and then added agents to the visualization (users → agents, third bar Fig. 21.a). Only two out of these nine participants were able to find relevant talks. Six participants added tags in a next step (users → agents → tags, fifth bar Fig. 21.a). This strategy was successful in 50% of the sessions. The remaining five participants started with methods operating on tags or agents as opposed to starting with the default exploration of users. Four out of these five participants were able to find relevant items.
Fig. 21. Effectiveness of strategies: (a) overview of all strategies and their relative effectiveness, (b) effectiveness of strategies based on the number of different entities involved (users, agents or tags).

Fig. 21.b lists the effectiveness of different strategies based on the number of different entities that they use. A strategy operating only on users is a strategy in the first category. A strategy operating on users and tags is a strategy in the second category. A strategy that uses user, agent, and tag is a strategy in the third category. As presented in Fig. 21.b, strategies relying on one entity type were not successful. Fourteen sessions record a two-entity strategy (users → tags, users → agents or agents → users). In four sessions, subjects were able to find relevant items, resulting in an overall effectiveness of 22%. Nine sessions record a strategy with three entities. In five of these sessions, the strategy was successful. Thus strategies that combine three entity types could be the most successful strategies.

Fig. 22 presents the efficiency of successful strategies. Efficiency of strategies was measured as the total time spent by participants in a session using this strategy divided by the number of bookmarks that were made by these participants in these sessions. As illustrated in Fig. 22.b, the average time to bookmark an item with strategy that employs methods with three different types of entities is 2:45 minutes. The average time to bookmark an item with a strategy that employs two types of entities only is 5:28 minutes.

Fig. 22. Efficiency of successful strategies: (a) average time to bookmark an item with the different strategies, and (b) average time to bookmark items with strategies operating on two and three different entity types.
Although based on a small dataset, three preliminary findings stand out from this analysis of interaction patterns:

1. Participants who explored only bookmarks of users and no other entities (agents, tags) were not successful in finding relevant talks.
2. Strategies starting with adding and exploring tags or agents have a high success rate. This result hints that the default visualization showing three users as presented in Fig 6.b may not be the best way to introduce users to our approach.
3. Effectiveness of strategies that explored all types of entities is 55% compared to 22% for strategies that employ methods operating on two types of entities. On average, participants needed 2:45 minutes to bookmark items with a strategies operating on three entity types, compared to 5:13 minutes with strategies operating on two entities. This result may suggest that the combination of the different entity types is the key functionality that should be supported in future work.

6.7 Questionnaire results

To collect additional feedback, we again used a questionnaire to inquire about needs and issues that users have when attending a conference and the extent to which our visualization addresses those needs and issues. The same questionnaire was used than in the first study.

Fig. 23.a presents the results of this questionnaire on a five point Likert scale. The first column of bar charts presents answers to questions that inquire about the importance of issues and needs at a conference. The second column presents to which extent TalkExplorer addresses these needs and issues. These results indicate that conference attendees of the second user study perceive 'finding relevant talks at a conference' as an equally important need (median 5) than in the first study and that TalkExplorer addresses this need in a good way (median 4). Whereas this result is the same as in the first study, the other results are slightly different. ‘Being aware which talks my friends are attending’ is perceived as important (first column). Most participants indicate that this need is addressed by TalkExplorer (median 5). This result is higher than in the first study (median 4). ‘Being aware which talks people I follow are attending’ is considered less important (median 3), but most users indicate that this need is addressed by TalkExplorer (median 5). This result is again higher than in the first user study. ‘Gaining insight into why talks are recommended’ is perceived as important (first column). Most participants indicate that this need is addressed by TalkExplorer according to participants (median 5). This result is higher than in the first study (median 3). ‘Finding other users who have similar interests’ is equally important for many participants (median 4) and is addressed well by TalkExplorer (median 5). This result is again higher than in the first user study. ‘Gaining insight into why talks are recommended’ is important for many, but not all, participants. This need is not well addressed TalkExplorer according to participants (median 3). This result is lower than in the first user study. To summarize, results inquiring about addressing issues and needs to gain insight into which talks other users are attending and finding other users with similar interests are higher in this second study, whereas gaining insight into the rationale of recommenders is lower. We discuss in the next section how these results might be used to explain differences in the way users interacted with the visualization in both user studies.

