Improving the Discoverability of Indie Games by Leveraging their Similarity to Top-Selling Games

Identifying Important Requirements of a Recommender System

Quang N. Vu
vu.n.quang@ualberta.ca
Analytics of Software, GAmes And Repository Data (ASGAARD) lab, University of Alberta
Edmonton, Alberta, Canada

Cor-Paul Bezemer
bezemer@ualberta.ca
Analytics of Software, GAmes And Repository Data (ASGAARD) lab, University of Alberta
Edmonton, Alberta, Canada

ABSTRACT

Indie games often lack visibility as compared to top-selling games due to their limited marketing budget and the fact that there are a large number of indie games. Players of top-selling games usually like certain types of games or certain game elements such as theme, gameplay, storyline. Therefore, indie games could leverage their shared game elements with top-selling games to get discovered. In this paper, we propose an approach to improve the discoverability of indie games by recommending similar indie games to gamers of top-selling games. We first matched 2,830 `itch.io` indie games to 326 top-selling `Steam` games. We then contacted the indie game developers for evaluation feedback and suggestions. We found that the majority of them (67.9%) who offered verbose responses show positive support for our approach. We also analyzed the reasons for bad recommendations and the suggestions by indie game developers to lay out the important requirements for such a recommendation system. The most important ones are: a standardized and extensive tag and genre ontology system is needed to bridge the two platforms, the expectations of players of top-selling games should be managed to avoid disappointment, a player's preferences should be integrated when making recommendations, a standardized age restriction rule is needed, and finally, the recommendation tool should also show indie games that are the least similar or less popular.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Applied computing → Computer games**.

KEYWORDS

indie games, computer games, game discoverability, Steam, Itch.io

ACM Reference Format:
Quang N. Vu and Cor-Paul Bezemer. 2021. Improving the Discoverability of Indie Games by Leveraging their Similarity to Top-Selling Games: Identifying Important Requirements of a Recommender System. In *International Conference on the Foundations of Digital Games (FDG '21), August 3–6, 2021, Montreal, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/0000000.0000000

FDG '21, August 3–6, 2021, Montreal, Canada

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/0000000.0000000

1 INTRODUCTION

Indie games are games that are developed by a single developer or a small team who are independent of publishers [51]. With a limited marketing budget and manpower, indie game developers often face difficulty in getting their games known to gamers. To make matters worse, popular indie game distribution platforms, such as `itch.io`, may contain hundreds of thoudsands of games. Therefore, the discoverability problem is extremely challenging for indie game developers to solve. On the contrary, some of the most popular games (which are often developed and published by large game development studios) attract millions of players. These games often have their own franchise where a predecessor's success contributes significantly to its successor's discoverability.

In this paper, we propose to improve the discoverability of indie games from `itch.io` by leveraging their similarity to top-selling games from the `Steam` platform. There have been only a few prior studies on game discoverability. Past studies [47, 56] focus on game recommendation approaches for `Steam` games, or a small number of games from unspecified or less popular game stores [5, 13, 41]. To the best of our knowledge, our study is the first study to explore a cross-store approach to improve game discoverability that targets the indie games on the `itch.io` platform and to investigate whether indie game developers value such an approach.

In particular, we recommend indie games in which gamers are likely to be interested based on their preferred top-selling `Steam` game. The rationale is that players tend to like related games [42]. First, we matched 2,830 indie games with 326 top-selling games based on their similarity in tags, genres, and game descriptions. Then, we conducted a user study with 195 indie game developers to evaluate our matching results and elicit feedback and suggestions. We showed them a list of similar `Steam` games for which their indie games would be recommended so that we could collect all evaluations for each `itch.io` game. It is important to note that the eventual implementation of our approach should show the reverse direction: a list of similar `itch.io` games for each `Steam` game that a player chooses. Finally, we studied their responses to identify the most important requirements of our proposed indie game recommendation system. To find out the precision of our approach, we conducted a preliminary study of the following Preliminary Question (PQ):

PQ. How precise is our recommendation approach?

Motivation: We calculated several similarity scores where we assigned different weights to the similarity in tags, genres, and description keywords. We evaluated the final average precision based on the upvoted recommendations by the developers to understand which similarity component of the algorithm should be more heavily weighted.

Findings: We found that a more heavily weighted tag similarity in the overall score would give a slightly better average precision at the top five recommendations. However, the performance of our recommendation approach is still low. We also collected a ground truth dataset of 2,604 recommendations from the evaluations in our user study.

Then, we answer the following main Research Questions (RQs):

RQ1. How do indie developers feel about improving the discoverability of their games through our approach?

