6-7-2019

# Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance

Chun-Hua Tsai

Peter Brusilovsky

# Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance

Chun-Hua Tsai University of Pittsburgh Pittsburgh, USA cht77@pitt.edu

Peter Brusilovsky University of Pittsburgh Pittsburgh, USA peterb@pitt.edu

**ABSTRACT**

Recommender system helps users to reduce information overload. In recent years, enhancing explainability in recommender systems has drawn more and more attention in the field of Human-Computer Interaction (HCI). However, it is not clear whether a user-preferred explanation interface can maintain the same level of performance while the users are exploring or comparing the recommendations. In this paper, we introduced a participatory process of designing explanation interfaces with multiple explanatory goals for three similarity-based recommendation models. We investigate the relations of user perception and performance with two user studies. In the first study (N=15), we conducted card-sorting and semi- interview to identify the user preferred interfaces. In the second study (N=18), we carry out a performance-focused evaluation of six explanation interfaces. The result suggests that the user-preferred interface may not guarantee the same level of performance.

**CCS CONCEPTS**

• **Information systems → Recommender systems**; • **Human- centered computing → HCI design and evaluation methods**.

**KEYWORDS**

Visual Explanation; Recommendation; Similarity-Based

**ACM Reference Format:**

Chun-Hua Tsai and Peter Brusilovsky. 2019. Evaluating Visual Explanations for Similarity-Based Recommendations: User Perception and Performance. In *27th Conference on User Modeling, Adaptation and Personalization (UMAP '19), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3320435.3320465

## 1 INTRODUCTION

Recommender systems are typically optimized to produce a top-N list of the user-relevant items. It is a useful approach to solve the problem of information overload by exposing users to only a small set of options. However, a user may not accept the top recommendations blindly. Instead, it is common to see a user to perform a further search or compare the recommended items [27, 34]. An effective recommender system needs to consider factors beyond the recommendation accuracy aiming for better user interaction as the recommendation accuracy aiming for better user interaction as a whole [22].

In recent years, researchers in the field of recommender systems explored a range of advanced interfaces to support exploration, transparency, explainability, and controllability of recommendations [10]. *Controllability* enabled end users to participate in the recommendation process by providing various kinds of input [1, 18, 26], e.g., adjust preference or explore recommendations. *Transparency* features allowed interactive recommender systems to deal with the "black-box" problem, i.e., to explain the inner logic of the recommendation process to the end users [22, 32]. In the case when recommendation mechanism is too complicated to non-professional users to explain, some considerable transparency could be achieved by *explaiability*, i.e., the system may just need to *justify* why the recommendation was presented [3, 23, 27].

Enhancing explainability in recommender systems has drawn more and more attention. Explaining recommendations can achieve different *explanatory goals* by single-style or hybrid explanations [13, 20, 24]. A number of explanation interfaces and approaches have been proposed and studied to assess improvement of user satisfaction and other aspects [24]. However, most of the evaluation focus on solely user perception or preferences [13, 17]. In most cases, it remains unclear whether different kinds of explanations could improve the objective parameters of user performance rather than their perception of which option is better while *inspecting* the explanation interface [12].

This paper presents two user studies, which we conducted to develop and evaluate explanation interfaces for three similarity- based people recommendation models. In the first study (N=15), we introduced a card-sorting task to identify the

user preferred explanation interface design for multiple explanation goals. We assessed a total of fifteen explanation interfaces in the first study and selected the top-voted interface designs for each similarity- based recommendation model. In the second study (N=18), we used a performance-focused evaluation approach to investigate whether using two types of explanations in parallel could offer an advantage over a single type of explanation. We compared six explanation interfaces, one baseline plus one enhanced version for each of the three recommendation models. In each case, we use the top-rated interface as a baseline for each model and a com- bination of first and second preferred interfaces as an enhanced version. We implemented six explanation interfaces using the data from a state-of-the-art scholarly social recommender system. We evaluated the explanation interfaces by asking the participant to “sort” recommendation based on the relevance. We analyzed the findings combined with the sorting result, user perception survey, and NASA-TLX survey.

Our contribution is three-fold. First, we reported the participatory design process of developing the explanation interfaces for three similarity-based recommendation models. Second, we evaluated the explanation interfaces along multiple explanatory goals and discussed the changing performance across explanation inter- faces. Third, we introduced a correlation analysis to reveal hidden relationships between survey and user behavioral variables.

## 2 RELATED WORK

Enhancing explainability in recommender systems has drawn more and more attention. The first driving force of this process is the increased interest in the user-centered evaluation metric of a recommender system, i.e., evaluating the system from the point of the user instead of the offline accuracy-oriented algorithm refinements. Self-explainable recommender systems have been proven to increase user perception on system transparency [25], trust [19] and acceptance rate of system suggestions [12]. The second driving force is the newly initiated European Union’s General Data Protection Regulation (GDPR), which required the manager of any data-driven application to include a “right to the explanation” of algorithmic decisions [5],

thus urging to increase transparency in all existing intelligent systems. Explaining recommendations (i.e., algorithmic decisions produced by recommender systems) can achieve different *explanatory goals* helping users to make a better decision or persuading them to accept the suggestions from a system [20, 24]. Tintarev and Masthoff [25] reviewed seven explanatory goals: *Transparency*, *Scrutability*, *Trust*, *Persuasiveness*, *Effectiveness*, *Efficiency*, and *Satisfaction*. Since it is hard to have a single explanation interface that achieves all these goals equally well, the designer needs to make a trade-off while choosing or designing the form of interface [25]. For instance, a highly interactive interface could increase the user trust and satisfaction but may prolong the decision and exploration process while using the system (i.e., lead to a decrease of efficiency) [27].

Explanations can be categorized by their styles, reasoning models, paradigms, and information [6]. For instance: 1) *Styles:* Kouki et al. [13] conducted an online user survey to explore user preferences of nine explanation styles. They found that Venn diagrams outperformed all other visual and text-based interfaces. 2) *Reasoning Model:* Vig et al. [32] used tags to explain the recommended item and the user's profile that emphasized the factor of why a specific recommendation is plausible, instead of revealing the process of recommendation or data. *3) Paradigm:* Herlocker et al. [11] presented a model for explanations based on the user's conceptual model of the recommendation process. The result of the evaluation indicated that two interfaces - Histogram with grouping and Presenting past performance - improved the acceptance of recommendations. 4) *Information:* Pu and Chen [19] proposed explanations tailored to the user and recommendation, i.e., although one recommendation is not the most popular one, the explanation would justify the recommendation by providing the reasons.

