









junior scholars, such as graduate students or research assistants. The participants came from 15 different countries; their ages ranged from 20 to 50 ( $M=37$ ,  $SE=7.07$ ). All of them could be considered as knowledgeable in the area of the intelligent interface for at least one academic publication from IUI 2017. To control for any prior experience with the recommender system, we included a question about in the background questionnaire. The average answer score was ( $M=3.28$ ,  $SE=1.13$ ) on a five-point scale.

### Experiment Design and Procedure

To assess the value of the proposed interface, we compared the dual interface with the scatter plot and the ranked list (SCATTER) with a baseline interface using only a ranked list (RANK) with Section A (in Figure 1) removed. The study used a within-subjects design. All participants were asked to use each interface for three following tasks and to fill out a post-stage questionnaire at the end of their work with each interface. At the end of the study, participants were asked to compare interfaces regarding their explicit preference. The order of using interfaces was randomized to control for the effect of ordering. In other words, half of the participants started the study with the SCATTER interface. To minimize the learning effect (becoming familiar with data), we used data from two years of the same conference: the SCATTER interface used papers and attendees from IUI 2017, while the RANK interface used the corresponding data from IUI 2016.

Participants were given the same three tasks for each interface. The tasks were explicitly designed as diverse but realistic tasks that could be naturally pursued by attendees at research conferences. **Task 1:** Your Ph.D. adviser has asked you to find four Committee Member candidates for your dissertation defense. You need to find candidates with expertise close to your research field while trying to lower their travel cost to your defense. **Task 2:** Your adviser has asked you to meet four attending scholars, preferably from different regions across the world, who have a close connection to your research group. **Task 3:** You want to find four junior scholars (not yet faculty members) with reasonably similar interests among the conference attendees to establish networking.

The participants were asked to pick suitable candidates among conference attendees, based on their best judgment in each task. When designing the tasks, we attempted to make them realistic, yet focused on multiple aspects of relevance, as many real tasks are. We consider that task 1 is relevance-oriented and that tasks 2& 3 are diversity-oriented. For a relevance-oriented task, we expect to see if the proposed interface helps the user to coordinate different relevance aspects of the desired target efficiently. In contrast, for the diversity-oriented task, we expect the system to help to recognize the diversity of recommended items, as compared to the baseline interface.

### Action Analysis

Table 1 shows the system usage for two interfaces. The data indicate that participants extensively used both the control panel and explanation tabs to complete the tasks. The participants usually required more actions on the first task to familiarize themselves with the system. There is no significant

Task	Action	RANK		SCATTER		P
		M (SE)	User Count	M (SE)	User Count	
T 1	Control Panel	3.88 (2.40)	24	4.12 (2.02)	25	
	Explanation Tab	34.28 (29.50)	25	7.96 (7.48)	19	
	Click - Rank	26.28 (29.50)	25	4.92 (6.75)	15	*
	Click - Scatter	-	-	3.04 (5.45)	13	*
	Time Spent	345.44 (209.86)	25	389.12 (235.29)	25	
T 2	Control Panel	2.88 (1.64)	24	2.88 (1.12)	25	
	Explanation Tab	19.96 (17.47)	25	9.16 (6.28)	25	
	Click - Rank	16.96 (17.47)	25	2.68 (4.69)	15	*
	Click - Scatter	-	-	3.48 (5.41)	13	*
	Time Spent	216.6 (144.95)	25	190.84 (115.33)	25	
T 3	Control Panel	2.56 (1.04)	24	2.84 (1.10)	25	
	Explanation Tab	20.08 (20.29)	25	6.4 (7.22)	19	
	Click - Rank	19.08 (20.29)	25	3.48 (7.80)	9	*
	Click - Scatter	-	-	2.92 (2.95)	15	*
	Time Spent	345.95 (156.39)	25	369.2 (169.77)	25	

**Table 1.** User action summary of study 1: the table shows the user interaction statistics while performing each of the three tasks using two interfaces. (Statistical significance level: (\*)  $p < 0.05$ .)

difference on the action of change control panel and the click on the explanation tab between the interfaces in three tasks; although, in the SCATTER interface, the users tended to click the explanation functions less. The click frequency presented a significant difference between the two interfaces. This finding is not surprising because the RANK interface lacks the visualization information that pushes the participant to click more on the user profile page to inspect the necessary information. It is interesting to see that not every user clicked on the scatter plot. This data hints that some participants treated the scatter plot visualization as an explanation function rather than an interactive exploration interface. At the same time, we found no significant difference in the time spent on the tasks. The data hints that each action taken in the SCATTER interface delivered more interesting information with which to engage.