Motivation: We presented our approach to indie game developers from `itch.io` and asked for their overall thoughts. The goal is to find out the main challenges to such an approach to indie game recommendation.

Findings: The majority (67.9%) of the developers support our approach while 29.6% show a neutral sentiment. The remaining 2.5% do not support our idea. We extracted two reasons from their responses. First, the developers said that indie games are too unique to be matched by similarity. Second, players of top-selling games might be disappointed when playing indie games.

RQ2. What are the requirements of a future version of an indie game recommendation system?

Motivation: We qualitatively studied the feedback on downvoted recommendations to understand what are the main reasons for the downvotes. Additionally, we analyzed the suggestions given by indie game developers. Our goal is to lay out the requirements for future studies on the development of indie game recommendation systems.

Findings: We consolidated 10 requirements for future studies on indie game recommendation. Most importantly, a standardized and extensive tag and genre ontology system and age restriction rule are needed. Second, a gamer's expectations should be managed when recommending indie games that are more narrow in scope. Third, a gamer's preferences should be integrated when recommending similar indie games. Finally, the recommender should also recommend the indie games that are the least similar to showcase their uniqueness and less popular indie games.

The findings from our study are useful for future studies on indie game discoverability and the development of indie game recommendation systems. The most important contributions of our study are:

(1) A user study with 195 indie game developers that shows the positive support for our approach from the majority of the surveyed developers.
(2) A set of important requirements for indie game recommendation systems.
(3) A labelled dataset that shows the developer-perceived relevance of recommendations of `itch.io` indie games and top-selling `Steam` games [3].

Table 1: Number of games available on different distribution platforms (as of July 5th, 2020)

Type	Platform	# of games
Platforms with many AAA games	Steam	79,007
	Green Man Gaming	7,189
	GamersGate	6,374
	GOG.com	3,992
	Direct2Drive	2,454
	Epic Games	362
Platforms with almost no AAA games	Itch.io	264,179
	Kongregate	128,664
	Newgrounds	90,651
	Game Jolt	18,278

The rest of this paper is organized as follows. Section 2 provides background information and related work. Section 3 explains our methodology. Section 4 presents a preliminary analysis on the precision of our approach. Section 5 describes how indie game developers feel about our approach. Section 6 discusses the requirements for future studies on such an approach. The threats to validity are described in Section 7. Finally, Section 8 concludes the paper.

2 BACKGROUND AND RELATED WORK

2.1 An overview of the Steam and itch.io platforms

The `Steam` platform is a digital game distribution platform developed by Valve Corporation and currently has more than 79,000 games. It is considered one of the largest game distribution platforms with a peak number of concurrent players of more than 20 million [2]. Table 1 shows the current number of games available on `Steam` compared to several other game distribution platforms. The games distributed on `Steam` are mostly AAA games. Although there is no official definition of AAA games, they are usually regarded as games developed by major game development studios under a large development and marketing budget [15]. These games often hold a spot on `Steam`'s top-selling list [50].

On the other hand, `itch.io` focuses on small independent game developers to help them distribute smaller-budgeted games. In comparison to other indie game distribution platforms, `itch.io` is the largest (see Table 1). Due to the low cost of publishing games, `itch.io` is more suitable for small-scale independent developers.

2.2 Related work

Discoverability of games. The discoverability problem is faced by both indie and non-indie games alike. Indie games usually get discovered by being shared by their gamers or through word of mouth. Some examples of this sharing mechanism are: asking friends for help (Candy Crush Saga) [37], sharing game replay snippets (Poly Bridge) [4], or sharing quirky game videos (Untitled Goose Game) [17]. To the best of our knowledge, there have been no prior published studies on how effective these sharing mechanisms are in helping to increase the discoverability of indie games. `Steam Labs` has conducted many game discoverability experiments but their implementation details are not available to the public and they focus on games hosted on `Steam` itself. Ryan et al. proposed