## 3 SIMILARITY-BASED RECOMMENDATIONS

In this paper, we present the results of our attempts to design and evaluate visual explanations for three similarity-based people recommendation models: *text similarity*, *topic similarity* and *item similarity*. These models are widely adopted in many content-based recommender systems [27, 30, 32].

**Text similarity (E1)** is a metric that measures similarity or dis- similarity (distance) between two text strings [8]. The "strings" can consist of various information sources. For example, in a scholarly people recommender system, the string can be generated from scholar' academic publications. To measure the text similarity (distance), one promising approach is to convert the strings into a *term vectors* and then compute their *cosine similarity* [27]. A higher similarity (i.e., the shortest distance) between "strings" representing publications of two researchers indicates that the two researchers have a larger fraction of common terms in the text of their publications.

**Topic similarity (E2)** is a metric that measures the distance between topic distributions [4]. This is another approach to measure the similarity between the publications of two researchers. The approach assumes that a mixture of topics is used to generate a string (document), where each topic is a distribution of topical words. A social recommender engine, based on the topic-based approach, can represent the scholars' research interests through the learned *topics*. The topic similarity could be computed as the pairwise similarity of the topic distributions [30]. In our study, the *topics* were generated by topic modeling, Latent Dirichlet Allocation (LDA), by classifying their publication text [4]. A higher topic similarity means a shorter distance between the two scholars' research interests, i.e., the two scholars shared more common research topics.

**Item similarity (E3)** is a metric that measures the portion of shared items, which can be varied in a different context. For ex- ample, items shared by two users could be user-generated tags [32] or friends followed on social media. In a scholar recommender system, we can calculate the similarity between two scholars by measuring the intersection of papers bookmarked by these scholars [2, 14], e.g., using Jaccard similarity [27, 30]. A higher item similarity means that the two scholars have more similar interests in respect to the academic articles or conference presentation, i.e., they co-bookmarked a larger number papers at the same conference.

## 4 DEVELOPING EXPLANATION INTERFACES

We designed visual interfaces to explain three similarity-based models for recommending conference attendees to meet implemented in a conference support

system Conference Navigator (CN) [2]. All interfaces were selected to visually explain one type of “similarity” between the user and a recommended scholar described in the previous section. Our goal at this stage was to find visualizations that can better explain the similarity model as measured by user perception. Existing state-of-the-art explanation interfaces or models motivated our interface designs. Due to the page limit, this paper shows only top-performing designs (see Figure 1). The full set of designs can be found in [29].

**4.1 Study 1: Comparing Explanation Interfaces** We conducted the first user study to determine the *user preferred* visual interfaces of explaining the three similarity-based recommendation models (E1, E2 & E3). The participants were asked to complete *closed card-sorting tasks* to organize the proposed inter- faces into predefined groups. The tasks were designed to evaluate how well a visual interface supports the exploratory goal. A total of nineteen factor across seven explanatory goals were introduced in the study [24, 28]. The seven factors included Transparency (TP), Scrutability (SC), Trust (TS), Persuasiveness (PE), Effectiveness (ET), Efficiency (EF) and Satisfaction (SA). The detailed statement of each factor can be found in Table 1.

At the beginning of the study, we introduced the CN system and the recommendation models to the subjects. After the introduction, we asked the subjects to complete a closed card-sorting task for each recommendation model. In each task, we presented five ex- planation interfaces (paper mock-ups) and asked the subjects to assign the interfaces to group 1-5 (from Group 1: Strongly Agree; to Group 5: Strongly Disagree, or Not Applicable) based on the given exploratory factors (listed in Table 1). The experiment followed within-subject design, i.e., all participants required to perform three card sorting tasks (i.e., one for each group) with the same nineteen explanatory factors. The order of tasks and factors was the same to all participants. We continued with a semi-interview after each task to collect the qualitative feedback.

A total of 15 (6 female) participants (N=15) were recruited. They were first or second-year information science graduate students at the University of Pittsburgh with age ranged from 20 to 30 (M=25.73, SE=2.89). All participants had no previous experience of using the CN system. Each participant received USD$20 compensation

and signed an informed consent form. Subjects took between 40 and 60 minutes to complete the study.

**Table 1: The Explanation Factors**

| | Factor | Statement |
|---|---|---|
| 1 | TP, SA | The visualization presents the similarity between my interest and the recommended person. |
| 2 | TP | The visualization presents the relationship between the recommended person and me. |
| 3 | TP | The visualization presents where the data was retrieved. |
| 4 | TP | The visualization presents more in-depth information on how the scores sum up. |
| 5 | TP, TS, ET | The visualization allows me to see the connections between people and understand how they are connected. |
| 6 | SC | The visualization allows me to understand whether the recommendation is good or not. |
| 7 | SC | The visualization presents the data for making the recommendations. |
| 8 | SC | The visualization allows me to compare and decide whether the system is correct or wrong. |
| 9 | SC | The visualization allows me to explore and then determine the recommendation quality. |
| 10 | TS | The visualization presents a convincing explanation to justify the recommendation. |
| 11 | TS | The visualization presents the components (e.g., algorithm) that influenced the recommendation. |
| 12 | PE | The visualization shows me the shared interests, i.e., why my interests are aligned with the recommended person. |
| 13 | PE, SA | The visualization has a friendly, easy-to-use interface. |
| 14 | PE | The visualization inspired my curiosity to discover more information. |
| 15 | ET | The visualization presents the recommendation process clearly. |
| 16 | EF | The visualization presents highlighted items or information that is strongly related to me. |
| 17 | EF | The visualization presents aggregated, non-obvious relations to me. |
| 18 | SA | The visualization presents feedback from other users, i.e., I can see how others rated a recommended person. |
| 19 | SA | The visualization allows me to tell why does this system recommend the person to me. |

### 4.2 Explaining Text Similarity (E1)

The key component of text similarity is *terms* and *term frequency* of the publication as well as its mutual relationship (i.e., the common terms) between two scholars. We presented one text-based interface (**E1-1**) and four visual interfaces (**E1-2** to **E1-5**) for explaining text similarity.

**E1-1 Text-Based Explanation:** The text-based interface was presenting the explanation as: *You and [the scholar] have common words in [W1], [W2], [W3].*

**(Second-rated) E1-2 Two-way Bar Chart:** A bar chart is a common approach in analyzing the text mining outcome using a histogram of terms and term frequency [21]. We extended the design to a two-way bar chart to better compare two scholars' publication terms and term frequency, i.e., one scholar on the right and the other scholar on left (Figure 1b).