### User Feedback Analysis

To compare subjective feedback, we analyzed the responses of the post-stage questions using paired sample t-tests. Figure 3 shows the result of this analysis. We compared the twelve aspects of subjective feedback from the participants; among them, the SCATTER interface received a significantly higher rating for six aspects: Trust (Q4), Supportiveness (Q5), Interest (Q6), Satisfaction (Q8), Intention to Reuse (Q9), and Enjoyable (Q11). In two questions, facilitation (Q7) and the Reversed Benefit Question (Q12), the SCATTER interface scored higher, but not significantly so. It is interesting to see

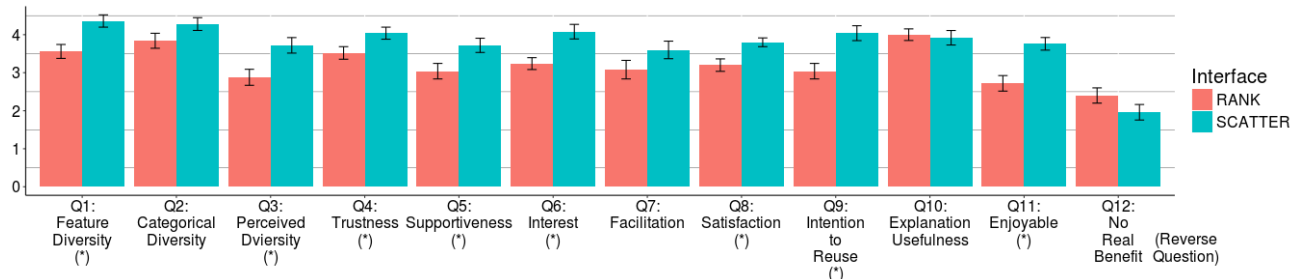


Figure 3. User feedback analysis results for Study 1 shows that the SCATTER interface received a significantly higher rating for six aspects. (A cut-off value was set at 3.5 on the 5 point scale. Statistical significance level: (\*)  $p < 0.05$ .)

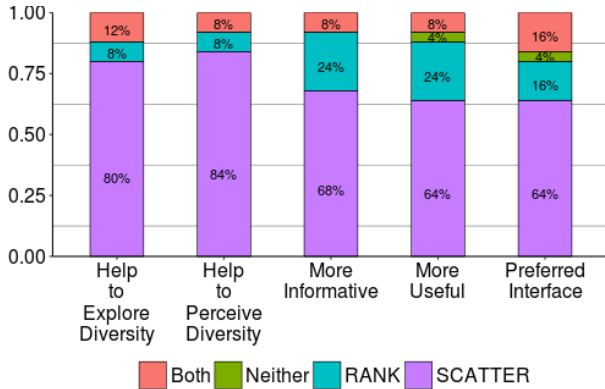


Figure 4. User preference analysis in study 1 (the preferences were collected after the users experienced both interfaces). The result shows the SCATTER interface was preferred by the users in all aspects.

that the RANK interface scored a bit higher (though not significantly so) on explanation usefulness, which hints that the lack of visualization made explanations more important in the RANK interface. In the final preference test, the SCATTER interface received much stronger support than the RANK interface in the user preference feedback (Figure 4). Most importantly, a majority of users (84%) considered the SCATTER interface to be a better system for recommending attendees and a better help in diversity-oriented tasks, as well as a better system for recommending.

**Diversity Analysis**

Table 2 shows the diversity analysis for each task and interface. The result shows the users’ responses to the tasks with a different pattern of exploration, which caused a variance of diversity and coverage measurements. All three tasks are shown a least one significance between two interfaces but in the different aspects of features. For the SCATTER interface, task 1 (relevance-oriented) shows significance statically on less difference between academic/social & social/interest features, but more coverage on the title category. Tasks 2 & 3 (diversity-oriented) show higher selection diversity in the interest/distance and social/distance features, respectively, as well as higher selection coverage in the title & country category features. The data supports the finding that the SCATTER interface helped the participants to accurately filter the attendees in the relevance-oriented task, as well as extend the selection diversity in the diversity-oriented tasks.

Task	Dimensions	RANK	SCATTER	P
		M (SE)	M (SE)	
T 1	Academic + Social	0.14 (0.06)	0.11 (0.03)	*
	Academic + Interest	0.16 (0.09)	0.13 (0.08)	
	Academic + Distance	0.13 (0.07)	0.12 (0.04)	
	Social + Interest	0.27 (0.13)	0.21 (0.09)	*
	Social + Distance	0.27 (0.12)	0.24 (0.08)	
	Interest + Distance	0.26 (0.13)	0.25 (0.13)	
	Title	0.17 (0.12)	0.22 (0.07)	*
	Position	0.26 (0.12)	0.23 (0.10)	
	Country	0.49 (0.25)	0.46 (0.15)	
	T 2	Academic + Social	0.14 (0.06)	0.13 (0.04)
Academic + Interest		0.14 (0.09)	0.14 (0.08)	
Academic + Distance		0.12 (0.07)	0.14 (0.04)	
Social + Interest		0.25 (0.13)	0.24 (0.11)	
Social + Distance		0.26 (0.13)	0.28 (0.10)	
Interest + Distance		0.23 (0.14)	0.27 (0.13)	-
Title		0.17 (0.14)	0.31 (0.12)	*
Position		0.29 (0.17)	0.25 (0.14)	
Country		0.46 (0.31)	0.68 (0.26)	*
T 3		Academic + Social	0.14 (0.06)	0.13 (0.03)
	Academic + Interest	0.12 (0.07)	0.12 (0.08)	
	Academic + Distance	0.14 (0.07)	0.16 (0.04)	
	Social + Interest	0.22 (0.12)	0.22 (0.14)	
	Social + Distance	0.24 (0.11)	0.31 (0.10)	*
	Interest + Distance	0.21 (0.14)	0.23 (0.11)	
	Title	0.17 (0.16)	0.32 (0.10)	*
	Position	0.19 (0.14)	0.15 (0.08)	
	Country	0.44 (0.26)	0.66 (0.26)	*