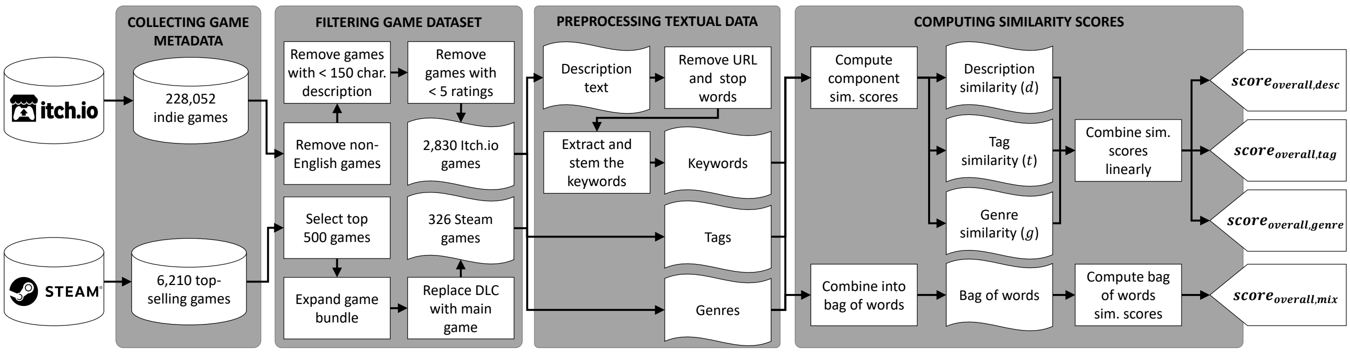

Figure 1: An overview of our methodology.

GameNet [43], GameSage [41], and GameSpace [40] which employ *latent semantic analysis* to train on Wikipedia game entry texts to build a game relatedness network of nearly 12,000 games to help game designers discover related games to their own. The study that is the closest to ours is one by Kholodylo and Strauss [22]. Its authors interviewed five respondents to analyze the relevance of recommender systems from the perspective of indie game developers. Our work differs from these prior studies in that we focus on recommending similar indie games to players of top-selling games on two game stores and we conducted a qualitative study on a much larger number (195) of indie game developers. We also provide a labelled dataset of 2,604 recommendations that shows the relevant matching between several `itch.io` and `Steam` games.

Game recommendation systems. Many prior works on game recommendation focus on in-game item recommendation ([6, 7, 11, 20, 31, 52]) or leverage the gamer profile to make recommendations. Yang and Huang [56] employed text mining to extract a gamer's personality traits from their game reviews and social messages to make a game recommendation based on the similarity between their personality and the aggregated personality of all game reviews. In the medical field, Catalá et al. [9] presented a recommendation method that uses a user's social network information to create personalized, user-friendly games for the elderly and disabled. Chow et al. [13] proposed an approach that leverages content-based and user-based information for mobile game recommendation and evaluated it on a local game mobile platform called WePlay. Sifa et al. [47] built a game recommendation system that employs an archetypal analysis approach to group similar gamers to make recommendations based on their shared preferences and tested on 500,000 users and more than 3,000 `Steam` games. Anwar et al. [5] implemented a collaborative-filtering (CF) game recommender using game ratings of a community of gamers and those of a particular gamer to recommend new games to that gamer. Smith [49] patented a method that combines gamer's profile information (e.g., previous downloaded or purchased games, playing pattern) and their friends' preferences to recommend new games to the gamer. Pérez-Marcos et al. [34] combined content-based and CF techniques to build a hybrid game recommender that was evaluated on a dataset of 3,600 `Steam` games and 11,350 gamers. Their approach showed an improvement over the baseline approach (a modified music recommender [33]). Cheuque et al. [12] proposed three `Steam` game recommendation approaches that use factorization machines, deep learning, and CF technique on game's and gamer's features and achieved better recommendation results than a baseline CF approach.

These prior studies focus on a single game store or in-game item recommendation. Our work explores an indie game recommendation approach that is based on the cross-store content-based similarity between two distinct platforms (`itch.io` and `Steam`).

Mining game distribution platforms. Most of the prior work on mining game distribution platforms focused on `Steam`. Lin et al. conducted empirical studies of several aspects of the `Steam` platform [26–29]. For example, they showed that positive game reviews (not just negative reviews) also contain useful feedback and developers should invest in the first few hours of the game, games that release frequently tend to have a higher proportion of urgent updates, developers should take advantage of the early access model to gather more positive reviews, and finally, game bug videos can be identified with high precision using a random forest classifier buit using several video features such as keyword matches in video description, tags, and title. Sifa et al. [48] studied the playtimes of 6 million `Steam` gamers to detect several archetypes of gamers. Blackburn et al. [8] studied cheaters on Steam Community and revealed that the number of cheater friends correlates with the likelihood of a gamer becoming a cheater. Another group of studies mined game mod distribution platforms (e.g., Nexus Mods and CurseForge) to study how games and their mods interact with each other and yield insightful results for both game developers and mod developers [16, 24, 25, 35]. Finally, Vu and Bezemer [53] performed an empirical study on `itch.io` game jams and their submissions and found that games with a better quality description, multi-platform and multi-genre support are more likely to be high-ranking. Our work is the first cross-store study that mines game data from both `itch.io` and `Steam`.

3 METHODOLOGY

In this Section, we describe our experiment setup. Fig. 1 gives an overview of our methodology.

