# The Dark Matter of Serendipity in Recommender Systems

Denis Kotkov
Department of Computer Science
University of Helsinki
Helsinki, Finland
denis.kotkov@helsinki.fi

Alan Medlar
Department of Computer Science
University of Helsinki
Helsinki, Finland
alan.j.medlar@helsinki.fi

Triin Kask
Metatellus
Tallinn, Estonia
triin.kask@metatellus.com

Dorota Głowacka
Department of Computer Science
University of Helsinki
Helsinki, Finland
dorota.glowacka@helsinki.fi

## ABSTRACT

Serendipity has been recognized as a valuable property of recommender systems. While there is a lack of consensus on the precise definition of serendipity, it is often conceptualized in terms of the relevance, novelty and unexpectedness of recommendations. However, the common understanding and original meaning of serendipity is conceptually broader, requiring serendipitous encounters to be neither novel nor unexpected. Recent work has highlighted the various ways in which serendipity can manifest, leading to a more generalized definition of serendipity. In this paper, we conducted an observational study where we collected 2002 survey responses from 397 users of an online article recommender system. In our study, we found a significant proportion of serendipitous recommendations were missed by the conventional definitions used in the recommender systems research literature, exposing the "dark matter" of serendipity that has been overlooked in prior studies. Interestingly, users' opinions of which articles should be considered serendipitous did not strongly align with any of the definitions investigated. Furthermore, despite several user behaviors being significantly associated with a majority of definitions of serendipity, the overall goodness of fit was very low. Our findings highlight the issues of evaluating serendipity in recommender systems and the challenge of reconciling serendipity with user expectations.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → **Field studies**.

## KEYWORDS

serendipity, field study, recommender systems

**ACM Reference Format:**
Denis Kotkov, Alan Medlar, Triin Kask, and Dorota Głowacka. 2024. The Dark Matter of Serendipity in Recommender Systems. In *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval*

## 1 INTRODUCTION

Recommender systems suggest items, such as movies and books, that are predicted to be of interest to users [28]. By design, recommender systems tend to recommend either popular items or items associated with users' consumption patterns, which means that users may already be familiar with recommendations or otherwise capable of finding them without a recommender system [17]. For these reasons, recommender systems are frequently evaluated on the basis of so-called "beyond accuracy" metrics, such as novelty, diversity, and serendipity, in addition to utility-based metrics like precision and nDCG [12]. In the case of serendipity, the recommender systems research community has operationalized the concept as a combination of relevance, novelty and unexpectedness, where relevance refers to whether an item was beneficial to the user, and novelty indicates whether the item was unfamiliar [14, 17, 34]. Unexpectedness has numerous definitions, the most common being that the user does not think they would have found the recommended item on their own [2, 8–10, 18, 36]. We argue that this conceptualisation is unnecessarily narrow and excludes many recommendations that should be considered serendipitous.

The term serendipity was coined in 1754 by Horace Walpole in reference to the Persian fairy tale The Three Princes of Serendip. In the story, the three princes were exploring the world and "*making discoveries, by accidents & sagacity, of things which they were not in quest of*" [27]. This usage is consistent with modern dictionary definitions. For example, Merriam-Webster[1] defines serendipity as "*the faculty or phenomenon of finding valuable or agreeable things not sought for*". While there is broad agreement between these definitions, they allow for multiple interpretations and cover a wide range of phenomena. Indeed, the sociologist Robert K. Merton spent decades collecting mentions of the term "serendipity" from magazines, newspapers, and journals to characterise its use context [26]. Merton's archive was subsequently used to create a typology of serendipity that encompasses four different types of scenario that can be considered serendipitous: Walpolian, Mertonian, Bushian and Stephanian serendipity [39]. Each type of serendipity is defined by (i) whether the discoverer had a specific goal in mind

---

[1] https://www.merriam-webster.com/dictionary/serendipity

and (ii) what type of solution the discovery lead to. Following [15], we collectively refer to these types of serendipity as **generalized serendipity** to distinguish them from how serendipity has been defined in the recommender systems literature, which we refer to as **RecSys serendipity**.

The focus on problem-solving in generalized serendipity makes it broader in scope than the definitions used in recommender systems and potentially easier to apply in experimental studies. Indeed, the RecSys definition has numerous shortcomings that complicate its use. First, while the most common definition of serendipity in recommender systems requires relevance, novelty and unexpectedness, a recent review showed that numerous studies have omitted one or more of these components [17]. Furthermore, even studies that define serendipity in the same way can differ in terms of how each component is measured [14]. This lack of consensus makes it difficult to compare results across studies and creates additional "researcher degrees-of-freedom" [32] that can lead to misuse. Second, despite being considered important in the RecSys definition of serendipity, there is no mention elsewhere of serendipitous encounters needing to be either novel or unexpected [15]. This suggests there are numerous recommendations that should be considered serendipitous (i.e. according to generalized serendipity), but are not due to the strictness of the RecSys definition. Last, when evaluating recommender systems based on user feedback regarding serendipity, it may be complicated by users' colloquial understanding of the term [15, 30]. Correspondingly, a recommender system designed to increase serendipitous discoveries may not meet users' expectations if their understanding of serendipity is misaligned with that of the system designers. For concision, we refer to this subjective definition of serendipity based on users' opinions as **user serendipity**.

In this paper, we investigate serendipity in the online article recommender system, Soulie[2]. We designed an observational study to identify serendipitous article recommendations using per-article surveys. We collected 2002 survey responses from 397 users and classified whether each article was serendipitous according to each definition of serendipity. First, we wanted to understand how well the definitions of serendipity used in recommender systems research and users' subjective opinions aligned with generalized serendipity. Next, as we observed considerable overlap between RecSys and generalized serendipity, we investigated the relative importance of relevance, novelty and unexpectedness in generalized serendipity. Last, we analyzed users' behavior in our system, including interface interactions (whether articles were "liked", added to "favourites" or marked as "read later"), reading progress and time spent in the system, to understand how these variables relate to different definitions of serendipity. In summary, we ask the following research questions:

- **RQ1 Alignment with Generalized and User Serendipity:** To what extent do the various definitions of RecSys serendipity capture items considered serendipitous under generalized serendipity? Furthermore, to what extent do users' opinions agree with generalized and RecSys serendipity?
- **RQ2 Relative Importance of RecSys Serendipity Components:** What is the relative importance of relevance, novelty

and unexpectedness in both generalized and user definitions of serendipity?
- **RQ3 Associations with User Behavior:** What aspects of user behavior are associated with generalized serendipity, user serendipity and the various definitions of RecSys serendipity?