Table 2. Diversity analysis for study 1: the table shows selection diversity for three tasks in the feature and category dimensions. The result shows that the SCATTER interface can help users to explore a more diverse set of recommendation in diversity-oriented tasks (T2 & T3). (Statistical significance level: (\*)  $p < 0.05$ ; (-)  $p < 0.1$ .)

**Discussion**

In study 1, we evaluated a dual visual interface for recommending attendees at a research conference. A research conference context introduces several dimensions of attendee relevance, such as social, academic, interest, and distance similarities. Due to these factors, a traditional ensemble ranked list makes it difficult to express the diversity of recommended items (attendees). By spreading rankings over two dimensions, the suggested interface helps users to explore recommendations and recognize their diversity in several aspects. To assess the visual approach, we conducted a user study in a real conference environment to compare our interface (SCATTER) with a traditional ranked list (RANK) in three practical tasks. Our experimental result shows a tangible incremental impact on the metrics of system usage, efficiency, and diversity.

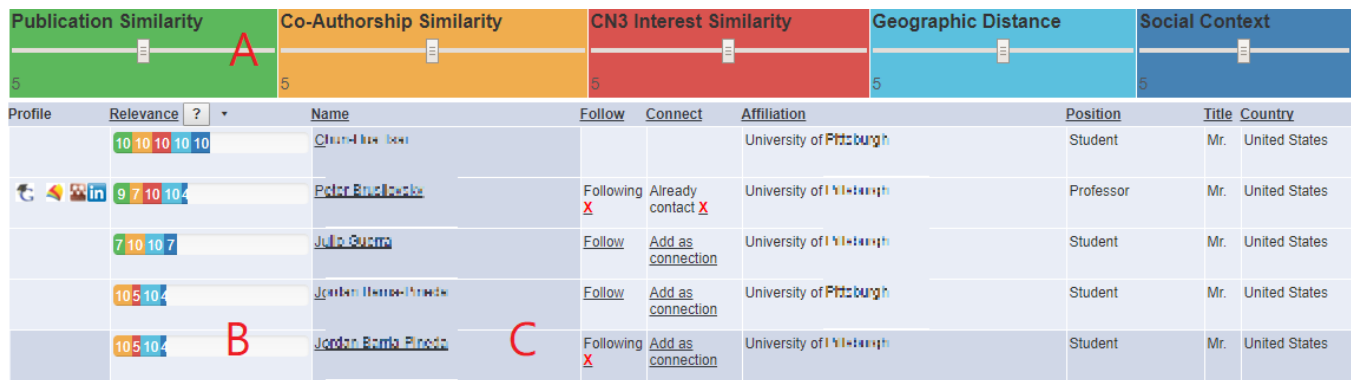


Figure 5. The design of the Relevance Tuner: (a) Relevance Sliders; (B) Stackable Score Bar; (C) User Profiles. The interface enables the user to adjust the feature weighting on-the-fly for retrieving a customized recommendation list. The user can examine the relevant aspects of the recommended item through the multicolored score bar. (The Name and Affiliation entities have been pixelated for privacy protection.)

We found that the **Scatter Viz** interface can improve user inspection on recommendations with multi-relevance, which leads to a higher selection diversity in the given tasks. However, we also noticed that some of the experiment participants still stick to the familiar ranked list, even when an enhanced visualization was provided. This finding helps us to realize a user preference on adopting the interface with lower learning efforts. Besides, the scatter visualization requires additional space to present, which may not be feasible in many real-world rank-based recommender systems. These findings lead to our second attempt at extending the ranked list with multi-aspect awareness, controllability, and diversity-aware designs.

## SECOND ATTEMPT: RELEVANCE TUNER

The ranked list is widely applied to recommender systems for presenting recommendations to users. Even in visual recommender systems, a basic ranked list is still essential for user interactions [44, 29, 40]. A qualified recommender system usually ranks the recommended items from high to low relevance, which may reduce the users' cognitive decision loading [2]. However, in a context with multiple relevance aspects, a statically ensembled ranked list makes it more difficult for the user to recognize the impact of different aspects and to adjust the recommendations to different needs. In addition, the persuasive design of the ranked list causes users to pay more attention to items on top of the list [8], which may cause the low selection diversity and the fitter bubble effect [25].