In addition, we inquired about perceived usefulness of the visualization to support transparency and exploration. These results are presented in Fig. 23.b and indicate that participants perceive the visualization as useful because it gives more insight than a plain list of talks (median 4). In addition, most participants liked our idea of adding agents in parallel to real users as a means to find interesting talks (median 4). However, the perceived usefulness to visualize the output of multiple recommenders as a basis to compare their relevancy is much lower than in the first user study (median 3 compared to median 5 in the first study). Compared to the first user study, users did not explore such intersections often. The intersection of two agents was explored more than 17 times in the first study. In the second study, this intersection was explored only 3 times.
In addition to these questions, we have collected additional feedback in two open questions. The first question inquired in which situations users would use TalkExplorer. Responses to these questions again point to interests related to other users, and include: “to know who I can meet at the talks I have bookmarked”, “to know where other people that I know are going”, “I found this tool really useful to find papers about my fields of interest. It may help you to discover people that work on similar fields” and “being aware of who is participating in talks and the conference”. Users are also interested to find out which other users are interested in their own work as indicated by “Browsing to see who else might be like me”.

Users also point to the usefulness of tags: “easier to find papers in visualization while exploring the tags”. In general, users indicate that the visualization is useful to find relevant talks, including talks that they would not consider – i.e. more diverse content that their main research topics. Several users also point to the usefulness for newcomers to a conference community. One user indicates she would like to use the tool as a retrospective tool after the conference: “I would find the tool more useful as a retrospective tool after a conference has taken part. It would serve as a kind of a visualization of notes and contacts.”

Further remarks point to usability issues that need to be resolved in future deployments: the checkboxes in the left side pane (see Fig. 3) to add entities to the visualization, such as a particular user, agent or tag, were difficult to use for some users. They indicate that these checkboxes should be made more prominent. Users indicate that they would like to know how similar they are to other users. Adding a similarity indicator would be a useful extension. Users also would like to see more details of talks, including the abstract.

### 6.8 Summary and Discussion

There are similarities and differences between the design and the results of the first and second user study. As presented in the previous section, the first study provided several evidences that multiple relevance prospects are useful. However, there were also several limitations to this study. Most importantly, we asked participants to explicitly explore users in their neighborhood, recommender agents and tags in three separate tasks. While results of these tasks give some interesting insights in the usefulness of these entities and the way users interacted with additional entities during these tasks, we cannot draw strong conclusions about the relative usefulness of tags, users and agents in this way.

In the second study, we addressed this limitation by asking users to freely explore our interactive visualization. The analysis of interaction patterns yields less biased data with respect to relative usefulness of tags, users and agents. Despite a considerable change of the study design, the main outcomes of the second study confirm the key findings of the first study. The first outcome of the second study confirmed that
users perceive the use of multiple relevance prospects instead of one as useful: users, agents and tags were explored 63, 83 and 80 times respectively. Data from the questionnaire confirms that users are interested to explore these entities to find relevant talks at a conference. These data indicate that users perceive the interface as useful because it gives more insight than a plain list of talks (median 4). In addition, most participants liked our idea of adding agents in parallel to real users as a means to find interesting talks (median 4). These results are the same as of outcomes of the first controlled user study. As users were free to explore the visualization and not explicitly asked to explore these entities, these data give a better indication about the usefulness of these entities compared to the first study.

The second outcome that is confirmed by the second study is that intersections between multiple entities increase effectiveness: there is again a significant difference in effectiveness when users explored talks in intersections of multiple entities as opposed to talks of a single entity. A set of items of a single entity was effective in 13% of explorations, as summarized in Fig. 17.b. Effectiveness increases to 50% when three entities are used. These results confirm that enabling end users to explore multiple dimensions of relevance increases the chance of finding a relevant item in personalized social systems.

The field study gives additional insight into strategies of different users. Results of the sequence-based analysis indicate that there are eight different strategies that were employed by participants. The first five strategies (bars 1-5, Fig. 21.a) start with exploring users and then add agents and/or tags. The last three strategies start with a different entity type: participants that used these strategies added first an agent or a tag, and then explored these entities and their combinations with other entity types. The effectiveness of these strategies is high, as indicated in Fig. 21.a Strategies that combine methods operating on all different entity types (users, agents, tags) were effective overall in 55% of the sessions.

Results of the event-based analysis indicate that both achievers (participants who were able to find relevant items) and non-achievers (participants not able to find relevant items) explored a similar number of entities (15 for non-achievers, 16.4 for achievers). An interesting outcome of this analysis is that achievers used significantly more methods operating on different states to inspect entities from more perspectives than non-achievers, for instance by combining users, agents and tags and inspecting overlaps between two or more of these entities. These results provide evidence that combining multiple relevance prospects is key to find relevant items.

At the same time, we found an interesting difference between the volume of user activity in the first controlled user study and the second field study. Despite their confirmed value, intersections of prospects were not explored as frequently in the field study as they were in the controlled study. Intersections between two entities were explored 28 times, intersections between three entities only 8 times. The visualization of intersections of four entities was not explored at all. The data of both user studies are summarized and compared in Fig. 18. Fig. 18.b shows the number of explorations in both studies.