Our study found that the precision of using RecSys and user serendipity to predict the set of generalized serendipity articles was 0.64, showing a moderate alignment between definitions. Recall was generally lower, with scores ranging from 0.35-0.71. Indeed, the most common definition of RecSys serendipity (i.e. relevant, novel and unexpected [1, 2, 14, 17, 25, 35]) had the lowest recall, exposing what we refer to as the "dark matter" of serendipity; serendipitous recommendations that would have been overlooked in prior studies. Conversely, while user serendipity aligned closest with RecSys serendipity, it still only had a precision of 0.57. Despite these findings, we found that the components of RecSys serendipity were significantly associated with both generalized and user serendipity[3]. However, model fit was low: 0.12 and 0.31 for generalized and user serendipity, respectively. Similarly, despite user behaviors being significantly associated with a majority of serendipity definitions, the model fit was generally very low and did not exceed 0.11. These results suggest a need to identify better explanatory factors in future work.

## 2 BACKGROUND

In background, we cover definitions, assessment and predictors of serendipity in recommender systems.

### 2.1 Definitions of Serendipity

Here, we describe the various definitions of RecSys serendipity as well as the definitions of generalized and user serendipity.

*2.1.1 RecSys Serendipity:* The first mention of serendipity in recommender systems was in 2004 by Herlocker et al., who stated that "*A serendipitous recommendation helps the user find a surprisingly interesting item he might not have otherwise discovered*" [9]. While numerous definitions of serendipity have been used since then, a majority require serendipitous items to be some combination of relevant, novel and/or unexpected [14, 17]. Relevance indicates user interest in an item, which can be captured explicitly via a survey question or implicitly if, for example, users consume the item [14]. Novel items have been defined as those the user is unfamiliar with and as those not thought of at the moment of recommendation [14]. Unexpectedness has also been defined in several ways, such as being dissimilar to what the user usually consumes [11, 14, 17, 40] and when the user enjoyed consuming an item they did not expect to enjoy [2, 14]. In general, there is no consensus on which components are integral to serendipity nor how each component should be measured [14, 15, 17, 42].

In this paper, we consider all combinations of relevance, novelty and unexpectedness as variations of RecSys serendipity as each combination has been used in prior studies (with the exception of relevance, that we include for comparison) [11, 13, 17, 21, 23, 41]. For relevance, we used whether the user liked a given item

---

[3]Relevance, novelty and unexpectedness were significantly associated with user serendipity, but only relevance and novelty were significantly associated with generalized serendipity.

and, for novelty, we required that an item be unfamiliar to the user [14]. Lastly, we used one of the most common definitions for unexpectedness: items the user thinks they would not have found without the recommender system [2, 8–10, 18, 36].

*2.1.2 Generalized Serendipity:* Generalized serendipity was introduced by Kotkov et al. as a broader definition of serendipity for recommender systems than what was previously used [15]. Generalized serendipity is based on Yaqub's typology that divides serendipitous encounters into four categories: Walpolian, Mertonian, Bushian and Stephanian serendipity [39]. Walpolian serendipity is the beneficial discovery of things not looked for by the discoverer. Mertonian serendipity extends the previous category by including beneficial discoveries looked for by the discoverer, but found via an unexpected route. Bushian serendipity further extends the definition by including things found without a specific goal in mind. Lastly, Stephanian serendipity also covers beneficial discoveries that help to solve problems that occur after the discovery was made [39].

Kotkov et al. operationalized Yaqub's work by considering users' goals with a recommender system [15]. Under generalized serendipity, an item is serendipitous if it helps the user to achieve at least one goal different from those they set out to achieve with the recommender system. For example, if a user wants to find an article to read during breakfast, but finds an article to share with a friend, then finding that article is considered serendipitous. If the user does not have any goals when using the system, then any goals achieved are considered serendipitous. The formal definition of generalized serendipity is as follows: "*Assume that all goals of the target user at moment in time $t_k$ is the set $G_{t_k}$. Meanwhile, the time-line is discrete $T = \{t_1, t_2, ..., t_n\}$, such that at each moment user goals are different compared to the goals in adjacent moments in time: $(G_{t_{k-1}} \neq G_{t_k}) \wedge (G_{t_k} \neq G_{t_{k+1}})$. Goals that the target user wants to achieve in a recommender system at $t_k$ are $R_{t_k} \subseteq G_{t_k}$, while $R_{t_k} = \emptyset$ if the user does not interact with the system or has no specific goals during the interaction. An item recommended by the system to the target user is serendipitous if it helps them to achieve any goals from the set $S_{t_k} = \bigcup_{j=k}^{n} G_{t_j} \setminus R_{t_j}$*". Generalized serendipity is similar to definitions used in other disciplines, such as information interaction [20] and the social sciences [15]. The key difference between generalized and RecSys serendipity is that generalized serendipity is based on the difference between what the user looks for and what they find, whereas RecSys serendipity is based on the user's interest in and awareness of the item prior to finding it.

*2.1.3 User Serendipity:* User serendipity is based on each users' subjective understanding of the term serendipity [15]. While user serendipity is ill-defined, it is an important point of comparison to understand users' perspectives and to interpret user feedback. Indeed, serendipity has been recognized as one of the most difficult words to translate into different languages [22] and is often misused in daily life [29].

## 2.2 Assessment of Serendipity

Serendipity is generally assessed in recommender systems using surveys tailored to a specific definition of serendipity [17].

*2.2.1 RecSys Serendipity:* Serendipity studies in recommender systems tend to ask users to interact with a set of items and use survey questions to identify whether those items should be considered serendipitous. For example, Kotkov et al. conducted an experiment in MovieLens, a movie recommender system, where the authors picked movies that were relevant to users based on their ratings and asked them to indicate whether each movie was novel or unexpected according to multiple definitions of each term. If a user indicated that a relevant movie was both novel and unexpected, then it was labeled serendipitous according to the RecSys definition [14].

As there is no consensus on precisely what constitutes RecSys serendipity, researchers have used various questions to assess relevance, novelty and unexpectedness. For example, Taramigkou et al. only asked about relevance and unexpectedness: *"Did you find artists you wouldn't have found easily on your own and which you would like to listen to from now on?"* [36]. Zhang et al. essentially defined serendipity as unexpectedness (i.e. no relevance or novelty) using the following semantic differential scale: *"Exactly what I listen to normally... Something I would never have listened to otherwise"* [40]. In contrast, Smets et al. used relevance and a variation of unexpectedness: *"How often do you find yourself pleasantly surprised by the recommended restaurants or bars on these websites and apps?"* [34].

*2.2.2 Generalized Serendipity:* There have been two study designs proposed to investigate generalized serendipity in recommender systems: a laboratory study [33] and a field study [15].

Smets et al. proposed a laboratory study for assessing generalized serendipity [33]. Users would be instructed to use a recommender system to shortlist items that fit specific criteria, such as finding appropriate books for a book discussion club. While looking for these items, users would be provided with another list to save items they found interesting, but were otherwise unrelated to the search task (that they would be forwarded after the study). After completing the task, participants would then be asked to fill out a survey to identify serendipitous recommendations in the non-task related list.

Kotkov et al. proposed a field study for measuring generalized serendipity [15]. In the experiment, participants would use the recommender system in their daily lives, but would regularly be asked to share their current goals, i.e. why they are using the system. After consuming each item, participants would be asked to indicate what goal the item helped them to achieve. An item would be considered to be generalized serendipitous if it helped a participant to achieve any goals different from those they set out to achieve. This paper implements and further develops this study design.