**E1-3 Word Clouds:** Word cloud is a universal design in explaining text similarity [7, 26]. We adopted the word cloud style from [33], which presented the term in the cloud and the term frequency by the font size. We used two-word clouds (one for each scholar), so the user can perceive the mutual relationship.

**(Top-rated) E1-4 Venn Word Cloud:** This interface could be considered as a combination of a word cloud and a Venn diagram [30], which presents term frequency using the font size. The unique terms of each scholar are shown in a different color (green and blue) while the common terms are presented in the middle, with red color, for determining the mutual relationship (Figure 1a).

**E1-5 Interactive Word Cloud:** A word cloud can be interactive. We extend the idea from [26] and used "Zoomdata Wordcloud" tool [35], which follows the common approach to visualize term frequency with the font size. The term color was selected to distinguish the scholars' terms, i.e., different term color for each scholar. A slider was attached to the bottom of the interface that provides a real-time interactive functionality to increase or decrease the number of terms in the word cloud.

### 4.3 Explaining Topic Similarity (E2)

The key component of topic similarity is *research topics* and *topical words* of the scholar as well as its mutual relationship (i.e., the common research topics) between

two scholars. We presented one text-based interface (**E2-1**) and four visual interfaces (**E2-2** to **E2-5**) for explaining topic similarity.

**E2-1 Text-Based Explanation:** The interface was presenting the explanation as: *You and [the scholar] have common research topics on [T1], [T2], [T3].*

**E2-2 Topical Words:** This interface extended the approach by McAuley and Leskovec [15], which attempted to enable topic interpretation by presenting topical words in a table. We adopted the idea as *E2-2 Topical Words* that present the topical words in two multi-column tables (each column contains ten topical words).

**(Second-rated) E2-3 FLAME:** This interface was proposed by Wu and Ester [33], which adopted a bar chart and two word-clouds in displaying the topical mining result. The user can interpret the topic model by the diagram (for the *beta* value of topic) and the table (for the topical words). We extended the idea as *E2-3: FLAME* that showed two sets of research topics (top 5) and the relevant topic words in two word-clouds (one for each scholar). (Figure 1d)

**(Top-rated) E2-4 Topical Radar:** The interface was introduced by [30], which presented a radar diagram with a topical word table. The radar filed top 5 topics (ranked by *beta* value) of the user and the corresponding value of the recommended scholar. The table with topical words was presented in the right so that the user can inspect the context of each research topic. (Figure 1c)

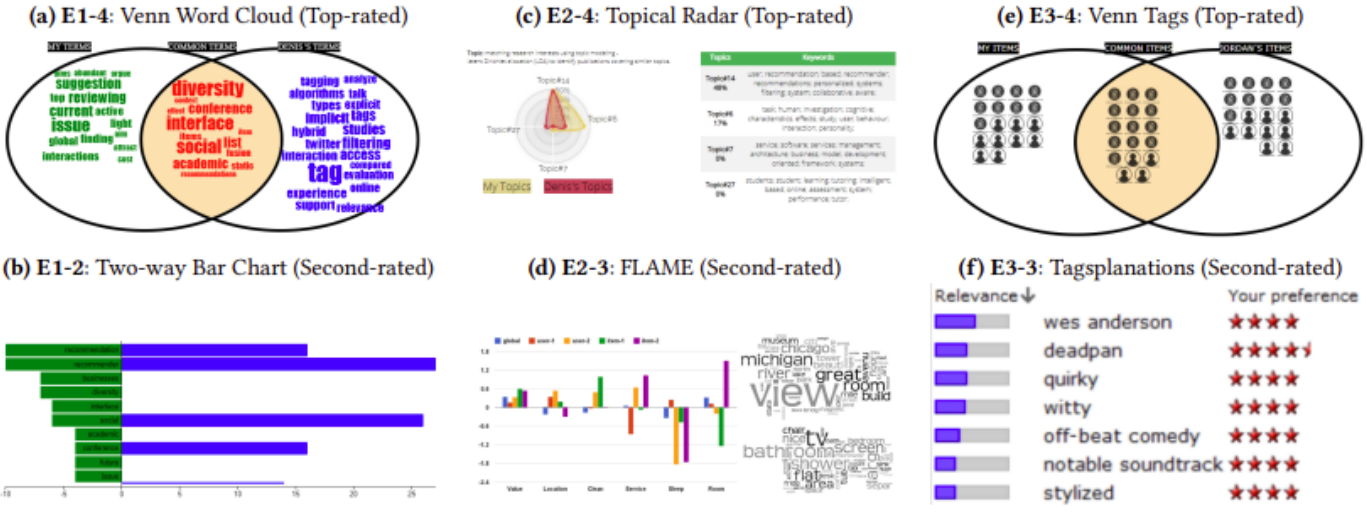


**Figure 1: The top-rated and second-rated visual interfaces in the first user study.**

**E2-5 Topical Bars:** Bar chart have been shown useful in analyzing the frequency across different topics [21]. In this interface, we adopted multiple bar charts to show six topic distribution of two scholars and the topical information (words and topic *beta* value).

### 4.4 Explaining Item Similarity (E3)

The key component of item similarity is the *papers* and *authors* of the bookmarking as well as its mutual relationship (i.e., the common items) between two scholars. We present the five prototyping inter- faces for explaining item similarity. In addition to four visualized interfaces (E3-2 to E3-5), we also include one text-based interface (E3-1) for explaining item similarity.

**E3-1 Text-Based Explanation:** The interface was presenting the explanation as: *You and [the scholar] have common bookmarking, they are [B1], [B2], [B3].*

**E3-2 Similar Keywords:** The interface was proposed and deployed by the CN system [16]. We extended the interface to explain common bookmark between two scholars. The interface represents the scholars in two sides and the common co-bookmarking items (e.g., the five common co-bookmark papers or authors) in the middle. A strong (solid line) or weak (dash line) tie will be used to connect the item was bookmarked by the one-side or two-sides.

**(Second-rated) E3-3: Tagsplanations:** The visualization was proposed by Vig et al. [32]. The idea is to show tag and relevance in an ordered bar. We extended the interface to explain the co- bookmarking information. In our design, the co-bookmarked item will be listed and ranked by its social popularity, i.e., how many users have followed/bookmarked the item? (Figure 1f)

**(Top-rated) E3-4: Venn Tags:** The study [13, 18] pointed out that users preferred Venn diagrams as a way to explain recommendation. In the interface, presented items bookmarked by compared scholars as icons on the Venn diagram. Two sides of the diagram show bookmarked item belonging only to one of the compared scholars. The co-bookmarked item are presented in the middle. The users can mouse over the icon for detail information, i.e., paper title. (Figure 1e)

**E3-5: Itemized List:** An itemized list has been adopted to explain the bookmark

in [31]. We extended the design in presenting the bookmarked or followed items in two comparable itemized lists.