3.1 Collecting game metadata

We developed customized web crawlers to retrieve game metadata from `itch.io` and `Steam`, taking note of the crawling rules set

Table 2: An overview of our game dataset.

Total # of `itch.io` games	228,052
Total # of studied `itch.io` games	2,830
Total # of top-selling `Steam` games	6,210
Total # of studied `Steam` games	326

by both sites in their respective `robot.txt` file. Table 2 shows an overview of our game dataset.

Collecting `itch.io` game metadata. We ran a crawler to retrieve a snapshot that includes the metadata of 228,052 games from `itch.io` [21] on April 15th, 2020. We collected the game name, developer(s), language, description, tag(s), genre(s), number of ratings, and URL.

Collecting `Steam` game metadata. We ran another crawler to retrieve the metadata of top-selling `Steam` games [50] on April 9th, 2020. There were 6,210 games on this list. For each of those games, we collected the game name, short description, long description, tag(s), genre(s), and URL.

3.2 Filtering game dataset

Filtering the `itch.io` game dataset. We removed games that are not in English based on their language metadata. We also removed games with a description length of less than 150 characters (the third quartile of the distribution of the description length). We consider the remaining games to have enough descriptive text for our text-similarity calculation. Finally, we removed games with fewer than five ratings from players; we deem these games likely to be toy or unfinished projects. Our final `itch.io` game dataset has 2,830 games.

Filtering the `Steam` game dataset. Among the 6,210 top-selling `Steam` games, we kept the top 500 games; we consider these as the most popular games. We manually went through these games and checked if they are a bundle or downloadable content (DLC). If the game is a bundle of multiple games, we expanded the bundle into its component games. If it is a DLC, we replaced it with its original game. There are instances where both a game and its DLCs are in the dataset; we removed the DLCs and kept the original game. We ended up with 326 `Steam` games in our final dataset.

3.3 Preprocessing textual data

We preprocessed the `itch.io` game descriptions and `Steam` game main descriptions to standardize them into a common input space for our algorithms. In particular:

(1) We removed HTTP links because they may be unique for each text and thus do not contribute to similarity.
(2) We used the Rapid Automatic Keyword Extraction (RAKE) [39] technique to extract words that appear in candidate keyword phrases, which can represent the content of the game description concisely. In this study, we used the `rake-nltk` implementation and used the `nltk` corpus's list of English stopwords to remove words that provide little meaning to the documents (e.g., "the", "of"). The minimum length of candidate keyword phrases was set to three. Therefore, the algorithm would produce a list of words with a degree of at least three (i.e, words that co-occur with at least three words, including themselves, in all candidate keyword phrases).
(3) We used the Snowball stemmer [36] to reduce the extracted words to their root form (e.g., "connected", "connection", and "connecting" are reduced to "connect").

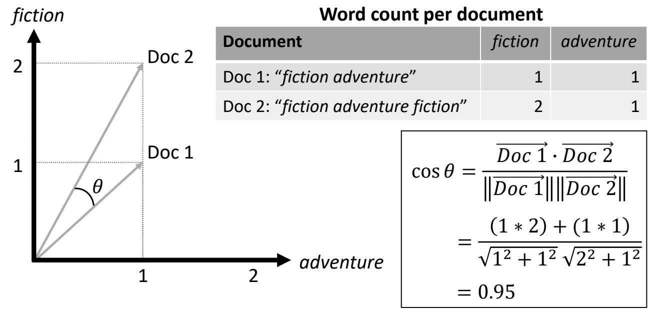

Figure 2: Two simplified documents are represented as vectors of word count. Their cosine similarity score is the cosine of the angle θ between the vectors.

3.4 Computing similarity scores

Computing component similarity scores. To assess the similarity between two games, we computed their similarity in tag, genre, and description. We used these metadata because (1) games that share one or more tags are likely to share several similar game elements described by the tags, (2) games in the same genre are likely to have a similar gameplay, and (3) games that have similar descriptions are likely to have similar storylines, themes, characters, etc. For each pair of `itch.io` and `Steam` game, we computed their tag similarity score (denoted as t), genre similarity score (g), and description similarity score (d). To compute the score between two textual inputs, we converted them into numerical vector representations using the `CountVectorizer` [44]. Finally, we calculated the pairwise cosine similarity score between any two games. The cosine similarity value measures the cosine of the angle between two document vectors: the smaller the angle, the greater its cosine value, the more similar the two documents are. Fig. 2 shows how to calculate the cosine similarity score between two sample text documents.