*2.2.3 User Serendipity:* Similar to RecSys serendipity, user serendipity is assessed by asking participants a survey question after they have interacted with a recommender system. This question takes the form of a direct question that contains the word serendipity, but provides no further explaination with regards to specific criteria. For example, Said et al. simply asked the question: *"Are the recommendations serendipitous?"* [30].

(a) Hub page     (b) Article viewing     (c) Article survey     (d) Selecting interests     (e) Goal survey     (f) Collections (favorites, read later, history)
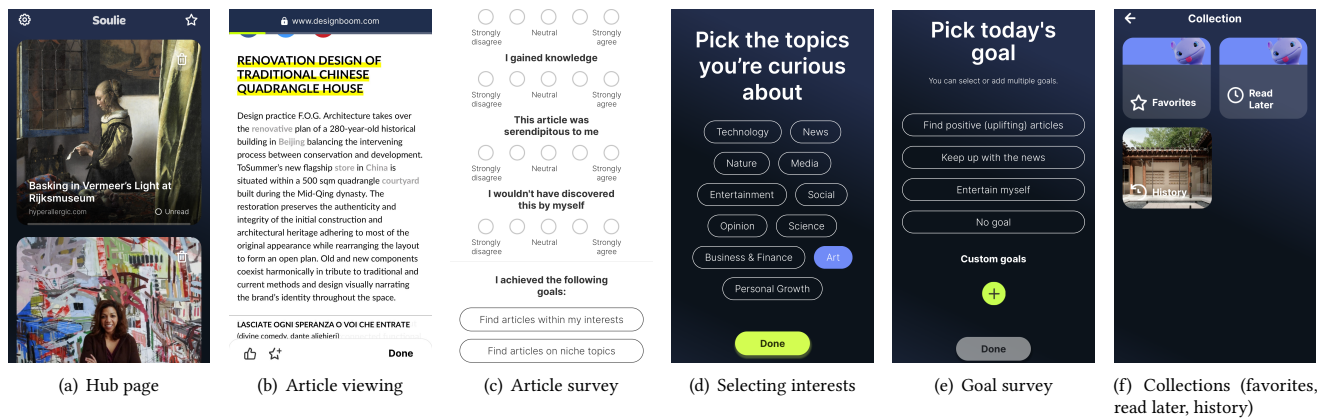
**Figure 1: User interface of the article recommender system.**

## 2.3 Predictors of Serendipity

Recently, several user studies have identified predictors of RecSys serendipity. Kotkov et al. studied serendipity in MovieLens and found it to be correlated with preference broadening. Furthermore, ratings predicted by MovieLens, item popularity, content-based and collaborative similarity to users' profiles were all predictors of serendipitous recommendations [14]. Chen et al. conducted a survey on serendipity in e-commerce and showed it to be correlated with user satisfaction, purchase intention and timeliness [6].

Wang and Chen [37] compared serendipity in the movie [14] and e-commerce domains [6]. They found that lower item popularity and smaller user profiles (items consumed in the past) result in higher chances of serendipity in both domains. However, other factors were domain-specific: diversity of items consumed was correlated with serendipity in the movie domain, but anti-correlated in e-commerce. Whereas, timeliness and collaborative similarity were correlated with serendipity in e-commerce, but anti-correlated in the movie domain. Interestingly, they showed that in e-commerce certain user groups (men, older people) and personality types (high on curiosity, openness to experience, conscientiousness, extraversion, and neuroticism, but low on agreeableness) were more inclined to perceive recommendations as serendipitous.

Lastly, Smets et al. conducted a survey on venues in an urban recommender system. They found that the more frequently users visited venues, the higher the chances of them finding serendipitous venues [34].

## 2.4 Serendipity Outside of Recommender Systems

Serendipity has been widely researched in other disciplines using similar definitions to those investigated in our study. For example, Bao and Yang investigated serendipity in management using a definition similar to one previously used in recommender systems: *"serendipity has two fundamental properties: unexpectedness and value"* [5]. The authors surveyed 353 participants and found that serendipity was one of the core factors that influenced impulsive buying behavior [5]. Sawaizumi et al. introduced the concept of "serendipity cards" in education and used a definition of serendipity

similar to generalized serendipity [31]. Agarwal conducted a literature review on information behavior and suggested that serendipity can result in beneficial discoveries, but also disappointment, depending on the information encountered [4]. Lastly, Makri et al. conducted an observational study of serendipity in information seeking with 45 participants and also used a definition similar to generalized serendipity: *"[w]e therefore define 'coming across information serendipitously' as 'finding useful or potentially useful information unexpectedly – either when not looking for information at all, when looking for information about something else or when looking for information with no particular aim in mind.'"*[20]. The study suggested that serendipity-related information interaction behavior can be observed in a research setting. Considering the similarities between the definitions of serendipity used in other disciplines and those we have investigated in our study, we believe that the issues we highlight could also have implications for the study of serendipity in other contexts.

## 3 METHODOLOGY

In this work, we wanted to directly compare examples of generalized serendipity, the various definitions of serendipity used in the recommender systems literature and participants' own understanding of serendipity. Our study was based on Soulie, an application where participants could browse through recommendations of online articles related to their interests and provide feedback using brief surveys (see Figure 1). The data collected allowed us to assign which, if any, definition of serendipity could be applied to a given article. Additionally, we logged numerous interaction events between participants and the application to analyze which aspects of user behavior are associated with different definitions of serendipity.

### 3.1 Study Design

We designed an observational study based on Soulie, an article recommender system currently under development for Android and iOS. This study was part of the development process.

The application loads articles from the Internet and displays them in an internal web browser (see Figure 1(b)). The articles were

chosen from popular websites, including Pitchfork[4], Ars Technica[5] and Live Science[6]. These websites were crawled each day and new articles added to the database. At the time of writing, the database contains links to 97,583 articles categorized using the following topics: news (76,652), media (7,186), technology (3,123), business and finance (2,596), opinion (1,945), social (1,331), art (1,304), science (1,228), nature (945), personal growth (739) and entertainment (534). Articles can only belong to a single topic. After two weeks, the articles expire and become unavailable for recommendation.

The application was advertised through social media as a tool for providing personalized serendipitous article recommendations for learning useful information and to encourage personal development. Participants were notified that the application was currently under development and were invited to join a mailing list and Slack channel. When an early version of the application was ready, participants were invited to enroll in the study. Participants were informed that the study was dedicated to serendipity and their participation would support further development. All participants were aged 18 or over and gave informed consent for their data to be used for research purposes. The study ran for 16 weeks from December 15, 2022 to April 6, 2023.

## 3.2 Survey Design

To identify serendipitous articles, we asked participants to complete two types of survey throughout the study: goal surveys (1 question) and article surveys (5 questions). The survey questions used a combination of 5-point scales and short free-form text. The survey questions were all derived from the recommender systems literature and are shown in Table 1. We used the minimum number of survey questions to increase response rate and minimize disruptions while users interacted with the system.