### 4.5 Card-Sorting Analysis

The card-sorting results are presented in Table 2. The first (top- rated) and the second (second-rated) most-preferred interfaces are highlighted in red and blue color, respectively. In general, the result indicated that the participants preferred visual explanations over text-based explanations. The pattern was consistent in all three tasks; the text-based explanations were always received the most *Group 5* and *Not Applicable* votes.

In explaining text similarity (E1), the *E1-4 Venn Word Cloud* was preferred by the participants (received 117 votes in Group 1) outperforming other four interfaces. According to the post-stage interview, 13 subjects agreed that E1-4 is the best interface due to the following reasons. 1) the Venn diagram provided common terms in the middle, which highlighted the common terms and shared relationship; 2) it is useful to show non-overlapping terms on the sides (N=5) and 3) the design is simple, easy to understand and require less time to process (N=3). Two subjects particularly mentioned they preferred *E1-2: Two-way Bar Chart* (received 47 votes in Group 1, second-rated interface) since histograms gives them the “concrete numbers” for “calculating” the similarity, which was harder when using word clouds.

In explaining topic similarity (E2), the *E2-4 Topical Radar* received 137 votes in Group 1 outperforming all other interfaces. *E2-3 FLAME* ended up second in Group 1 as well as received the most votes in Group 2. According to the post-stage interview, 13 subjects agreed E2-4 is the best interface among all examined interfaces. The supporting reasons for E2-4 can be summarized as 1) It is easy to see the relevance through the overlapping area from radar chart and the percentage numbers from the table (N=12). 2) It is informative to compare the shared research topics and topical words (N=9). One subject specifically preferred E2-3, and one subject suggested a mix of E2-3 and E2-4 as the best design.

**Table 2: The Card-sorting Result of Study 1**

| | R1 | R2 | R3 | R4 | R5 | Not Applicable | Total Votes |
|---|---|---|---|---|---|---|---|
| E-1 | 32 | 39 | 30 | 30 | 76 | 78 | 285 |
| E1-2 | 47 | 55 | 31 | 56 | 35 | 61 | 285 |
| E1-3 | 12 | 33 | 68 | 68 | 34 | 70 | 285 |
| E1-4 | 117 | 57 | 44 | 8 | 2 | 57 | 285 |
| E1-5 | 38 | 50 | 56 | 45 | 31 | 65 | 285 |
| | | | | | | | |
| E2-1 | 18 | 9 | 23 | 29 | 113 | 93 | 285 |
| E2-2 | 18 | 18 | 22 | 119 | 49 | 59 | 285 |
| E2-3 | 48 | 125 | 56 | 13 | 2 | 41 | 285 |
| E2-4 | 137 | 55 | 26 | 25 | 2 | 40 | 285 |
| E2-5 | 30 | 39 | 108 | 35 | 13 | 60 | 285 |
| | | | | | | | |
| E3-1 | 11 | 13 | 6 | 48 | 113 | 94 | 285 |
| E3-2 | 32 | 103 | 87 | 22 | 2 | 39 | 285 |
| E3-3 | 91 | 75 | 66 | 14 | 4 | 35 | 285 |
| E3-4 | 101 | 48 | 68 | 15 | 2 | 51 | 285 |
| E3-5 | 17 | 10 | 12 | 116 | 63 | 67 | 285 |

In explaining item similarity (E3), the *E3-4 Venn Tags* received 101 votes in Group 1 outperforming the other four interfaces. *E3-3 Tagsplanations* finished as a very close second receiving 91 votes. According to the post-stage interview, eight subjects agreed that E3-4 is the best interface among the five interfaces. The supporting reasons can be summarized as 1) the Venn diagram is more familiar or clear than other interfaces (N=4); 2) The Venn diagram is simple and easy to understand (N=4). Three subjects particularly mentioned that they preferred E3-3 since this interface provide extra information without requiring extra mouse-hovering efforts while inspecting the details.

## 5 ASSESSING VISUAL EXPLANATIONS

Based on the card-sorting result of study 1, we implemented the top- rated designs to assess visual explanation interfaces more reliably. The screenshot of the

most-preferred interfaces can be found in Figure 1. We proposed 1) *E1-4 Venn Word Cloud* to explain text similarity (**Sim1**), 2) *E2-4 Topical Radar* to explain topic similarity (**Sim2**) and 3) *E3-4: Venn Tags* to explain item similarity (**Sim3**).

At the same time, the result of the post-stage interview indicated that while second-rated interfaces collected fewer votes than top-rated interfaces, participants mentioned different reasons for preferring these interfaces. We hypothesized that the features of first and second design choices could complement each other and decided to explore whether we could improve the value of the most- preferred interface by enhancing it with the second-most-preferred design. That is, we added the *E1-2 Two-way Bar Chart* to *E1-4 Venn Word Cloud* to provide additional term comparison information (**Sim1+**), attached two word clouds to *E2-4 Topical Radar* to mix up the user preferred component of E2-3 (**Sim2+**), and provided an extra list to *E3-4: Venn Tags* (**Sim3+**) to decrease the need for mousing-over while getting the item details. We aimed to answer the following research questions (RQs):

- How does the visual interface reach the explanation goals?
- How does user perception vary with the enhanced interface?
- How does the explanation interface affect the user performance (inspectability) across recommendations?

**5.1 Study 2: Evaluating Explanation Interfaces** To answer the research questions, we conducted a controlled user study to evaluate and compare the selected interfaces for explaining the three similarity-based recommendation models. We introduced a total of six explanation interfaces (three baseline and three enhanced interfaces) in the context of the attendee recommender component of the Conference Navigator (CN) (same as study 1) [2]. A total of 18 (11 female) participants (N=18) were recruited for this study. There were 16 information science graduate students and 2 graduate from nursing and linguistics programs at the University of Pittsburgh. Their age ranged from 21 to 35 years ($M$ = 24.94, $SE$ = 3.24). All participants had no previous experience of using the CN system. Each participant received USD$20 compensation and signed informed consent.