The reduced use of intersections didn’t allow the users to fully benefit from the power of this feature. This might also explain why users were less prone to agree that it was useful to visualize the output of multiple recommenders as a basis to compare their relevancy (Fig. 23.b, question 3, median 3). In the first study, users explored these combinations more often and were more positive about the usefulness of this concept. This difference can be explained by two reasons:

- First, the subjects in the first study were specifically requested to explore different relevance prospects. As a result, they were literally pushed to explore more options for each relevance prospect including combinations that they might not discover on their own and had more chances to experience and appreciate the value of this feature. The “push” vs. “no-push” issue was further magnified by a relatively complex visualization approach implemented in TalkExplorer where set overlaps were shown in a more advanced but less intuitive way in comparison with more common Venn diagrams. The complexity of the interface was likely to less affect the users of the first study who had a dedicated task to explore each of the relevance prospects than the users of the second study who were left free to explore the interface.
Second, the audience of the second study is less technical. The study was conducted at two conferences in the Technology Enhanced Learning domain. It is likely that participants in this community are less familiar with recommendation and visualization technologies than participants of the first. While the value of prospect intersection holds for this audience as well, the lower familiarity with the underlying technology could further reduce their motivation to explore complex intersections. The complexity of the TalkExplorer interface could further increase the difference between more and less technical users.

7. GENERAL DISCUSSION AND FUTURE WORK

Results of both studies illustrate the usefulness of visualizing multiple prospects. Users are interested to explore users, agents and tags and indicate that these multiple prospects are useful as a basis to find relevant talks. Exploring intersections increases effectiveness, although a less motivated and technical audience may require more guidance to understand the rationale of such intersections. In general, these users were more interested to find out which talks other users are attending and to find interesting talks than to gain insight into recommendations.

Before we discuss limitations of our studies, we summarize results. There are two outcomes that can be confirmed by both studies. A first outcome confirmed by an analysis of interaction patterns indicates that the use of multiple prospects increases effectiveness of finding relevant items. This effectiveness increases from 22% in study 1 and 13% in study 2 when a single entity is used to 57% when four entities are used in study 1 and 50% when three entities are used in study 2. The difference is statistically significant. The results are summarized in Fig. 18. A second outcome is that multiple prospects are perceived as useful: users indicate in both studies that the approach gives more insight than a plain list of recommendations (median 4) and that providing agents in parallel to real users is helpful. In the first study, we explicitly asked users to explore all entity types (tags, users, agents). In the second study, users were free to explore the visualization. The fact that users still explored all entity types as frequently is another indication that users are interested in multiple prospects of relevance: users, agents and tags were explored 63, 83 and 80 times respectively in the field study. A comparison with baseline data also indicates that acceptance of recommendations is higher when users can examine which users have bookmarked items that are recommended to them. Results of the field study also indicate that acceptance of recommendations in TalkExplorer is significantly higher than a plain list of recommendations.

7.1 Limitations of user studies

Although our user studies delivered interesting results, there are some limitations that should be articulated. The first user study was conducted in a controlled format and can be biased due to specific exploration tasks assigned in a fixed order. The second user study addresses this limitation, but at the same time the data in this study with respect to explorations of intersections is limited. Thus, although the general trend can be confirmed, the data may be too limited to draw strong conclusions.

A second limitation of our user studies is that only two agents were shown in the visualization: a content-based recommender system and a tag-based recommender system were both embodied as agents and shown in parallel. In a follow up study, we will explore whether the use of more agents has an effect on the perceived usefulness of interrelating these agents. We will add more recommendation techniques that are already implemented in CN3 and embody them all as agents.

A third limitation is that the number of items represented is not that high: we deployed our visualization on top of relatively small, 1 – 3 stream conferences where a total number of talks in the visualization panel was not exceeding 59 – 129 (see Table 1). The added value of such a visualization may be higher when the number of items that needs to be filtered is higher: if a user needs to select from a list of 50 talks, she may not need advanced technology to help her find relevant items. When the list is much longer, for instance between 500 – 1000 items, the need to visualize this data to find items may be higher. This may also explain
why users did not explore intersections of three entities and four entities a lot. If for example only 10 items are represented in an intersection of two entities, there is no need to add a third entity to filter out the more relevant items from this set: the user can just browse this set of 10 items. In a follow up study, we will deploy our visualization on top of data from a larger conference with more tracks. In addition, we are exploring the use of this visualization in a different domain such as music information retrieval.

7.2 Scalability Issues

A cluster map visualization was chosen in this study to evaluate whether combinations of multiple relevance prospects are useful. Aduna [Aduna 2014] is a tool that we picked for its ability to integrate many prospects, but future research is needed to find the most efficient design. The particular visualization has important limitations: the design seems complex for a non-technical audience. Whereas multiple entities result in the best display to find relevant items, these intersections were not explored often by non-technical users. In the first user study, we inquired about the maximum number of entities on the screen that users considered manageable: results indicate that users consider four entities (st. dev. 0.47) as the maximum of a readable design.