*3.2.1 Goal surveys:* The goal survey appeared the first time participants opened the application each day and asked them to state their current goal for using the application (see Figure 1(e)). The goal survey included a list of predefined goals (described below) that participants could add to with their own custom goals. Participants could change their goals at any time, indicate that they had *"No goal"* and set multiple simultaneous goals (excluding *"No goal"*, which could not be combined with any other goals). The goal surveys were only used to identify whether articles could be considered serendipitous according to generalized serendipity.

*3.2.2 Article surveys:* The article surveys were used for measuring each serendipity type and appeared after the participants had read each article (see Figure 1(c)). Participants had the option to skip the article survey and, if they skipped three surveys in a row, were not asked to complete the survey for the rest of the day. If a participant completed an article survey for the same article more than once, then we only considered the most recent response.

*3.2.3 Predefined Goals:* To help participants fill out the surveys quickly, we included a list of predefined goals for both the goal and article surveys. We created the list of predefined goals by asking members of the application's Slack channel to indicate the goals

they would like to achieve with the application and to describe the kinds of article they would like recommended. We received 36 replies and extracted goals using the following procedure [7]. First, we filtered out goals that were inapplicable to our study design (e.g. *"spending less time on the phone"*) or too specific (e.g. *"finding articles on financial freedom"*). Next, we summarized the remaining answers, resulting in the list shown in Figure 2. Following [15], we also added a *"No goal"* option.

*3.2.4 Generalized Serendipity:* We used the two goal-related questions from the goal and article surveys to understand whether a given article was serendipitous according to the definition of generalized serendipity. We compared the goals the participant stated they wanted to achieve with our system with the goals that they stated had been achieved with the recommended article [15]. If the participant achieved at least one goal different from the list of planned goals, then we labeled the article as an example of generalized serendipity. We note that for the purpose of our analysis, we only required each participant to have a consistent interpretation of each goal and did not require different participants to have the same interpretation of a given goal [15].

*3.2.5 RecSys Serendipity:* We included three survey questions related to relevance, novelty and unexpectedness in the article survey to understand whether each article was serendipitous according to any of the previously used definitions of serendipity in the recommender systems research literature (see Table 1). Following [14], an article was considered, for example, relevant, if participants responded with either 4 or 5 on a 5-point scale (corresponding to "agree" and "strongly agree", respectively). If the participant indicated that each component necessary for a given type of RecSys serendipity was true for the current article, then we labeled the article as an example of RecSys serendipity.

*3.2.6 User Serendipity:* The article survey included a single question to understand whether participants considered a given article serendipitous according to their own understanding of the term. If the participant responded with either 4 or 5 on a 5-point scale (corresponding to "agree" and "strongly agree", respectively), then we labeled the article as an example of user serendipity.

## 3.3 Measures

We logged numerous interactions in the application to understand the degree to which user behavior was impacted after encountering serendipitous articles. We used the following variables in our analysis:

- **User interactions:** We modelled whether users interacted with the application after reading an article as a binary variable. The variable was set to 1 if the participant had either "liked" or added the article to their "Favorites" or "Read Later" lists, and 0 otherwise. Among the articles with survey responses, 20% had at least one type of interaction: 17% were liked, 5% were added to "Favorites" and 4% were added to "Read Later". Participants were able to "unlike" an article as well, so we only considered the final status at the end of the study.
- **Reading progress:** We modelled the reading progress of each viewed article as a continuous variable between 0 and 1 indicating the proportion of the web page scrolled through by the

**Table 1: Survey questions and response scales. The 5-point Likert scale corresponds to the options: strongly disagree, disagree, neutral, agree, strongly agree.**

| Type of Serendipity | Question | Response options | Adopted from |
|---|---|---|---|
| Goal survey | | | |
| Generalized serendipity | Pick today's goal | Learn new things, Improve myself, etc. | [15] |
| Article survey | | | |
| Relevance | I liked this article | 5-point Likert scale | [14, 17, 34] |
| Novelty | I gained knowledge | 5-point Likert scale | [14] |
| Unexpectedness | I wouldn't have discovered this by myself | 5-point Likert scale | [2, 8–10, 18, 36] |
| User serendipity | This article was serendipitous to me | 5-point Likert scale | [30] |
| Generalized serendipity | I achieved the following goals | Learn new things, Improve myself, etc. | [15] |
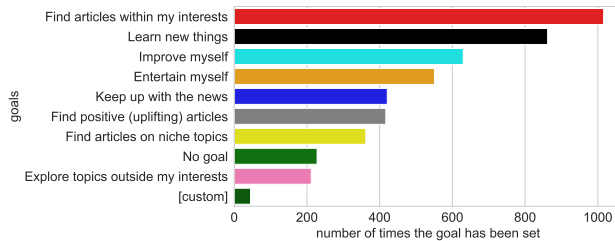


(a) Distribution of goals set by participants to be achieved with the application.
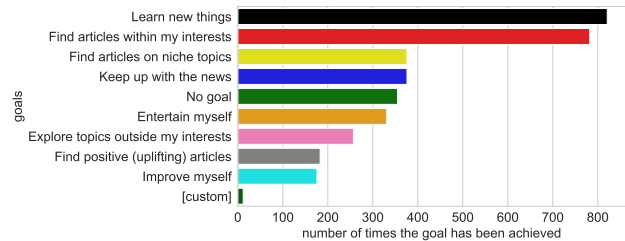
(b) Distribution of goals achieved with an article. *"No goal"* indicates that the article has not helped the participant to achieve any goals.

**Figure 2: Planned and achieved goals. Custom refers to goals added by participants.**
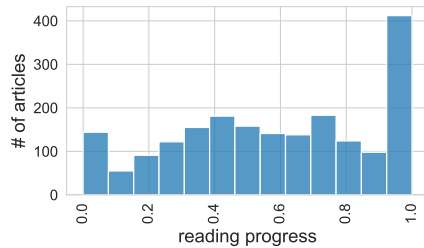


(a) Distribution of read articles per reading progress
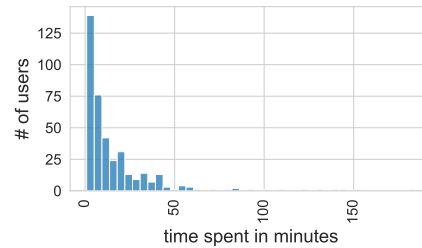
(b) Distribution of time spent in system in minutes per participant. For clarity, we omitted four participants from the graph who spent more than 200 minutes using the application.

**Figure 3: Distributions of user behaviors**

participant. Figure 3(a) shows the most frequently occurring value of reading progress was close to 1. It is worth noting that articles are often followed by online advertisements, so participants may not have scrolled to the end of the page even though they read the whole article.