We first introduced the CN system and the recommendation models to the

subjects. After the introduction, we asked the sub- jects to complete a "recommendation-sorting task" using the given explanation interface, i.e., the subjects were required to *rank the recommendation relevance solely based on the visual explanation*. The tasks were designed to evaluate *how well an explanation interface supports the user performance of comparing the relevance across recommendations*. The experiment adopted a within-subject design, i.e., all participants were asked to perform six sorting tasks using the proposed explanation interfaces. In each task, the subject received five people recommendations generated by one recommendation model. To make the conditions equal, all users received the same recommendation generated using data of a scholar who used the CN system for at five conferences. The subjects can click the recommendation link to open the corresponding explanation interface. The five people recommendations were displayed as five links with names of the recommended scholars. All related back- ground information (e.g., list of publications, affiliation, title, etc.) was hidden to reduce the bias. The order of recommendation and explanation interfaces were randomized to avoid the ordering effect.

**Table 3: Log Activity Analysis**

| Sim1 | | Sim1+ | Sim2 | Sim2+ | Sim3 | Sim3+ |
|---|---|---|---|---|---|---|
| Variable | M (SE) | M (SE) | M(SE) | M (SE) | M (SE) | M (SE) |
| Clicks | 11.16 (1.68) | **18.22 (5.74) **** | 5.17 (0.39) | **10.37 (2.14) **** | 6.00 (1.57) | 6.94 (2.38) |
| Time (Secs) | 383.27 (206.70) | 382.94 (165.03) | 346.58 (122.23) | 399.88 (132.56) | 308.00 (176.43) | 348.88 (172.77) |

To reduce the learning bias, we used data from different conferences to generate recommendations, i.e., IUI 2017 for the baseline interfaces and UMAP 2017 for the enhanced interfaces.

After each task, the subjects were asked to fill-in a three-part post-stage questionnaire. First, the subjects were asked to rank the five recommendations by

relevance (from high to low relevance). We measured the correct rate by *Levenshtein Distance*, i.e., given correct order as “ABCDE” and submitted answer as “ABDCE”, the Levenshtein distance is 2 and the correct rate is 60% ((5−2)/5 = 0.6). Second, the subjects answered the nineteen-factor questions (shown in Table 1). Third, the subjects answered four NASA-TLX questions [9]. The NASA-TLX question included: *(TLX1) Mental Demand: How mentally demanding was the task? (TLX4) Performance: How successful were you in accomplishing what you were asked to do? (TLX5) Effort: How hard did you have to work to accomplish your level of performance?* and *(TLX6) Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?* The order of question was the same to all participants with a 5-point scale (1=Strongly Disagree/Very Low, and 5=Strongly Agree/Very High).

**Table 4: Task Survey Analysis**

| Sim1 | | Sim1+ | Sim2 | Sim2+ | Sim3 | Sim3+ |
|---|---|---|---|---|---|---|
| Goal | M (SE) | M (SE) | M(SE) | M (SE) | M (SE) | M (SE) |
| TR | 3.37 (0.77) | **3.92 (0.71)*** | 3.64 (0.91) | 4.00 (0.78) | 3.71 (0.83) | 3.80 (0.74) |
| SC | 3.22 (0.76) | **4.26 (0.77)**** | 4.02 (0.81) | 4.05 (0.96) | 3.76 (0.80) | 3.81 (1.04) |
| TS | 3.33 (0.77) | **3.88 (0.64)*** | 3.70 (0.92) | 4.01 (0.75) | 3.75 (0.88) | 3.72 (0.85) |
| PE | 3.62 (0.78) | 3.87 (0.83) | 4.15 (0.60) | 4.09 (0.66) | 3.92 (0.87) | 4.14 (0.77) |
| ET | 3.25 (0.66) | **3.92 (0.64)**** | 3.68 (0.85) | 4.00 (0.80) | 3.59 (1.04) | 3.64 (0.93) |
| EF | 3.41 (0.82) | 3.66 (0.98) | 3.38 (0.92) | 3.63 (0.92) | 3.13 (0.70) | 3.25 (0.89) |
| SA | 3.40 (0.70) | 3.62 (0.72) | 3.70 (0.69) | 3.91 (0.66) | 3.65 (0.61) | 4.04 (0.71) |

5.2 **Behavior difference among visualizations** User activity while performing the recommendation-sorting tasks was logged. All explanation interfaces were static, so the user behavior was relatively simple. We tracked the number of *mouse clicks* (click to view explanation interface) as well as the *time spent* in each task. The result of the log analysis is reported in Table 3.

To analyze behavior difference among treatments, we performed Wilcoxon Rank Sum and Signed Rank Test on log activity variables. The normality assumption did not hold in our analysis. In the text similarly group, there was a significant difference in the number of clicks for *Sim1* (M=11.16, SD=1.68) and *Sim1+* (M=19.22, SD=5.74) interface; W(18)=9.5, p < 0.01. We did not find a significant effect on the time spent, but the time variance of *Sim1+* was smaller.

**Table 5: NASA-TLX Survey Analysis**

| Sim1 | | Sim1+ | Sim2 | Sim2+ | Sim3 | Sim3++ |
|---|---|---|---|---|---|---|
| Variable | M (SE) | M (SE) | M(SE) | M (SE) | M (SE) | M (SE) |
| TLX1 | 2.77 | 2.55 | 2.00 | 2.33 | 2.22 | 2.22 |
| | (1.16) | (1.38) | (1.32) | (1.28) | (1.35) | (1.55) |
| TLX4 | 3.66 | **4.22** | 4.52 | 4.22 | 4.27 | 4.44 |
| | (0.76) | **(0.94)*** | (4.22) | (0.73) | (0.82) | (0.70) |
| TLX5 | 2.61 | 2.16 | 1.52 | 1.88 | 1.66 | 1.88 |
| | (1.09) | (1.38) | (0.79) | (0.90) | (1.02) | (1.32) |
| TLX6 | 2.05 | 1.77 | 1.23 | 1.50 | 1.50 | 1.16 |
| | (1.16) | (1.06) | (0.75) | (0.70) | (0.92) | (0.51) |

In the topic similarly group, we also found a significant difference in the number of clicks for *Sim2* (M=5.17, SD=0.39) and *Sim2+* (M=10.37, SD=2.14) interface; W(18)=0, p < 0.01. We did not find a significant difference for the time spent, but the subjects took a longer time at average to complete the sorting task while using the *Sim2+* interface. In the item similarly group, we did not find significant differences for clicks or time spent,

but we can observe that the enhanced interface (*Sim3+*) required on average slightly more clicks and time to complete the task.

In general, we found adding additional visual component resulted in more clicks and time spent to complete the sorting tasks. The combined explanation interface produced more user interactions than a single explanation. Furthermore, the tasks were demanding to the subjects since they spent at average 5 to 6 minutes to complete the sorting. The subjects faced more difficulties while interacting with the *Sim1* interfaces, which took the longest time and the most clicks to complete the task.