According to the first study, we knew the users were able to correlate the multiple relevances in a two-dimensional scatter plot visualization, but that this ability came with a high learning curve. In our second attempt, we would like to see if the user can still correlate the multi-relevance within a ranked list extended with a controllable tuner and stackable color bars. We proposed the **Relevance Tuner** - a visual interface with user-driven control function and meaningful visual encoding. This design aims to help the users to inspect and filter recommendations of multi-relevance from a ranked list, which may facilitate the exploration of diversity-oriented tasks. The interface is extensible to various types of recommender systems without the limitation of dual interfaces. This design reduces the user's difficulties in getting familiar with the interface.

## Visual Design

The design of the Relevance Tuner is shown in Figure 5. **Section A** contains five controllable sliders with the different colors representing the features of the Personalized Relevance Model. The scale of the slider ranges from 0 to 10. The user can change the weighting on the fly to re-rank the ranked list below. It provides controllability for the user to adjust the ranking to different recommendation needs and preferences. The interface also adds one new feature: Social Context. This feature computes the Google search result based on the scholar's name and affiliation information; that is, the text similarity of the homepage and other related search results. **Section B** shows the stackable relevance score bar of each recommended item in the ranked list. The color corresponds to the features in section A. It would adaptively adjust the bar score (length) from 0 to 20, based on the weighting percentage of the sliders. A stackable color bar interface is known for its ability to enhance controllability and transparency in a multi-aspect ranking [9]. In our system, the stackable color bars help the user to see how different relevant aspects of a recommended item are coordinated while adding transparency to the multi-aspect recommendation process. **Section C** shows the recommended scholar's meta-data, including name, social connection, affiliation, position, title, and country. The user can sort the ranked list by clicking the head of each column, or can inspect the explanation tabs (same as Section C in Figure 1) by clicking the name entities.

## STUDY 2: SCATTER VIZ V.S. RELEVANCE TUNER

### Data and Participants

Study 2 was conducted through the Conference Navigator 3 (CN3) system. The data was extended from Study 1 to a new conference: the 25th Conference on User Modeling, Adaptation, and Personalization (UMAP 2017). A total of 20 participants (7 female) were recruited for the user study. All of the participants were attendees at the UMAP 2017 conference. They were from 15 different countries; their ages ranged from 20 to 40 ( $M=31.19$ ,  $SE=4.97$ ). All of them had at least one publication from UMAP 2017. The background knowledge of recommender systems score was ( $M=3.85$ ,  $SE=0.79$ ) on a five-point scale.

Task	Action	TUNER		SCATTER		P
		M (SE)	User Count	M (SE)	User Count	
T 1	Control Panel	38.4 (37.71)	20	2.85 (2.23)	18	
	Explanation Tab	9.35 (8.28)	20	22.95 (23.71)	20	*
	Click - Rank	5.05 (2.45)	20	9.8 (8.43)	17	*
	Click - Scatter	-	-	4.1 (6.03)	11	
	Time Spent	357 (289.04)	20	537 (596.98)	20	
T 2	Control Panel	15.2 (13.63)	19	2 (1.71)	17	*
	Explanation Tab	6.5 (8.74)	20	8.45 (6.79)	20	
	Click - Rank	4.3 (1.21)	20	5.2 (3.76)	18	
	Click - Scatter	-	-	1.8 (2.94)	8	*
	Time Spent	201 (235.43)	20	294 (470.78)	20	
T 3	Control Panel	12.2 (11.67)	17	2.25 (1.80)	19	*
	Explanation Tab	9.25 (8.75)	20	12.9 (14.38)	20	
	Click - Rank	5.15 (2.51)	20	10.2 (16.93)	17	
	Click - Scatter	-	-	2.05 (3.13)	10	*
	Time Spent	153 (92.28)	20	285 (470.78)	20	

**Table 3. User action summary of study 2: the table shows the statistics of user interaction while solving each of the three tasks using each interface. (Statistical significance level: (\*)  $p < 0.05$ .)**

### Experiment Design and Procedure

In Study 2, we compared the interface of the ranked list plus the relevance tuner (TUNER) with a baseline of the scatter plot plus ranked list (SCATTER). The experiment design and procedure repeat the setting of Study 1. We manipulated the new proposed interface and adapted data from different conferences: the SCATTER interface used papers and attendees from UMAP 2017, while the TUNER interface used the same data from UMAP 2016, to minimize the learning effect between the two manipulations.

### Action Analysis

Table 3 shows the system usage for the two interfaces of Study 2. In TUNER interface, the control panel usage is defined as each time the user moves the sliders. The data supports the users interacting more frequently (it shows significance in all three tasks) with the control panel in TUNER than in SCATTER. Conversely, the users clicked more on explanation tabs in SCATTER than in TUNER. The data implies that the information listed on the table was sufficient for the users to inspect and make decisions in three proposed tasks. In task 1, the SCATTER has a significantly higher clicking frequency and longer time spent (not significant) than the TUNER interface. The same pattern repeats in task 2 & 3, which shows that the users took more time to get familiar with the SCATTER interface. The users were gaining familiar with the TUNER interface more rapidly than with the SCATTER interface.