- **Time spent:** We logged the number minutes each participant spent in the application by calculating the total duration of all sessions. Participants were considered to be in an active session if there was a timestamped interaction within the last 10 minutes. Interactions included, for example, navigation between pages, survey responses and goal updates. Participants spent an average of 20.5 minutes (median 8.4 minutes) in the application

(see Figure 3(b)). We log transformed time spent as it could only take a positive value.

## 3.4 Procedure

As part of the registration process for the study, participants gave informed consent and provided demographic information. Only participants over the age of 18 were allowed to take part in the study. Participants were then given access to the application, which they needed to download and install on their mobile device.

Participants were not given an explicit task to complete other than to use the application as a way to discover articles to read.

We describe how participants could interact with the application during that process:

- **Define interests:** Participants could define their interests during registration and at any point afterwards. Interests were selected from the list of article categories, i.e. news, technology, etc. (see Figure 1(d)). Each participant had to pick at least one topic of interest.
- **Set goal(s):** Participants stated their current goals by answering the goal survey. The goal survey was opened at the start of the first session of the day and could be accessed at any point by the participant to redefine their goals. Goals could be selected from the list of predefined goals or a custom goal added by the participant (see Figure 1(e)).
- **Browse articles:** Participants could browse articles from the hub screen (see Figure 1(a)). The application recommended 15 articles at a time with 12 articles corresponding to the participant's stated interests and 3 articles from topics outside of their interests. Articles were sampled randomly from each category. If participants modified their interests, then the recommended articles were updated accordingly. Participants could select articles for reading or remove an article from the list of recommendations. Articles that were read or removed were replaced by another article from the same category.
- **Read articles:** While reading articles (see Figure 1(b)), participants could "like" or add it to their "Read Later" and "Favorites" lists. After reading an article, participants were asked to complete an article survey (see Figure 1(c)). Articles that participants had read or added to a list were available on the collection screen (see Figure 1(f)).

## 3.5 Participants

We recruited 616 participants for our study (339 male, 218 female, 59 other). Participants were recruited via social media. The median age of participants was 21, but ranged from 18 to 44. As this was an observational study with no explicit task, many participants interacted with the application, but did not complete any goal or article surveys. To answer our research questions, we focused on the subset of 397 participants (207 male, 147 female, 43 other) who provided at least one response to a pair of goal and article surveys. In this subset, participants provided 2002 article survey responses on 1320 unique articles. A majority of participants only rated a single article, while the median was 2 articles and the mean was 5 articles.

## 4 RESULTS

We conducted a quantitative analysis to contrast the sets of articles that met the definitions of generalized, RecSys and user serendipity. In terms of participants' goals, the most popular were *"finding articles within my interests"* and *"learn new things"*, which were also the most frequently achieved goals (see Figure 2). Custom goals added by participants included, for example, *"keep up with economics"*, *"get information to make a decision what to buy/do"* and *"regain focus"*. Based on the responses to the article surveys, we identified 1075 articles for generalized serendipity, 687 articles for RecSys serendipity (using the most common definition: relevant, novel and unexpected) and 789 articles for user serendipity. These

**Table 2: Classification metrics of how well each definition of serendipity predicts the set of generalized serendipity articles. The highest scores for each metric are in bold.**

| Serendipity def. | Precision | Recall | Accuracy |
|---|---|---|---|
| user | **0.64** | 0.40 | 0.59 |
| rel | 0.59 | 0.67 | **0.62** |
| nov | 0.58 | **0.71** | 0.61 |
| unexp | 0.50 | 0.64 | 0.51 |
| rel_nov | 0.61 | 0.53 | 0.61 |
| rel_unexp | 0.61 | 0.43 | 0.59 |
| nov_unexp | 0.59 | 0.47 | 0.58 |
| recsys (rel_nov_unexp) | **0.64** | 0.35 | 0.58 |

sets of articles have substantial overlaps with one another (see Figure 4). For example, around half of the generalized serendipity articles are also considered serendipitous by users (Figure 4(d)), while most of the user serendipity articles belong to at least one component of RecSys serendipity (Figure 4(b)).

## 4.1 Alignment with Generalized and User Serendipity

For RQ1, we wanted to investigate the extent to which the various definitions of RecSys serendipity captured the set of articles considered serendipitous according to generalized and user serendipity. We used the classification evaluation metrics: precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$) and accuracy ($\frac{TP+TN}{TP+FP+TN+FN}$), where TP is true positives, TN is true negatives, FP is false positives and FN is false negatives. To compare ResSys and generalized serendipity, we considered the set of true articles (i.e. TP + TN) to be the 1075 generalized serendipity articles and the set of positive articles (i.e. TP + FP) to be those considered serendipitous according to a given RecSys definition. As the number of articles viewed by each participant was not equal, we report the average precision, recall and accuracy per participant. To investigate user serendipity, we repeated this analysis, but replaced the set of generalized serendipity articles with the set of user serendipity articles.

For concision, we used the following shortcuts to describe each component of RecSys serendipity: *rel* for relevance, *nov* for novelty and *unexp* for unexpectedness. We indicated the intersection between serendipity components with an underscore, e.g. *rel_nov* corresponds to articles that were both relevant and novel. We considered all definitions of RecSys serendipity that have previously appeared in the literature [11, 13, 17, 21, 23, 41] in addition to relevance for comparison.

*4.1.1 Generalized serendipity.* Table 2 shows the precision, recall and accuracy of how well each definition of RecSys serendipity and user serendipity predict the set of generalized serendipity articles. Both user serendipity and rel_nov_unexp had the highest precision (0.64). However, these two definitions also had the lowest recall scores (0.4 and 0.35 for user serendipity and rel_nov_unexp, respectively). These findings are in accordance with our intuitions: RecSys serendipity is conceptually narrow as demonstrated by low recall, but its moderately high precision may still make it appropriate for ensuring the presence of serendipitous items in top-n
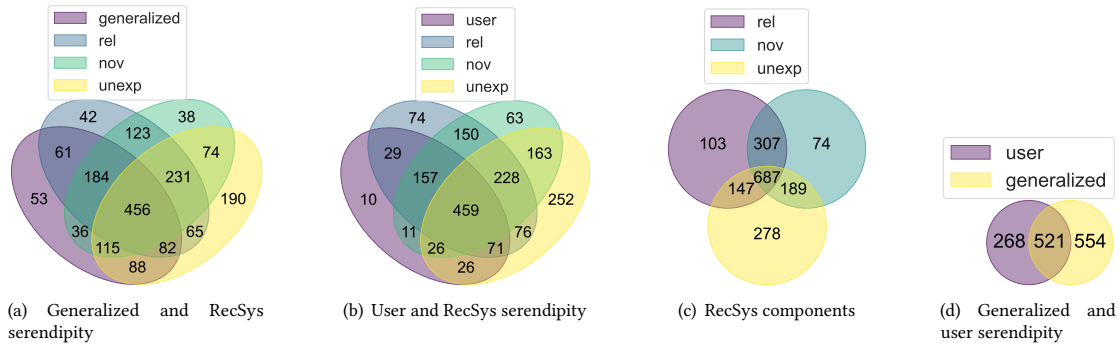
(a) Generalized and RecSys serendipity