Computing overall similarity scores. We combined the three component scores into an overall similarity score. Since each component may contribute differently to the similarity, we assigned a weight for each score and added them linearly. Formally, we define the overall similarity score between an `itch.io` game and a `Steam` game as:

$$score_{overall} = \alpha * t + \beta * g + \gamma * d \quad (1)$$

where α, β, and γ are the weights of the score components. To investigate which component contributes to better similarity matching between an `itch.io` game and a `Steam` game, we computed three overall scores, each with a different component score having a weight of 0.5 and the other two component scores having an equal weight of 0.25 (the sum of all three weights is 1). We denote $score_{overall,tag}$, $score_{overall,genre}$, $score_{overall,desc}$ as the overall scores for the three scoring methods with the tag, genre, description having heavier weights respectively. Furthermore, we devised

Table 3: Component score weight of different overall similarity scores.

Overall score	α	β	γ
$score_{overall,tag}$	0.50	0.25	0.25
$score_{overall,genre}$	0.25	0.50	0.25
$score_{overall,desc}$	0.25	0.25	0.50
$score_{overall,mix}$	*NA*	*NA*	*NA*

a fourth method where we put all game tags, genres, and description words together in a bag of words. We then computed a single overall similarity score (denoted as $score_{overall,mix}$) between the two games based on the similarity of their bags of words. Table 3 shows the component weights for each overall scoring method.

3.5 Getting indie game developer feedback

Extracting the unique set of `Steam` game matches. For each of the 2,830 indie games, we extracted the top 5 `Steam` games with the highest $score_{overall,tag}$, $score_{overall,genre}$, $score_{overall,desc}$, and $score_{overall,mix}$. Theoretically, there could be 20 `Steam` game that are matched to each indie game. However, because there are overlaps in the matches across the 4 overall scores, each indie game was matched with at least 5 `Steam` games and at most 15. Fig. 3 shows the distributions of the number of matched `Steam` games per indie game: the majority are each matched with 10 `Steam` games.

Collecting developer feedback. To evaluate the matching between an `itch.io` game and a `Steam` game, we collected feedback from the developers of the indie games in our dataset because they best understand the content of their games. We built a website[1] to display a set of matching `Steam` games for each indie game. For each `Steam` game, there is a short description and a link to its main page. We showed the indie game developers the reverse direction of recommendations: for each developer, we showed them their `itch.io` game and the unique set of similar `Steam` games for which their game would be recommended. We did this to collect all evaluations for each `itch.io` game. Note that the eventual implementation of our approach should allow gamers to select their preferred `Steam` game to get a list of recommended indie games. The order of the matched `Steam` games was randomized when displayed to the developers to reduce the bias when evaluating the first few matches (with higher scores) versus matches further down the list. For each match, we asked the developers to give "thumbs-up" if they thought it is similar to the selected indie game, and "thumbs-down" otherwise, and elaborate on their choice. Finally, we asked them to provide their thoughts on the idea of improving the discoverability of indie games by providing suggestions to the players of top-selling games about indie games that they may enjoy. A screenshot of an example of the form to collect developer's feedback is available online [3].

We obtained the contact details of the indie game developers from their public `itch.io` profile page. These details include Twitter handles, email addresses, Facebook pages, and Discord servers. We managed to contact 1,526 developers and received 195 responses which equals to a response rate of 12.8%. These responses include evaluations of 2,604 `Steam` game matches of 262 `itch.io` games.

[1] http://www.recommendindie.games (see sample screenshot in [3])

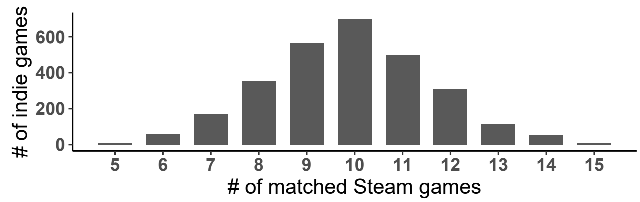

Figure 3: The distribution of the number of matched `Steam` games per `itch.io` game.

4 PRELIMINARY STUDY ON THE PRECISION OF OUR RECOMMENDATION APPROACH

Motivation: We studied the contributions of tags, genres, and description similarity toward good recommendations to understand which component should be more heavily weighted in the overall similarity score calculation. The findings are useful for our subsequent qualitative analysis.