### 5.3 Survey difference among visualizations

The survey feedback was collected after performing each of the recommendation-sorting tasks. The subjects were asked to answer questions for nineteen explanation factors and four NASA- TLX questions. We summarized the factor questions into seven exploratory goals (shown in Table 1), e.g., the goal of *Transparency (TP)* was consisted by the average score of Q1, Q2, Q3, Q4, and Q5, etc. The results of task survey and NASA-TLX survey were reported in Table 4 and Table 5, respectively. To analyze behavior difference among treatments, we performed Wilcoxon Rank Sum and Signed Rank Test on log activity variables. The normality assumption did not hold in our analysis.

In the text similarity group, the enhanced interface (*Sim1+*) received significantly higher ratings in the goal of *Transparency (TP)*; $W(18)=97.5$, $p < 0.05$, *Scrutability (SC)*; $W(18)=54$, $p < 0.01$, *Trust (TS)*; $W(18)=96.5$, $p < 0.05$, and *Effectiveness (ET)*; $W(18)=73.5$, $p < 0.01$. The result indicated that the baseline explanation interface (*Sim1*) benefited from the additional explaination component. We further analyzed the result of NASA-TLX survey and found similar effects. The subjects perceived significantly better performance (*TLX4*) in accomplishing the sorting task. We did not find significant differences in other questions, however, the users reported lower mental demand (*TLX1*), effort (*TLX5*) and frustration (*TLX6*) while interacting with the enhanced explanation interface (*Sim1+*).

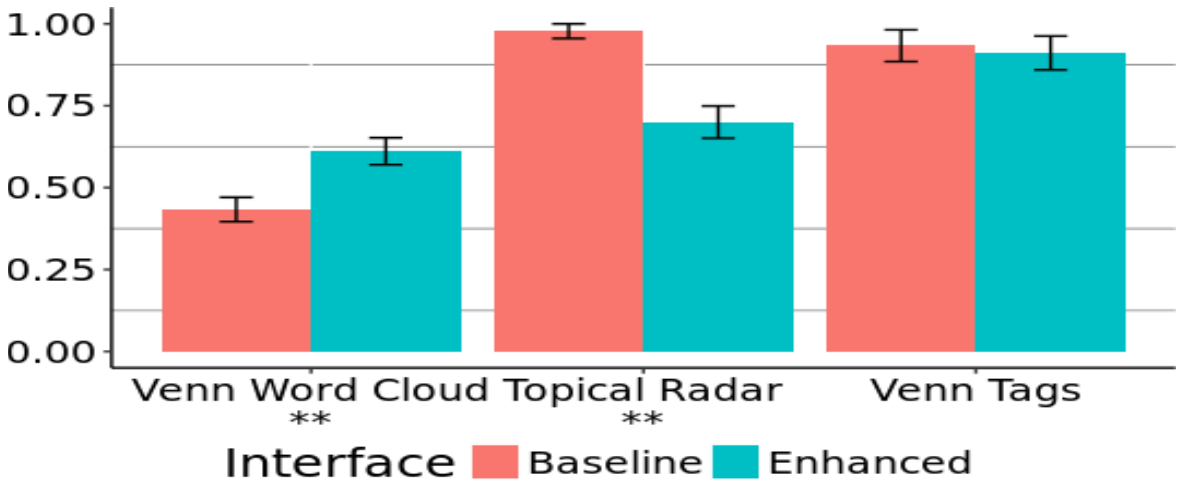


**Figure 2: The correct rate of the recommendation-sorting tasks. Statistical significance level: (**) p< 0.01; (*) p< 0.05)**

In the topic similarly group, we did not find any significant differences in explanation goals. However, we can still observe the similar improving tendency when adding a visual component to the baseline interface. The subjects perception increased on average for almost all explanation goals, except *Persuasiveness (PE)*. The result hints that the additional explaining component might improve the explainability of the baseline interface (*Sim2*). Interesting, in the survey of mental questions, we found the enhanced interface led on average to higher mental demand (*TLX1*), lower performance (*TLX4*), higher efforts (*TLX5*) and higher frustration (*TLX5*), however none of the differences were significant.

We did not find any significant differences in the item similarly group, but observed the same tendency of improved user perception when adding a visual component to the baseline interface. The subjects average perception increased for almost all explanation goals, except the goal of Trust (TS). The result indicated that the explainability of baseline interface (*Sim3*) could be improved by adding a paper or user list to the Venn Tag interface. The survey of mental questions provided an interesting finding. We found the subjects perceived comparable mental demand (*TLX1*), higher performance (*TLX4*), higher efforts (*TLX5*) and lower frustration (*TLX5*). That is, the subjects did not feel higher mental demand, yet adding an extra list still made them perceive the sorting task as harder to accomplish.

In general, we found that adding a visual component lead to a higher user perception score in the explanation goals. However, the improvement in the explanation goals did not guarantee a better user mental model. We found the interface *Sim1* was benefited the most by the additional visual component, either in user perception or mental survey. Adding a visual component to *Sim2* improved user perception but impaired the user mental model. In the interface *Sim3*, adding the extra component improved the user perception while maintaining a comparable user mental model.

### 5.4 Sorting difficulty among visualizations

In addition to the subjective feedback, we are interested in the question of how do the interfaces help the user to compare (sort) the relevance across recommendations. In each interface, we generated five recommendations using the associated recommendation model with a sample scholar profile. We then asked the subjects to sort the relevance among the five given recommendations and compared the answer with the ground truth. We used the *correct rate* to define the *sorting difficulty* among the explanation interfaces. It was an essential metric of *performance* when the user adopted the explanations interfaces in the exploring recommendations. The result was reported in Figure 2.

In the text similarly group, there was a significant difference in the correct rate for *Sim1* (M=0.43, SD=0.15) and *Sim1+* (M=0.61, SD=0.17) interface; $W(18)=75.5$, $p < 0.01$. The result was surprising to show the subjects achieve only 43% correct rate when attempting to sort the given five recommendation with relevance. In this case, adding visual component can be pretty helpful in assisting the sub- jects to complete the sorting task. However, a 61% correct rate may not be considered as an effective explanation interface, in particular, when the users have a chance to browse multiple recommendations and compare the explanations. The inconsistency would hurt the user trust and satisfaction to the explanation interfaces.