(b) User and RecSys serendipity

(c) RecSys components

(d) Generalized and user serendipity

Figure 4: Euler diagrams showing the overlaps between different definitions of serendipity

**Table 3: Classification metrics of how well each definition of serendipity predicts the set of user serendipity articles. The highest scores for each metric are in bold.**

| Serendipity def. | Precision | Recall | Accuracy |
|---|---|---|---|
| generalized | 0.40 | 0.64 | 0.59 |
| rel | 0.46 | **0.86** | 0.67 |
| nov | 0.41 | 0.78 | 0.58 |
| unexp | 0.35 | 0.69 | 0.49 |
| rel_nov | 0.50 | 0.72 | 0.71 |
| rel_unexp | 0.52 | 0.59 | 0.71 |
| nov_unexp | 0.44 | 0.54 | 0.64 |
| recsys (rel_nov_unexp) | **0.57** | 0.49 | **0.73** |

recommender systems. Generalized serendipity appears to be well captured by novelty: it has the highest recall (0.71) and the second highest accuracy (0.61). However, we note that these scores are conditional on participants' decision to read a given article and, therefore, implicitly depend on some aspect of relevance as well. Indeed, relevance had the highest accuracy (0.62), which is interesting because recommender systems are designed to predict which items are relevant to users' interests and this is thought to be antagonistic to serendipity [2, 11, 17].

*4.1.2 User serendipity.* Table 3 shows the precision, recall and accuracy of how well each definition of RecSys serendipity and generalized serendipity predict the set of articles considered serendipitous by participants. From the user perspective, rel_nov_unexp, the most common definition of RecSys serendipity, had both the highest precision (0.57) and accuracy (0.73). Moreover, relevance had exceptionally high recall (0.86) for user serendipity. Generalized serendipity performs comparatively poorly across metrics versus all other definitions (with the exception of unexpectedness), suggesting that participants' understanding of serendipity is more inline with the components of RecSys serendipity than the other types of scenario in which the term could also be applied. We elaborate further on the implications of this finding in discussion.

**Table 4: Coefficients for fixed effects from two mixed-effect logistic regression models with relevance, novelty and unexpectedness as independent variables (with corresponding P-values in brackets). Significant P-values (P < 0.0015) are in bold.**

| Serendipity def. | Relevance | Novelty | Unexpectedness | pseudo-$R^2$ |
|---|---|---|---|---|
| generalized | **0.51 (2e-15)** | **0.24 (2e-4)** | 0.08 (0.09) | 0.12 |
| user | **1.39 (2e-16)** | **0.44 (7e-8)** | **0.38 (7e-9)** | 0.31 |

## 4.2 Relative Importance of RecSys Serendipity Components

Given the significant overlap between articles that meet multiple definitions of serendipity, we wanted to understand the relative importance of relevance, novelty and unexpectedness in both generalized and user serendipity. To address RQ2, we fitted two mixed-effect logistic regression models. Each model included three independent variables corresponding to participants' 5-point ratings of relevance, novelty and unexpectedness statements from the article surveys. The dependent variable for each model corresponded to whether an article was serendipitous according to the generalized or user definitions. To account for repeated measures, we included a random intercept for each participant. Table 4 shows the estimated regression coefficients for both models. We used McFadden's pseudo-$R^2$ as a measure of model fit [24]. We applied Bonferroni correction to the significance threshold as follows: 0.05/33 = 0.0015 (based on the 33 statistical tests conducted in this paper).

Table 4 shows that all components of RecSys serendipity are associated with user serendipity, whereas only relevance and novelty were associated with generalized serendipity. For both user and generalized serendipity, however, the strongest and most statistically significant association was with relevance, highlighting its importance to serendipity in general. Furthermore, the relative ordering of components by the magnitude of their coefficients is the same for user and generalized serendipity[7]. Lastly, the goodness of fit for each model using RecSys serendipity components was much higher for user serendipity than for generalized serendipity ($R^2$ = 0.31 and 0.12, respectively). This shows that (i) user serendipity is

---

[7]This assumes that the coefficient for unexpectedness in the model for generalized serendipity is zero as it was not statistically significant.

**Table 5: Coefficients for fixed effects from nine mixed-effect logistic regression models with serendipity definitions as dependent variables and user behavior measures as independent variables (with corresponding P-values in brackets). Significant P-values (P < 0.0015) are in bold.**

| Serendipity def. | Interactions | Reading progress | Time spent | pseudo-$R^2$ |
|---|---|---|---|---|
| generalized | 0.35 (0.01) | **1.01 (3e-8)** | 0.17 (6e-3) | 0.06 |
| user | **1.22 (2e-16)** | **0.66 (5e-4)** | 0.24 (2e-3) | 0.09 |
| rel | **2.30 (2e-16)** | **1.29 (7e-13)** | 0.17 (2e-3) | 0.11 |
| nov | **1.19 (4e-15)** | **0.93 (8e-8)** | 0.08 (0.09) | 0.05 |
| unexp | 0.31 (0.04) | 0.39 (0.03) | 0.02 (0.6) | 0.08 |
| rel_nov | **1.64 (2e-16)** | **1.05 (2e-9)** | **0.17 (7e-4)** | 0.08 |
| rel_unexp | **1.47 (2e-16)** | **1.11 (2e-9)** | 0.17 (7e-3) | 0.10 |
| nov_unexp | **0.92 (2e-11)** | **0.82 (5e-6)** | 0.11 (0.06) | 0.07 |
| rel_nov_unexp | **1.44 (2e-16)** | **1.11 (2e-8)** | 0.21 (2e-3) | 0.10 |

closer than generalized serendipity to RecSys serendipity as relevance, novelty and unexpectedness explain a greater proportion of variance and (ii) a substantial proportion of the variance in both user and generalized serendipity remains unexplained.

## 4.3 Impact of Serendipity on User Behavior

In RQ3, we wanted to understand how user behavior was impacted after encountering serendipitous articles. We fitted nine mixed-effect logistic regression models. Each model included independent variables for user interactions, reading progress and time spent in the application. The dependent variable for each model corresponded to whether an article was serendipitous according to a given definition of serendipity. To account for repeated measures, we included a random intercept for each participant.

Table 5 shows the estimated regression coefficients for each model. User interactions and reading progress were associated with a majority of serendipity definitions (7/9 definitions), confirming that participants valued serendipitous encounters (otherwise they would not have read or bookmarked the articles). However, generalized serendipity was only associated with reading progress, showing that information needs to be consumed to achieve new goals, but, as the most common goal achieved was to *"learn new things"* (see Figure 2(b)), it does not necessarily need to be bookmarked. Time spent was only associated with rel_nov, suggesting that higher user engagement with the application did not generally result in more serendipitous encounters. Lastly, model fit was uniformly low across all serendipity definitions. User serendipity, for example, had better model fit using RecSys serendipity components as predictors (0.31, from Table 4) than user behaviors (0.09). While serendipity appears to have a measurable impact on user behavior (as demonstrated by significant regression coefficients), there is limited potential for passively monitoring serendipity using interaction data (as indicated by extremely low $R^2$ values).