A first important line of research is to find better designs that enable overlapping multiple sets in a scalable and intuitive way. The most common set representations are Venn and Euler diagrams [Riche and Dwyer 2010]. We compared how users interact with a Venn diagram in comparison with the cluster map representation of TalkExplorer [Verbert et al. 2014]. Results indicate that users explore intersections more often when a Venn diagram is used. Nevertheless, a Venn diagram has scalability issues, in contrast to Aduna, as presenting an overlap of more than three prospects is a problem. Alternative representations that enable set comparisons have been presented in [Steinberger et al. 2011, Collins et al. 2009, Lex et al. 2014, Riche and Dwyer 2010, Alper et al. 2011]. Of particular interest is research on scalability and readability of these visualization techniques. Riche and Dwyer [2010] researched how readability of (complex) Euler diagrams can be improved: the authors present two algorithms to simplify the representation of set regions. The approach is promising to enable scaling to larger and more complex set combinations. Scalability of set visualizations is also addressed by work of Lex et al. [2014]. The authors present UpSet, a novel visualization technique that enables to scale to a large number of sets – while still enabling to explore individual elements in these sets. UpSet combines a set view for tasks related to set operations (intersections, unions, etc.) with an element view for element-related tasks, such as selecting elements of an intersection for a detailed exploration. In our ongoing research, we have deployed the UpSet visualization technique on top of data from CN3 and are evaluating whether this technique can be used by non-technical users to interact with recommendations and to improve their acceptance.

A second line of research to enable scaling to a larger number of entities is to pre-filter entities in a personalized way with intelligent filtering methods. Particularly when deploying the visualization on top of data of larger conferences, such filtering is needed to pre-filter the more relevant tags and people. We currently select 15 users that are similar to the active user. Further research is needed to evaluate the best strategies to filter these entities, or to aggregate them and provide filter and drill down functionalities to the details [Shneiderman 1996]. In addition, data of both studies show that it is not clear to the users why these users are shown. Work of Guy et al. [2009] suggests that familiarity is more effective than similarity to increase acceptance of recommendations. Assessing the effect of different pre-filtering or aggregation strategies, as well as explaining such strategies to users, is part of our future work.

8. CONCLUSIONS

In this article, we have presented an interactive visualization of social data and recommendations of the CN3 conference navigator. In addition, we have described and discussed the results of two user studies that have been conducted with this interface. The research contribution of this work is threefold:
1) First, we present a novel and synergetic approach to combine multiple relevance prospects that include personalized relevance as offered by different recommenders and social relevance as offered in social bookmarking systems. In this approach, recommender systems are presented as agents and their interrelationship can be explored - i.e. a user can explore which items are suggested by multiple recommender agents. In parallel, real users and their bookmarks are shown and users can explore both interrelationships between users as well as interrelationships between agents and users. Third, a user can explore tags and interrelationships between tags, users and agents.

2) Second, we present a visual recommendation interface TalkExplorer, which serves to both explain the provenance of recommendations in a transparent way and to support exploratory search. Users can browse bookmarks of other users, tags and suggestions of recommender agents as a basis to find relevant items.

3) Third, we present the results of two user studies that evaluated the usefulness and effectiveness of this interactive interface. In these studies we attempted to assess how the ability to control and combine entities can influence effectiveness of finding relevant items and usefulness. Evidence of our user studies demonstrates that combining multiple relevance prospects is productive; the success rate of finding a relevant item increases when multiple entities are explored in combination. Results of our field study show that particularly users who used more methods operating on combinations of entity types were successful in finding relevant items. These results illustrate that the approach is useful. Nevertheless, in an open setting such combinations are only explored to a limited extent.

Although interesting findings can be derived from our user studies, there are also important limitations that require further research. We have seen from our data that intersecting multiple entities increases effectiveness of finding relevant items, but the current visualization seems too complex to use without guidance. We will leverage this evidence, but research different representations that are more intuitive and that can scale to a large number of entities. A follow up study will also include multiple agents and assess the added value of our visualization on top of larger data collections.

ELECTRONIC APPENDIX
The electronic appendix for this article can be accessed in the ACM Digital Library.

ACKNOWLEDGMENTS
We thank all the participants of the user studies for their participation and useful feedback. Research of Katrien Verbert was supported by postdoctoral fellowship grant of the Research Foundation – Flanders (FWO). The authors would like to thank the anonymous reviewers for their valuable feedback and suggestions.

REFERENCES


ACM Transactions on xxxxxxx, Vol. xx, No. xx, Article xx, Publication date: Month YYYY


Rashmi Sinha and Kirsten Swearengen. 2002. The role of transparency in recommender systems. In CHI ‘02 Extended Abstracts on


Received ...; revised ...; accepted ...