In the topic similarly group, there was a significant difference in the correct rate for *Sim2* (M=0.97, SD=0.09) and *Sim2+* (M=0.70, SD=0.20) interface; $W(18)=286$, $p < 0.001$. We found adding extra visual component impaired the judgment on sorting the

recommendation relevance. In the baseline interface (*Sim2*), the subjects can achieve a 97% correct rate, which is strong evidence to support the explanation interface did help the users to sort the recommendation relevance. However, when adding the extra two topical word clouds, we found the correct rate was significantly decreased to 70%, which indicated the users might be "mislead" by the extra information. The result implied that adding the extra visual component can mis- inform the user, although the explanation interface was preferred and received higher user perception ratings by the user.

We did not find a significant difference in the correct rate, in the item similarly group. Both of the interfaces helped the user to achieve a high correct rate (>90%): *Sim3* (M=0.93, SD=0.20) and *Sim3+* (M=0.91, SD=0.21). The result implied adding an extra list to the Venn Tag diagram may not impair or improve the user inspectability (performance) of sorting the recommendations.

### 5.5 Relations between survey and log variables

To better understand the relationship between the survey, log activities and sorting result. We aggregated the variables in all three tasks (N=54). We then performed a correlation (using *Pearson's r*) analysis between task survey items and log variable revealed some interesting associations. The result was reported in Table 6. In general, when subjects did more mouse click activities, the recommendation- sorting correct rate was decreased (*Correct Rate*, r=-0.44, p<0.01) and the subjects will feel more frustrated (*TLX6*, r=0.20, p<0.05). The mouse click means spent more time (*Time*, 0.15, p=0.12) in completing the tasks. The longer time of completion negatively correlated to all explanation goals, e.g., lower the user perception in system transparency (*Transparency*, r= -0.21, p<0.05).

The better inspectability means the subjects can correctly sort the recommendation by relevance. We found the subjects can better understand (*Scrutability*, r=0.21, p<0.05), be convinced by (*Persuasiveness*, r=0.20, p<0.05) and be satisfied (*Satisfaction*, r=0.25, p<0.01) the explanation interface more when they can achieve high correct rate of recommendation-sorting task. Furthermore, the subjects tended to feel less mental demand (*TLX1*, r=-0.22, p<0.05), less effort (*TLX5*, r=-0.39, p<0.01), less frustration (*TLX6*, r=-0.24, p<0.01) but feel more confident in answering the

sorting question (*TLX4*, r=0.43, p<0.01).

**Table 6: Correlation Analysis (N=54)**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Correct Rate | 1.00 | | | | | | | | | | | | | |
| 2. Clicks | **-0.44**** | 1.00 | | | | | | | | | | | | |
| 3. Time | -0.04 | 0.15 | 1.00 | | | | | | | | | | | |
| 4. Transparency | 0.10 | 0.03 | **-0.21*** | 1.00 | | | | | | | | | | |
| 5. Scrutability | **0.21*** | 0.06 | -0.03 | **0.58**** | 1.00 | | | | | | | | | |
| 6. Trust | 0.14 | -0.01 | -0.11 | **0.83**** | **0.67**** | 1.00 | | | | | | | | |
| 7. Persuasiveness | **0.20*** | -0.09 | 0.11 | **0.49**** | **0.39**** | **0.58**** | 1.00 | | | | | | | |
| 8. Effectiveness | 0.07 | 0.02 | -0.09 | **0.77**** | **0.58**** | **0.93**** | **0.58**** | 1.00 | | | | | | |
| 9. Efficiency | -0.03 | 0.15 | -0.12 | **0.47**** | **0.30**** | **0.51**** | **0.31**** | **0.48**** | 1.00 | | | | | |
| 10. Satisfaction | **0.25**** | -0.10 | -0.13 | **0.65**** | **0.45**** | **0.59**** | **0.68**** | **0.51**** | **0.43**** | 1.00 | | | | |
| 11. TLX1 | **-0.22*** | 0.07 | -0.14 | -0.04 | -0.10 | -0.09 | **-0.29**** | -0.17 | 0.00 | -0.05 | 1.00 | | | |
| 12. TLX4 | **0.43**** | -0.06 | -0.05 | 0.13 | **0.26**** | 0.16 | **0.41**** | 0.17 | 0.00 | 0.17 | **-0.49**** | 1.00 | | |
| 13. TLX5 | **-0.39**** | 0.15 | 0.06 | -0.14 | **-0.27**** | **-0.23*** | **-0.40**** | **-0.29**** | -0.01 | -0.15 | **0.66**** | **-0.72**** | 1.00 | |
| 14. TLX6 | **-0.24*** | **0.20*** | 0.15 | -0.05 | -0.11 | -0.16 | **0.35**** | **-0.24*** | -0.02 | -0.13 | **0.44**** | **-0.56**** | **0.65**** | 1.00 |

We also found high internal consistency among all seven ex- planation goals, which implied the post-experiment survey was reliable. The goal of transparency, trust and effectiveness were highly correlated with each other, which was reasonable because they shared one common factor (the Q5 in Table 1). That is, the correlation analysis suggested that if we can provide an explanation interface with high transparency rating, then we can assume the user may tend to trust and feel the effects in the recommendations.

Higher user perception in the goal of scrutability (r=-0.27, p<0.01), trust (r=-0.23, p<0.05), persuasiveness (r=-0.40, p<0.01), and effectiveness (r=-0.29, p<0.01) can reduce the storing difficulty (*TLX5*). Since the mental variable of *TLX5* and *TLX6* were highly correlated with each other (r=0.65, p<0.01), it was reasonable to expect if one explanation goal was negatively correlated with *TLX5* then it should matintain the same pattern with *TLX6*, e.g., between *Effectiveness* and *TLX6*. However, in the explanation goal of *persuasiveness*, we found a positive correlation with *TLX6* (r=0.35, p<0.01), i.e., when the recommendations were very persuasive, the user tend to frustrate more in completing the sorting tasks. We believe this is due to the explanation interface required the users to inspect more details (i.e., the Q14 in Table 1), which leaded to a higher cognitive load.

## 6 CONCLUSIONS

In this paper, we presented two user studies of explanation inter- faces for three similarity-based recommendation models. In study 1, we compared 15 explanation interfaces (twelve visual explanations and three text-based explanations) through nineteen explanation factors. The experiment results suggested that participants preferred visual explanation interfaces over text-based explanation interface. We selected top-rated interfaces to explain the recommendation model, i.e., *E1-4 Venn Word Cloud* to explain text similarity (**Sim1**), *E2-4 Topical Radar* to explain topic similarity (**Sim2**) and *E3-4: Venn Tags* to explain item similarity (**Sim3**). Based on the post-stage user interview, we further proposed *enhanced* visual component to each explanation interface.