## 5 DISCUSSION AND LIMITATIONS

The main goal of this research was to compare the various definitions of serendipity used to assess recommender systems with generalized serendipity, a conceptually broader definition focused on problem-solving. We also wanted to understand how users' subjective understanding of serendipity compared to these more formal definitions. We conducted an observational study over a period of 16 weeks. In the study, we collected 2002 survey responses from 397 users of an online article recommender system. In our results, we found a significant proportion of serendipitous recommendations were missed by the conventional definitions used in the recommender systems research literature. Furthermore, we identified a disconnect between both RecSys and generalized serendipity and what users believed to be serendipitous encounters. Lastly, our attempts to model generalized and user serendipity fell short, with relevance, novelty and unexpectedness, and different types of behavioral data having very low goodness of fit. We discuss the implications of these results in the following subsections.

## 5.1 The Dark Matter of Serendipity

Our study highlights the "dark matter" of serendipity in recommender systems research; recommendations that should be considered serendipitous and are, therefore, beneficial to users, but have been overlooked due to the narrow definitions of serendipity used in prior work. We presented numerous results to substantiate this claim. First, more articles fitted the definitions of generalized and user serendipity compared to RecSys serendipity in terms of raw counts (687 versus 789 and 1075 for RecSys, user and generalized serendipity, respectively). Second, even the best performing RecSys definitions had only moderate precision for both generalized and user serendipity. Furthermore, the most common variant of RecSys serendipity had the worst performance in terms of recall (see Tables 2 and 3). Lastly, the explanatory power of relevance, novelty and unexpectedness was shown to be limited for both generalized and user serendipity (see $R^2$ values in Table 4).

Unobserved serendipity has numerous implications. The lack of consistency between user studies in terms of the selection of definition and the composition of surveys already makes it difficult to compare results. However, the possibility of unobserved serendipity has the potential to dramatically alter the conclusions of past studies. Indeed, even in system development, an algorithm that has been optimized to recommend serendipitous items is likely to under perform as a result of being trained on data that was labelled based on RecSys serendipity.

In future work, we plan to investigate the implications of this unaccounted for serendipity. First, our current study looked at

serendipity from the perspective of sets of items, however, recommendations are usually generated from a ranked list. Therefore, we want to understand how different definitions of serendipity affect the number of serendipitous encounters in top-n recommendations. Next, Wang and Chen showed that serendipity can manifest in domain-specific ways [38]. We want to understand whether this is the case with generalized serendipity, or if it is more robust to differences in domain.

## 5.2 Serendipity or User Experience?

We showed that user serendipity – users' colloquial understanding of serendipity – resulted in a different set of serendipitous articles compared to generalized and RecSys serendipity. In particular, we showed that the most common version of RecSys serendipity had only moderate precision when classifying user serendipity (see Table 3). Furthermore, while the explanatory power of relevance, novelty and unexpectedness was higher for user serendipity than generalized serendipity, the model fit still low (0.31, see Table 4). Lastly, relevance appears to capture most of the items users identify as serendipitous, whereas novelty identifies the majority of items with respect to generalized serendipity (recall scores of 0.86 and 0.71, respectively, see Tables 2 and 3).

These findings have consequences for recommender systems design: optimizing a recommender for either generalized or RecSys serendipity will not align the resulting recommendations with users expectations of serendipity, which may negatively impact user experience. In future work we want to understand the interplay between serendipity and user experience, and investigate whether it would actually be beneficial to narrow the RecSys definition of serendipity further by attempting to filter out recommendations that would not meet users' subjective understanding of the term.

## 5.3 Beyond Relevance, Novelty and Unexpectedness

Our results showed that neither the RecSys serendipity components nor the user behavior data we collected could be used to model generalized or user serendipity with even moderate goodness of fit (see Tables 4 and 5). If we want to produce descriptive models of serendipity, then we need to identify additional candidate variables to account for the unexplained variance. Based on prior work, our study design could be extended to investigate the following: context-dependency, user and item characteristics, and the impact of user interfaces. We cover each in turn:

- **Context-dependency:** Contextual information, such as weather, time of day and location, have long been known to impact the efficacy of recommender systems [3] and, therefore, will impact all definitions of serendipity due to the general importance of relevance (see Table 4).
- **User and item characteristics:** RecSys serendipity has been shown to be associated with user characteristics, such as demographic information and personality traits [37], and item characteristics, such as item popularity and similarity to items already consumed [14]. Both are likely to have an impact on user and generalized serendipity as well.
- **Impact of user interface:** Both user interface [33] and recommendation list composition [16, 19] have been shown to

alter user perceptions of recommendations. Indeed, intra-list diversity has already been shown to alter RecSys serendipity [16]. We believe that these factors could also benefit user and generalized serendipity.

## 5.4 Limitations

There are several limitations with our study. First, our results are based on an observational study and not a randomized controlled trial, limiting us to identifying correlations related to serendipity. This limitation is common among serendipity studies because the short duration of laboratory studies may not include any serendipitous encounters (for comparison, our study ran for 16 weeks). Second, previous studies have highlighted domain-specific factors in serendipity, but we only investigated recommendations in a single domain: online articles, which could limit the generalizability of our findings. Third, we investigated recommendations that were sampled at random from user-selected topics. While this had the advantage that it could not suffer from the inherent biases in top-n recommender systems, such as popularity and sparsity bias, it is not clear whether this affected our results. Fourth, as we notified participants that the study was about serendipity, this might have had a priming effect on survey responses. Lastly, our study implicitly assumes that all serendipitous encounters should be weighted equally, but this is not necessarily the case from the user perspective.

## REFERENCES

[1] Panagiotis Adamopoulos and Alexander Tuzhilin. 2011. On unexpectedness in recommender systems: Or how to expect the unexpected. In *Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011), at the 5th ACM International Conference on Recommender Systems.* ACM, 11–18.

[2] Panagiotis Adamopoulos and Alexander Tuzhilin. 2014. On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected. *ACM Transactions on Intelligent Systems and Technology* 5, 4 (2014), 1–32.

[3] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-Aware Recommender Systems. In *Recommender Systems Handbook.* Springer US, 217–253.

[4] Naresh Kumar Agarwal. 2015. Towards a definition of serendipity in information behaviour. *Information research: an international electronic journal* 20, 3 (2015), n3.

[5] Zheshi Bao and Jing Yang. 2022. Why online consumers have the urge to buy impulsively: roles of serendipity, trust and flow experience. *Management Decision* (2022).

[6] Li Chen, Yonghua Yang, Ningxia Wang, Keping Yang, and Quan Yuan. 2019. How serendipity improves user satisfaction with recommendations? a large-scale user evaluation. In *The world wide web conference.* 240–250.