### User Feedback Analysis

Figure 6 shows the analysis of the post-stage survey. The high rating in both interfaces shows the positive user acceptance in Study 2 (no significance on all the factors). However, the user tends to favor the TUNER interface when considering the factors of Supportiveness (Q5), Interest (Q6), Facilitation (Q7), Satisfaction (Q8), Intend to Reuse (Q9), and Usefulness (Q10). The SCATTER interface performs better on the measures of Trustiness (Q4) and Enjoyable (Q11). This result supports the users in favor of rank-based list more than the visual-based interface, but the visualization shows an increased level of usability on gaining trust and enjoyment in using the interface. Surprisingly, the feedback also indicates that the TUNER interface would be better for the user to fulfill the task on the feature diversity (Q1) and category diversity (Q2), but that the SCATTER interface is outperformed on the ability to Perceive Diversity (Q3).

This result shows that a user tends to use the ranked list with better controllability and transparency to conduct diversity-oriented tasks. The scatter visualization would play the role of helping the users to perceive diversity in multiple areas of relevance. The final preference result in Figure 7 also confirms this conclusion. About half of the users select the TUNER interface as the one with an advantage at helping to explore diversity, providing more informative information, being more useful, and fitting their preference - but the users also agree that the SCATTER interface could better help to perceive diversity after they finished the three tasks on two interfaces.

### Diversity Analysis

Table 4 shows the diversity analysis of Study 2. In task 1 (relevance-oriented), there is no significant difference between the two interfaces on the diversity measurement, showing both of the interfaces can support the user to fulfill a relevance-oriented task. However, in the diversity-oriented tasks 2&3, we found the TUNER group could achieve higher entropy than the SCATTER group. This finding hints that even with a ranked-list interface the user can achieve a good level of selection diversity if controllability and transparency for each considered dimension of relevance is available. At the same time, the SCATTER interface performed slightly (but not significantly) better than the TUNER interface in the *category* diversity metrics. This finding helps to highlight the value of color-coding data in the SCATTER interface, a feature not supported by TUNER. The diversity analysis of Social Context is omitted due to the page limitation.

### Discussion

In Study 2, we presented a new rank-based interface for recommender attendees at a research conference. A total of five dimensions of relevance were proposed from the Personalized Relevance Model. We conducted a user study in a real conference environment to compare the two interfaces of an enhanced ranked list (TUNER) and the visualization interface (SCATTER). Our experimental results suggested the different suitable scenarios for the two interfaces. We found that, even in diversity tasks with multi-relevance settings, the users were still able to fulfill the diversity task with a rank-based interface, but it required the support of interface controllability



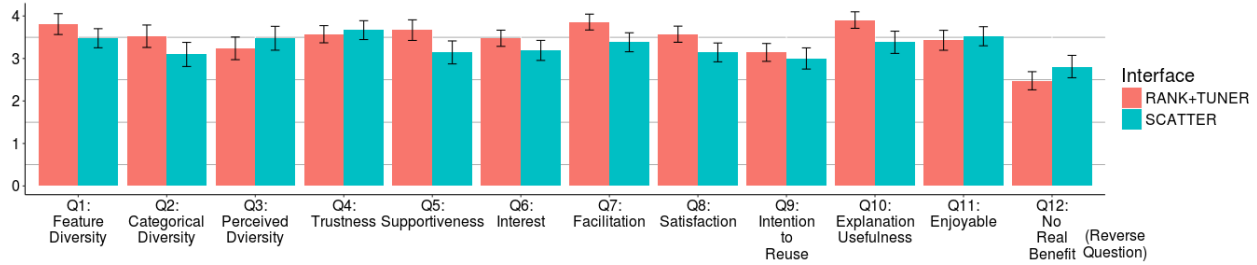


Figure 6. User feedback analysis results for study 2: we did not find significant difference in all aspects, which indicates that the usability of two interfaces was comparable. (A cut-off value was set at 3.5 on the 5 point scale. Statistical significance level: (\*)  $p < 0.05$ .)

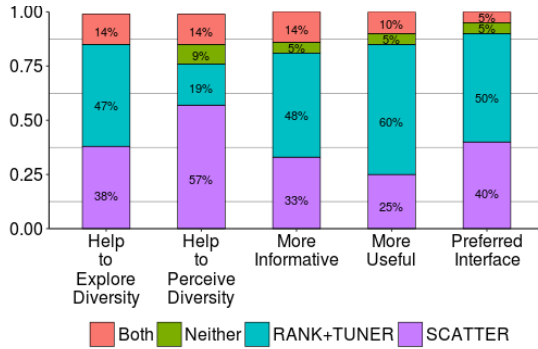


Figure 7. User preferences for study 2 collected after the users experienced both interfaces. The result shows that the TUNER interface was preferred by users in all aspects except perceived diversity.