Approach: **Evaluating the precision.** We received 682 upvoted recommendations. We used the *Average Precision at n (AP@n)* [23] metric to evaluate the precision. The *AP@n* measures the number of relevant recommendations (upvotes) in the top *n* recommended items while also taking into account their positions in the ranked list. We used the following variant of the formula:

$$AP@n = \frac{1}{n}\sum_{k=1}^{n}(P@k \text{ if } k^{th} \text{ item is relevant}) \qquad (2)$$

where $P@k = \frac{r}{k}$ (k is the number of recommendations, r is the number of upvotes within the first k recommendations). Since we extracted the top five most similar games, we calculated $AP@5$ for each scoring method. If there is one upvote out of five `Steam` games for an `itch.io` game, the $AP@5$ is 0.46 if the upvote is for the first game in the list, and 0.04 if it is the last.

Additionally, to understand whether there is a significant difference between the $AP@5$ values of different similarity scores, we statistically compared the distribution of the $AP@5$ values using the one-sided Mann-Whitney U test [54]. The test has the null hypothesis that one distribution is not greater than the other. If the test returns a p-value that is less than 0.05, we conclude that one distribution is statistically larger than the other and reject the null hypothesis. To understand the magnitude of the difference, we used Cliff's delta d [30], with the thresholds suggested by Romano et al. [38]: negligible if $|d| \leq 0.147$; small if $0.147 < |d| \leq 0.33$; medium if $0.33 < |d| \leq 0.474$; and large if $0.474 < |d| \leq 1$.

Results: **Tag similarity contributes the most to relevant recommendations.** Fig. 4 shows the distributions of the $AP@5$ values of the four scoring methods. The recommendation method in which tag similarity is weighted heavier ($score_{overall,tag}$) has the highest median $AP@5$ at 0.17. The Mann-Whitney U test shows that this scoring method has a statistically significantly larger $AP@5$ (both p-values are 0.03) than the genre-heavy-weight ($score_{overall,genre}$) and the description-heavy-weight ($score_{overall,desc}$) method. However, the effect sizes for both comparisons are negligible (0.09). The median $AP@5$ for the latter two methods are both 0.1. Comparing between the tag-heavy-weight scoring method and the bag-of-word

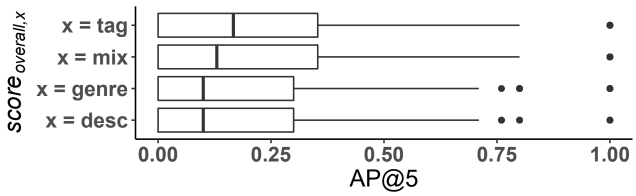

Figure 4: The distribution of the *AP*@5 values of the four similarity scoring methods.

scoring method ($score_{overall,mix}$), there is no statistically significant difference between their *AP*@5 values (p-value > 0.05). Finally, the bag-of-word scoring method also has a statistically significantly higher *AP*@5 than the genre-heavy- and the description-heavy-weight method, though with a negligible effect size.

All of the median *AP*@5 values remain low. However, since many indie games strive to be unique (as later found in Section 6), they might not have any relevant top-selling game matches (in our list, or at all). This is, therefore, an extremely difficult recommendation problem and a high average precision is hard to achieve. Additionally, to the best of our knowledge, there exists no pre-defined ground truth of relevant matchings between indie games from `itch.io` and top-selling `Steam` games. Therefore, we strived to establish an initial labelled dataset for future research in indie game discoverability and recommendation (publicly available at [3]).

Summary: Tag similarity contributes the most to relevant recommendations between an `itch.io` game and a `Steam` game. Further studies are required to optimize the weights assigned to the component similarity scores.

5 HOW DO INDIE DEVELOPERS FEEL ABOUT OUR RECOMMENDATION APPROACH?

Motivation: Indie games tend to lack visibility due to a low marketing budget. Therefore, we presented our recommendation approach to help increase their visibility. First and foremost it is important to understand whether indie game developers support this idea, as there may be considerations that we overlooked. Hence, first we studied whether indie game developers would support the recommendation approach.

Approach: **Open coding the developer's sentiment in their overall feedback responses.** To understand developers' sentiment about our approach, we studied the overall feedback provided by the indie game developers. We were able to extract 159 non-empty responses to the sentiment question. Next, independently, the first and the second author manually read through these responses and categorized the sentiment of the developers into three categories: "Positive" (the developer supports the idea), "Negative" (the developer does not think it is a good idea), and "Neutral" (the developer's response does not show how they feel about the idea). Then the authors compared the results and found that both agreed in the categorization of 133 responses (83.6%). There are 23 cases where one author categorized the response as "Neutral" and the other author put "Positive" or "Negative", the reason for which is that some responses are not obviously positive or negative, making it difficult to categorize them. There are only three responses where both authors categorized the opposite. The disagreements were resolved through discussion until a consensus was reached.