[7] Jessica T DeCuir-Gunby, Patricia L Marshall, and Allison W McCulloch. 2011. Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field methods* 23, 2 (2011), 136–155.

[8] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity. In *Proceedings of RecSys'10* (Barcelona, Spain). ACM, New York, NY, USA, 257–260.

[9] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. 2004. Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004), 5–53.

[10] Yoshinori Hijikata, Takuya Shimizu, and Shogo Nishida. 2009. Discovery-oriented Collaborative Filtering for Improving User Satisfaction. In *Proceedings of IUI'09* (Sanibel Island, Florida, USA) *(IUI '09).* ACM, New York, NY, USA, 67–76.

[11] Marius Kaminskas and Derek Bridge. 2014. Measuring Surprise in Recommender Systems. In *Proceedings of the Workshop on Recommender Systems Evaluation: Dimensions and Design (Workshop Programme of the 8th ACM Conference on Recommender Systems).*

[12] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. Interact. Intell. Syst.* 7, 1, Article 2 (Dec. 2016), 42 pages.

[13] Aleksandra Karpus, Iacopo Vagliano, and Krzysztof Goczyła. 2017. Serendipitous recommendations through ontology-based contextual pre-filtering. In *Beyond Databases, Architectures and Structures. Towards Efficient Solutions for Data Analysis and Knowledge Representation: 13th International Conference, BDAS 2017, Ustroń, Poland, May 30-June 2, 2017, Proceedings 13.* Springer, 246–259.

[14] Denis Kotkov, Joseph A Konstan, Qian Zhao, and Jari Veijalainen. 2018. Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing.* 1341–1350.

[15] Denis Kotkov, Alan Medlar, and Dorota Glowacka. 2023. Rethinking Serendipity in Recommender Systems. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23).* Association for Computing Machinery, New York, NY, USA.

[16] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. 2020. How does serendipity affect diversity in recommender systems? A serendipity-oriented greedy algorithm. *Computing* 102, 2 (2020), 393–411.

[17] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192.

[18] Denis Kotkov, Qian Zhao, Kati Launis, and Mats Neovius. 2020. ClusterExplorer: Enable User Control over Related Recommendations via Collaborative Filtering and Clustering. In *Proceedings of the 2020 ACM conference on Recommender systems.*

[19] Jiqun Liu. 2023. Toward A Two-Sided Fairness Framework in Search and Recommendation. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval.* 236–246.

[20] Stephann Makri, Jaffor Bhuiya, Jermaine Carthy, and Jermaine Owusu-Bonsu. 2015. Observing serendipity in digital information environments. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–10.

[21] Andrii Maksai, Florent Garcin, and Boi Faltings. 2015. Predicting Online Performance of News Recommender Systems Through Richer Evaluation Metrics. In *Proceedings of RecSys'15.* ACM, New York, NY, USA, 179–186.

[22] David Martin. 2008. Most untranslatable word. http://www.todaytranslations.com/blog/most-untranslatable-word/ [Online; accessed 21-April-2023].

[23] Christian Matt, Alexander Benlian, Thomas Hess, and Christian Weiß. 2014. Escaping from the Filter Bubble? The Effects of Novelty and Serendipity on Users? Evaluations of Online Recommendations. In *Thirty Fifth International Conference on Information Systems.*

[24] Daniel McFadden. 1974. Conditional logit analysis of qualitative choice behavior. In *Fontiers in Econometrics*, Paul Zarembka (Ed.). Academic press, New York, 105–142.

[25] Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being Accurate is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (Montreal, Quebec, Canada) *(CHI EA '06).* ACM, New York, NY, USA, 1097–1101.

[26] Robert K Merton and Elinor Barber. 2011. *The travels and adventures of serendipity: A study in sociological semantics and the sociology of science.* Princeton University Press.

[27] Theodore G Remer. 1965. *Serendipity and the three princes: From the Peregrinaggio of 1557.* Norman, U. Oklahoma P, 20.

[28] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. *Recommender Systems Handbook.* Springer US, Chapter Introduction to Recommender Systems Handbook, 1–35.

[29] Isabelle Rivoal and Noel B Salazar. 2013. Contemporary ethnographic practice and the value of serendipity. *Social Anthropology* 21, 2 (2013), 178–185.

[30] Alan Said, Ben Fields, Brijnesh J. Jain, and Sahin Albayrak. 2013. User-centric Evaluation of a K-furthest Neighbor Collaborative Filtering Recommender Algorithm. In *Proceedings of CSCW'13* (San Antonio, Texas, USA). ACM, New York, NY, USA, 1399–1408.

[31] Shigekazu Sawaizumi, Osamu Katai, Hiroshi Kawakami, and Takayuki Shiose. 2007. Using the concept of serendipity in education. (2007).

[32] Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science* 22, 11 (2011), 1359–1366.

[33] Annelien Smets, Lien Michiels, Toine Bogers, and Lennart Björneborn. 2022. Serendipity in Recommender Systems Beyond the Algorithm: A Feature Repository and Experimental Design. In *16th ACM Conference on Recommender Systems.* CEUR Workshop Proceedings, 44–66.

[34] Annelien Smets, Jorre Vannieuwenhuyze, and Pieter Ballon. 2022. Serendipity in the city: User evaluations of urban recommender systems. *Journal of the Association for Information Science and Technology* 73, 1 (2022), 19–30.

[35] Suboojitha Sridharan. 2014. *Introducing serendipity in recommender systems through collaborative methods.* Ph. D. Dissertation. UNIVERSITY OF RHODE ISLAND.

[36] Maria Taramigkou, Efthimios Bothos, Konstantinos Christidis, Dimitris Apostolou, and Gregoris Mentzas. 2013. Escape the Bubble: Guided Exploration of Music Preferences for Serendipity and Novelty. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong, China). ACM, New York, NY, USA, 335–338.

[37] Ningxia Wang and Li Chen. 2022. How do item features and user characteristics affect users' perceptions of recommendation serendipity? A cross-domain analysis. *User Modeling and User-Adapted Interaction* (2022), 1–39.

[38] Ningxia Wang, Li Chen, and Yonghua Yang. 2020. The Impacts of Item Features and User Characteristics on Users' Perceived Serendipity of Recommendations. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization.* 266–274.

[39] Ohid Yaqub. 2018. Serendipity: Towards a taxonomy and a theory. *Research Policy* 47, 1 (2018), 169–179.

[40] Yuan Cao Zhang, Diarmuid Ó Séaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of WSDM'12.* ACM, New York, NY, USA, 13–22.

[41] Reza Jafari Ziarani and Reza Ravanmehr. 2021. Deep neural network approach for a serendipity-oriented recommendation system. *Expert Systems with Applications* 185 (2021), 115660.

[42] Reza Jafari Ziarani and Reza Ravanmehr. 2021. Serendipity in recommender systems: a systematic literature review. *Journal of Computer Science and Technology* 36, 2 (2021), 375–396.