Task	Dimensions	TUNER	SCATTER	P
		M (SE)	M (SE)	
T1	Academic + Social	0.12 (0.05)	0.12 (0.05)	
	Academic + Interest	0.17 (0.10)	0.16 (0.07)	
	Academic + Distance	0.13 (0.06)	0.12 (0.04)	
	Social + Interest	0.23 (0.11)	0.23 (0.10)	
	Social + Distance	0.23 (0.10)	0.22 (0.09)	
	Interest + Distance	0.24 (0.16)	0.29 (0.12)	
	Title	0.12 (0.04)	0.16 (0.13)	
	Position	0.29 (0.15)	0.27 (0.12)	
	Country	0.35 (0.20)	0.57 (0.29)	*
T2	Academic + Social	0.15 (0.02)	0.12 (0.04)	*
	Academic + Interest	0.17 (0.07)	0.13 (0.07)	*
	Academic + Distance	0.16 (0.06)	0.12 (0.05)	*
	Social + Interest	0.25 (0.07)	0.19 (0.08)	*
	Social + Distance	0.28 (0.07)	0.22 (0.10)	*
	Interest + Distance	0.31 (0.15)	0.24 (0.13)	*
	Title	0.15 (0.02)	0.17 (0.12)	
	Position	0.28 (0.12)	0.32 (0.15)	
	Country	0.41 (0.24)	0.58 (0.22)	*
T3	Academic + Social	0.16 (0.03)	0.13 (0.05)	
	Academic + Interest	0.14 (0.07)	0.14 (0.06)	
	Academic + Distance	0.17 (0.05)	0.14 (0.05)	*
	Social + Interest	0.32 (0.09)	0.26 (0.16)	
	Social + Distance	0.32 (0.08)	0.26 (0.09)	*
	Interest + Distance	0.28 (0.12)	0.26 (0.13)	
	Title	0.15 (0.02)	0.14 (0.05)	
	Position	0.16 (0.07)	0.22 (0.16)	
	Country	0.48 (0.19)	0.54 (0.32)	

Table 4. Diversity analysis for study 2: the table shows selection diversity for three tasks for each feature combination and category dimensions. The result indicates that the TUNER interface enabled users to explore a more diverse set of recommendation in diversity-oriented tasks T2 & T3. (Statistical significance level: (\*)  $p < 0.05$ .)

and transparency through visual encoding. Besides, while we found that the user would better perceive the diversity in the SCATTER interface, the user would prefer to adopt the TUNER interface to fulfill the diversity tasks.

Furthermore, when the user was interacting with the TUNER interface, the user spends more time on inspecting the information on each row, instead of checking the explanation functions. This result shows that the higher level of diversity exploration was not triggered by the diversity-enhanced visualization or explanation (fewer clicks on the explanation tabs), but was instead contributed by the intention of the user reaction to the simulated diversity-oriented tasks. In the SCATTER interface, the user relies more on the explanation function and multi-relevance visualization to explore diversity-oriented tasks. Although the result showed lower entropy measurement when the user adapted to the SCATTER interface, the SCATTER interface can better help the user to perceive the diversity among multiple-relevance dimensions, based on the user feedback analysis through post-study questionnaires.

### USER-CENTRIC EVALUATION

To better understand the mediation effects across the two interfaces, we conducted a structural equation model (SEM) analysis [20] to inspect the effects of the proposed interfaces on the user experiences. We used the logged data and questionnaire feedback from the two studies. There are two conditions and five summarized factors in the model (as shown in Figure 8). In objective system aspects (OSA), there are two manipulations based on the proposed interfaces. In subjective system aspects (SSA) and user experience (EXP), we proposed four factors based on the classification by [30] and our post-experiment questions. In interaction (INT), we listed the entropy of the participant’s selection diversity (average of Task 2&3 that shown in Table 2 and 4). The model fit the statistics of  $\chi^2(96) = 234.68, p < 0.01, RMSEA = 0.18, 90\%CI : [0.152, 0.211], CFI = 0.941, TLI = 0.922$ .

The model shows that the two manipulations have different positive effects on the system. The TUNER condition has a positive effect on the Perceived Usefulness and Perceived Diversity factors. It supports providing controllability and transparency on the ranked list, which helps the users to perceive the variety of recommendation results and increase the overall usefulness rating. The SCATTER condition only affects the Trust factor, which supports the argument for providing multi-relevance visualization to help the user to gain trust in the interface, but not mediates to the ratings of satisfaction or