Results: **67.9% of the indie game developers that offered verbose responses support our idea of an indie game recommendation approach.** The majority (108 out of 159) of the responses indicated that the developers think it is a good idea to help increase the discoverability of indie games based on their similarity with top-selling `Steam` games. In 47 responses (29.6%), developers did not show their support nor rejection of the idea. For example, one wrote *"Could work for some indie games, but based on the above recommendations a lot of smaller weirder indie games (which make up a fair percentage of the total) are completely their own thing and have little similarity to most AAA games."* Many developers gave suggestions on how to improve the recommendation algorithm. We discuss the suggestions in-depth in Section 6.

2.5% of the developers do not support our approach for indie game recommendation, due to gamer expectation management or a niche market. Four developers (2.5%) have a negative feeling about our approach. We manually read through them and extracted the following reasons:

- *Gamer expectation management:* the developers mentioned that players of top-selling games usually have high expectations in terms of the level of polish, which is difficult to achieve for small indie game development teams. Indie games tend to be small and short therefore gamers may feel underwhelmed and not enjoyable. Prior work by Chambers et al. [10] showed that gamers are difficult to please: they have specific gameplay preferences.
- *Niche market:* the developers wrote that indie games strive to be different from other games and should not be considered as similar to top-selling games.

Although these are valid reasons, we felt that future iterations of our approach could leverage on gamers' preferences, on top of content similarity, when recommending indie games. As found in Section 6, many developers made such a suggestion to improve our indie game recommendation approach.

Summary: The majority (67.9%) of the developers that offered verbose responses support our idea to improve the discoverability of indie games by recommending similar indie games to top-selling `Steam` games' players. 2.5% do not support such an approach because the indie games might not meet the expectations of gamers or the indie games strive to be in a niche market.

6 WHAT ARE THE REQUIREMENTS OF A FUTURE VERSION OF AN INDIE GAME RECOMMENDATION SYSTEM?

Motivation: We received thumbs-down on some recommendations with the reasons given by the indie game developers. We studied these reasons to understand the shortcomings of our approach. Additionally, we analyzed the suggestions given by the developers to understand the potential improvement points to achieve higher quality recommendations. The findings will be useful for future developers and researchers of indie game recommendation systems.

```
1. Inputs = All feedback responses, a list of categories
     of reason (intially an empty list)

2. For each feedback response:
    - Read the response manually for this match
    - If the reason raised in the response matches
         existing category(-ies):
        Label the response with that/those category(-ies)
    - Else:
        Add a new category to the category list
        Restart the process with the new category list

3. Outputs = All feedback responses labelled with
     categories and a list of categories of reason
```

Listing 1: The categorization process.

Approach: **Open coding the feedback on downvoted recommendations.** We received 1,959 downvoted recommendations, of which 873 were accompanied by the developer's reason for downvoting. To have an overall understanding of the reasons for the bad matches, we extracted a random statistically-representative sample of 87, which has a 95% confidence level and 10% confidence interval. We manually categorized the reasons by performing an iterative process that is similar to *Open Coding* [45, 46]. Listing 1 shows the details of this process. The first and second authors categorized the reasons independently and then compared the results. To calculate the agreement between the two authors, we used the following method. First, for each category of reason, we calculated the agreement rate by counting the number of matches (we consider a match happens when both authors tag or do not tag the same category for a reason) and dividing it by the sample size. Second, we took the average of the agreement rates across all categories. The average agreement rate is 92.3%, which shows a high agreement between authors. Any disagreement was discussed to reach a consensus. We extracted seven reasons for the bad recommendation matches between `itch.io` indie games and top-selling `Steam` games. Table 4 shows these reasons and their description, an example response, and the number of responses that were tagged. Since a response can be tagged with multiple reasons, the sum of all counts is larger than the sample size.

Open coding the developer's suggestions in their overall feedback responses. We used the same 159 extracted responses as in Section 5. Following the same *Open Coding* approach, the first and the second authors independently read and categorized the suggestions raised in those responses. The average agreement rate was 97.7% across the categories. Any discrepancy was discussed until we reached a consensus. We extracted 17 categories of suggestions from the responses. Table 5 shows the description of the identified suggestions, along with an example response.

Results: The remainder of this Section discusses the consolidated requirements R1–R10 identified from Table 4 and 5.

R1: A standardized and extensive tag and genre ontology system. Table 5 shows that 34 developers (C1) suggested that indie games and top-selling games should be matched on specific similar elements. These elements range from gameplay (e.g., "resource management", "platformer") to theme (e.g., "nature related") to graphics (e.g., "art style"). We observed that gameplay mismatch, and look and feel mismatch are the second and third dominant reasons for

Table 4: Identified reasons for bad recommendation matches.