In study 2, we conducted a performance-focused evaluation of six explanation interfaces. For each model, we compared the top- rated design (**baseline**) with a combination of top and second-rated interfaces (**enhanced**). We expected that the complementary nature of the top designs could make their combination even stronger than the top choice alone. We found, however, that adding another visual component may result in increasing the cognitive overload and even creating a mental conflict. The findings were varied of each recommendation models: in the group of text similarity, we found adding new visual component (*Two-way Bar Chart*) to the original explanation interfaces significantly improves user performance. However, in the group of topic similarity, we found that adding new visual component (*Word Clouds*) might impair the user perception and performance of the recommendation-sorting task. In the group of item similarity, the extra explanation (list) did not change the user perception or performance scores.

Based on the outcome of two user studies, we found the proposed explanation interfaces did reach the explanation goals. The result of task survey suggested that adding a visual component (enhanced explanation interface) might contribute to a higher user perception score in the explanation goals. However, the improved explanation goals did not guarantee a better user mental model, based on the index of NASA-TLX. The result of recommendation-sorting tasks further suggested the inspectability (performance) can be improved by adding the extra visual component,

but the user-preferred interface may not guarantee the same level of performance. Finally, we introduced a correlation analysis to discuss the relationships between survey and user behavioral variables.

There are several limitations of the presented work. First, the scale of conducted studies was small. Larger-scale studies are needed for more definitive conclusions. Second, user rating and post-stage question ordering are not normalized to control the potential bias. Third, we do not consider the user personality that may influence the user interaction. Fourth, the recommendations were generated for the same sample system user rather than for subjects themselves. All these issues will be addressed in our future work through a larger-scale, lab controlled study.

## REFERENCES


1. 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.
2. 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.
3. Cecilia di Sciascio, Peter Brusilovsky, and Eduardo Veas. 2018. A Study on User-Controllable Social Exploratory Search. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, NY, USA, 353–364. https://doi.org/10.1145/3172944.3172986
4. Jianguang Du, Jing Jiang, Dandan Song, and Lejian Liao. 2015. Topic modeling with document relative similarities. IJCAI.
5. Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces*. ACM, 211–223.
6. Gerhard Friedrich and Markus Zanker. 2011. A taxonomy for generating explanations in recommender systems. *AI Magazine* 32, 3 (2011), 90–98.
7. Fatih Gedikli, Mouzhi Ge, and Dietmar Jannach. 2011. Understanding recommen-

dations by reading the clouds. In *International Conference on Electronic Commerce and Web Technologies*. Springer, 196–208.

8. Wael H Gomaa and Aly A Fahmy. 2013. A survey of text similarity approaches. *International Journal of Computer Applications* 68, 13 (2013), 13–18.
9. Julio Guerra-Hollstein, Jordan Barria-Pineda, Christian D Schunn, Susan Bull, and Peter Brusilovsky. 2017. Fine-Grained Open Learner Models: Complexity Versus Support. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 41–49.
10. Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications* 56 (2016), 9–27.
11. 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.
12. 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.
13. Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User preferences for hybrid explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 84–88.
14. Danielle Lee and Peter Brusilovsky. 2017. How to Measure Information Similarity in Online Social Networks: A Case Study of Citeulike. *Information Sciences* 418-419 (2017), 46–60.
15. Julian McAuley and Jure Leskovec. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*. ACM, 165–172.
16. Conference Navigator. 2018. Paper Tuner. http://halley.exp.sis.pitt.edu/cn3/portalindex.php
17. Alexis Papadimitriou, Panagiotis Symeonidis, and Yannis Manolopoulos. 2012. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery* 24, 3 (2012),

555–583.

18. Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.
19. Pearl Pu and Li Chen. 2007. Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems* 20, 6 (2007), 542–556.
20. Amit Sharma and Dan Cosley. 2013. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 1133–1144.
21. Julia Silge and David Robinson. 2016. tidytext: Text mining and analysis using tidy data principles in r. *The Journal of Open Source Software* 1, 3 (2016), 37.
22. Kirsten Swearingen and Rashmi Sinha. 2001. Beyond algorithms: An HCI perspective on recommender systems. In *ACM SIGIR 2001 Workshop on Recommender Systems*, Vol. 13. Citeseer, 1–11.
23. Nava Tintarev and Judith Masthoff. 2011. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*. Springer, 479–510.
24. 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.
25. Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*. Springer, 353–382.
26. Chun-Hua Tsai and Peter Brusilovsky. 2017. 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.
27. Chun-Hua Tsai and Peter Brusilovsky. 2018. Beyond the Ranked List: User-Driven Exploration and Diversification of Social Recommendation. In *23rd International Conference on Intelligent User Interfaces*. ACM, 239–250.
28. Chun-Hua Tsai and Peter Brusilovsky. 2018. Explaining Social Recommendations to Casual Users: Design Principles and Opportunities. In *Proceedings of the 23rd*

*International Conference on Intelligent User Interfaces Companion*. ACM, 59.

29. Chun-Hua Tsai and Peter Brusilovsky. 2019. Designing Explanation Interfaces for Transparency and Beyond. In *Workshop on Intelligent User Interfaces for Algorithmic Transparency in Emerging Technologies*.
30. Chun-Hua Tsai and Peter Brusilovsky. 2019. Explaining Recommendations in an Interactive Hybrid Social Recommender. In *Proceedings of the 2019 Conference on Intelligent User Interface*. ACM, 1–12.
31. Chun-Hua Tsai and Peter Brusilovsky. 2019. Exploring Social Recommendations with Visual Diversity-Promoting Interfaces. *TiiS* 1, 1 (2019), 1–1.
32. Jesse Vig, Shilad Sen, and John Riedl. 2009. Tagsplanations: explaining recommendations using tags. In *Proceedings of the 14th international conference on Intelligent user interfaces*. ACM, 47–56.
33. Yao Wu and Martin Ester. 2015. Flame: A probabilistic model combining aspect based opinion mining and collaborative filtering. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 199–208.
34. Qian Zhao, Gediminas Adomavicius, F Maxwell Harper, Martijn C Willemsen, and Joseph A Konstan. 2017. Toward Better Interactions in Recommender Systems: Cycling and Serpentining Approaches for Top-N Item Lists.. In *CSCW*. 1444–1453.
35. Zoomdata. 2018. Real-time Interactive Zoomdata Wordcloud. https://visual.ly/community/interactive-graphic/social-media/ real-time-interactive-zoomdata-wordcloud