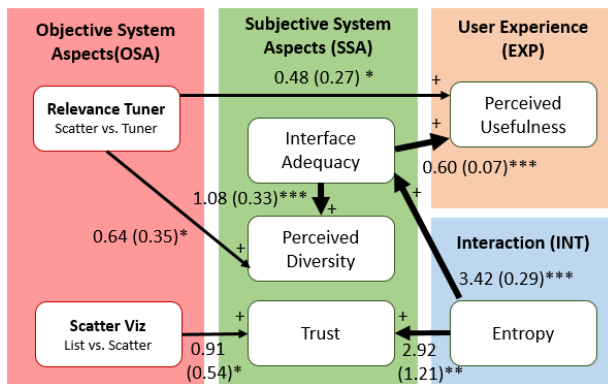


Figure 8. The structural equation model of the experiments. The number (thickness) on the arrows represents the  $\beta$  coefficients and standard error of the effect. Significance:  $*** p < 0.01$ ,  $** p < 0.05$ ,  $* p < 0.1$ .

usefulness. We also found the effect between Interface Adequacy and Perceived Diversity, which supports a better design interface that helps the user to perceive the recommendation diversity. Furthermore, a positive impact on Interface Adequacy to Perceived Usefulness implied that the usefulness is contributed by the rating of the interface design. The Entropy value mediates the additional effect of the two interfaces. We found that the Entropy factor has a positive impact on Interface Adequacy and Trust; that is, the users who can achieve higher selection diversity through the given interfaces gain more trust and interface adequacy ratings in the system. This finding explains why the rank-based interface, with a lower learning curve, is preferred by the participants on the proposed tasks. In other words, it also supports the multi-relevance visualization that is appreciated by the users only when they are capable of taking full advantage of the interface.

## CONCLUSION

In this paper, we presented experiments on three different interfaces: RANK, TUNER, and SCATTER. We first showed that providing a scatter plot (SCATTER) can help the user to better fulfill the diversity-related tasks, as compared to a simple ranked list (RANK). However, despite the benefits of the new two-dimensional presentation, the users still extensively use the ranked list component of the interface. Based on the results of the first study, we attempted to integrate the ability to coordinate multiple aspects of relevance within the ranked list rather than offering it in a separate component as in SCATTER. To compensate for the biasing nature of the ranked list, we also provided controllable fusion of relevance aspects. The resulting TUNER interface offered both controllability and visual encoding of multiple relevance aspects. We showed that the users could adopt a rank-based list to fulfill diversity-oriented tasks with higher selection diversity. The usability analysis reveals that both SCATTER and TUNER were ranked by the conference users with high subjective ratings. However, TUNER requires less learning effort. We also discussed the mediation effects of the proposed interfaces on the user experience. The analysis helps to describe the benefits of the two proposed interfaces in social recommender systems.

We aim to understand how the user adopts the recommender interfaces to diversity-oriented tasks. The given tasks indicated a concrete goal of exploring the conference attendees with multiple types of relevance. We found that both of the proposed interfaces were capable of helping the user to fulfill the assigned tasks. The experimental result supported the finding that the participants were able to correlate multiple aspects of relevance using two dimensions of visualization in SCATTER and the controllable ranked list with multi-aspect visualization in TUNER. We found that an extension of a traditional ranked list was better than a separate visualization component in the sense of on getting familiar with the new interface, which led to a higher rating on the user preference. In other words, a separate diversity-enhanced visualization can also achieve the goal, but it came at the cost of a steeper learning curve for the user. However, once the users were familiar with the interface, it brought an advantage of helping the users to perceive diversity and gain trust in the recommendations.

The findings shed light on designing diversity-enhanced and diversity-aware interfaces in a social recommender system. In a diversity-enhanced system, the target users would have a concrete goal of exploring the recommendation results. A similar scenario was reported in this paper; that is, the conference attendees explore the social recommendation within a particular conference. The user has specific goals on the exploration from multiple aspects of relevance. However, the users may not have specific goals when interacting with the recommender system. For example, when accessing news article on social media, the user may not have a strong desire to read the news article from a different perspective. In this case, we need a diversity-aware interface to help the user to perceive the prospect diversity, which may be a solution to help to solve the filter-bubble effect [36]. In this paper, we find that the effects of usability and perceived diversity are additive. The best user experience may happen when the users can take advantages from both of the proposed interfaces. It remains an open question as to how to design an interface to reflect the needs of users better.

## LIMITATIONS AND FUTURE WORK

Our study has some limitations. First, the within-subject user study was conducted using consecutive years of the same conference series. Some well-known and senior domain experts may appear in the recommendation list for both conferences. This repetition may cause bias in our user studies. Second, the data sparsity and cold-start problem may hurt the recommendation performance; for example, the Interest feature is less useful for users who never bookmarked any talks with CN3. We tried our best to send out emails both before and during the conference to improve interest-based recommendations. Third, the scale of reported user studies is relatively small. It may decrease the statistical power of the findings. Fourth, the experiment was conducted at mid-size conferences, so we were not able to explore scaling issues which might occur at conferences with much larger number of attendees or in a different recommendation context with a large set of items to explore. We hope to address some of these limitations in our future work.

## REFERENCES

1. Jisun An, Daniele Quercia, and Jon Crowcroft. 2013. Why individuals seek diverse opinions (or why they don't). In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 15–18.
2. Dirk Bollen, Bart P Knijnenburg, Martijn C Willemsen, and Mark Graus. 2010. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 63–70.
3. Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 35–42.
4. Engin Bozdag and Jeroen van den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17, 4 (2015), 249–265.
5. Peter Brusilovsky, Jung Sun Oh, Claudia López, Denis Parra, and Wei Jeng. 2016. Linking information and people in a social system for academic conferences. *New Review of Hypermedia and Multimedia* (2016), 1–31.
6. Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. 2009. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 201–210.
7. Henriette Cramer, Vanessa Evers, Satyan Ramalal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455.
8. Paolo Cremonesi, Franca Garzotto, and Roberto Turrin. 2012. Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 2 (2012), 11.
9. Cecilia di Sciascio, Vedran Sabol, and Eduardo E Veas. 2016. Rank As You Go: User-Driven Exploration of Search Results. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 118–129.
10. Michael D Ekstrand, Daniel Kluver, F Maxwell Harper, and Joseph A Konstan. 2015. Letting users choose recommender algorithms: An experimental study. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 11–18.
11. Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184.
12. Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Samuel Kaski, Giulio Jacucci, and others. 2013. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 117–128.
13. Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 833–842.
14. Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.
15. Rong Hu and Pearl Pu. 2011. Helping Users Perceive Recommendation Diversity.. In *DiveRS@ RecSys*. 43–50.
16. Marius Kaminskis and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2016), 2.
17. Hannah Kim, Jaegul Choo, Haesun Park, and Alex Endert. 2016. Interaxis: Steering scatterplot axes via observation-level interaction. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 131–140.
18. Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4189–4198.
19. Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. 2017. Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 57–66.
20. Bart P. Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. Inspectability and Control in Social Recommenders. In *6th ACM Conference on Recommender System*. 43–50.
21. Q Vera Liao and Wai-Tat Fu. 2013. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2359–2368.
22. Jennifer Moody and David H Glass. 2016. A Novel Classification Framework for Evaluating Individual and Aggregate Diversity in Top-N Recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 42.
23. Sean A Munson and Paul Resnick. 2010. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1457–1466.

24. Mark EJ Newman. 2001. Clustering and preferential attachment in growing networks. *Physical Review E* 64, 2 (2001), 025102.
25. Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. ACM, 677–686.
26. John O'Donovan, Brynjar Gretarsson, Svetlin Bostandjiev, Tobias Hollerer, and Barry Smyth. 2009. A visual interface for social information filtering. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, Vol. 4. IEEE, 74–81.
27. John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1085–1088.
28. Valeria Orso, Tuukka Ruotsalo, Jukka Leino, Luciano Gamberini, and Giulio Jacucci. 2017. Overlaying social information: The effects on users' search and information-selection behavior. *Information Processing & Management* 53, 6 (2017), 1269–1286.
29. Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.
30. Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.
31. Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, and others. 2013. Directing exploratory search with interactive intent modeling. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1759–1764.
32. J Ben Schafer, Joseph A Konstan, and John Riedl. 2002. Meta-recommendation systems: user-controlled integration of diverse recommendations. In *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, 43–51.
33. Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 990–998.
34. Robert Tarjan. 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing* 1, 2 (1972), 146–160.
35. Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and SVN Vishwanathan. 2016. Adaptive, Personalized Diversity for Visual Discovery. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 35–38.
36. Nina Tintarev. 2017. Presenting Diversity Aware Recommendations: Making Challenging News Acceptable. (2017).
37. Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (1 Oct. 2012), 399–439.
38. Chun-Hua Tsai. 2017. An Interactive and Interpretable Interface for Diversity in Recommender Systems. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion (IUI '17 Companion)*. ACM, New York, NY, USA, 225–228.
39. Chun-Hua Tsai and Peter Brusilovsky. 2016. A personalized people recommender system using global search approach. *ICConference 2016 Proceedings* (2016).
40. Chun-Hua Tsai and Peter Brusilovsky. 2017a. Enhancing Recommendation Diversity Through a Dual Recommendation Interface. In *Workshop on Interfaces and Human Decision Making for Recommender Systems*.
41. Chun-Hua Tsai and Peter Brusilovsky. 2017b. Leveraging Interfaces to Improve Recommendation Diversity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 65–70.
42. Chun-Hua Tsai and Peter Brusilovsky. 2017c. Providing Control and Transparency in a Social Recommender System for Academic Conferences. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 313–317.
43. Chun-Hua Tsai and Yu-Ru Lin. 2016. Tracing and Predicting Collaboration for Junior Scholars. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 375–380.
44. Katrien Verbert, Denis Parra, Peter Brusilovsky, and Erik Duval. 2013. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 351–362.
45. Weiquan Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246.
46. David Wong, Siamak Faridani, Ephrat Bitton, Björn Hartmann, and Ken Goldberg. 2011. The diversity donut: enabling participant control over the diversity of recommended responses. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1471–1476.