Category	Description (D) - Example (E)	Count
Genre mismatch	**D:** The genres of the games are different. **E:** *"While the content is horror related, the genres are wildly different."*	41
Gameplay mismatch	**D:** There are differences in the specific way that the games are played. **E:** *"Not a significantly bad match, I would just not connect the two games as their play styles are so different"*	32
Look and feel mismatch	**D:** The games are different in graphics, theme, or overall mood. **E:** *"does not have zombies, or dark themes"*	22
Target audience mismatch	**D:** The intended target audiences of both games are different. **E:** *"It seems to be targeted at different audience"*	6
Scope mismatch	**D:** The duration of time required to play and the content size of the games are different. **E:** *"Dash Connect is a short casual game, while this is an open world exploration game"*	3
Single-/multi-player mismatch	**D:** One game supports singleplayer or multiplayer and the other does not. **E:** *"We do not have multiplayer implemented yet"*	3
Not similar at all/Others	**D:** Developers said that the games are entirely different, or did not give a specific reason. **E:** *"Not similar at all"*	21

downvoted recommendations. Table 4 shows that 32 (36.8%) reasons for downvoting are categorized as gameplay mismatch. The feedback on these downvoted recommendations is largely about the difference in playstyle between the games. In an example match between the `You Are Undead` game from `itch.io` and the `Biped` game from `Steam`, the response indicated that *"The action/puzzle description kinda fits YAU's design, but Biped emphasizes the co-op, 'couch play' feature. I think it wouldn't be a good recomendation. But well, it's a good guess."* In the "look and feel mismatch" category, we categorized 22 responses (25.3%) (see Table 4). These responses mention words such as *mood*, *theme*, *atmosphere*, *tone*, and *vibe*. An example is the match between the `Tiler` game and the `Warhammer 40,000: Mechanicus` game, developers said that both games are strategy games but the latter one has a much heavier theme and tone while the former is simpler with flat colors. Since game tags reflect the primary game elements, having a relevant set of assigned tags that would, therefore, help improve the recommendation quality.

In addition, game genre terminology should be more fine-grained, extensively defined, and standardized between `itch.io` and `Steam` to augment the recommendation performance, as suggested by 12 developers (C4). For example, a developer suggested that their simulation game should be matched with "management simulation" games. However, both `itch.io` and `Steam` games in our dataset do not have this genre, but a common "simulation" genre. Another suggestion is to split the "Adventure" genre further into sub-genres such as "Action RPG", "Adventure RPG", "Turn-based RPG".

We observed that among the reasons for downvoting a recommendation, genre mismatch is the most prevalent reason. Table 4 shows that 41 (47.1%) responses are categorized with "Mismatched genre". We investigated these 41 responses further to see the distribution of genres. Fig. 5 shows the distribution of game genres of the `itch.io` games whose matched `Steam` game received downvotes

Table 5: Identified suggestions for our recommendation approach.

Category	Description (D) - Example (E)	Count
C1	**D:** The games should be matched on other similar elements such as content, theme, aesthetics. **E:** *"[...] the algorithm could also discover games which have a similar tone or aesthetic to the AAA games. For example, players who like Animal Crossing also like Stardew Valley. Even if the gameplay isn't necessarily similar, both games evoke a similar lighthearted, cozy feeling that scratches a similar itch."*	33
C2	**D:** Developers suggest specific `Steam` game examples to be recommended. **E:** *"Staxel, Eco are quite similar."*	29
C3	**D:** The indie game is too unique or caters to a niche target audience. **E:** *"A great idea, though Leximan is quite unique and the only thing I would find it comparable to is Undertale - which did not appear on the list."*	27
C4	**D:** More fine-grained genres or more weight to certain genres are needed. **E:** *"[...] I would look into separating RPGs further, between Action RPGs, Adventure RPGs, Turn-based RPGs, etc [...]"*	12
C5	**D:** The gamer's expectation should be raised right when it comes to indie games due to their usually smaller scope than that of top-selling games. **E:** *"[...] I think one issue in comparing small games on itch to AAA titles on steam is that the scope of games is often so different that players used to AAA scope may be disappointed in the scope of small independent games [...]"*	10
C6	**D:** Indie games should only be matched with other indie games. **E:** *"[...] Our game is perhaps not the best to try to find a match for - all similar games we know of are indie [...]"*	7
C7	**D:** Developers suggest we improve our study setup such as having a similarity scale instead of just thumbs-up vs thumbs down, contacting via email instead of Twitter. **E:** *"[...] I found the thumbs up/thumbs down system a bit too restrictive, as sometimes I found myself wanting to rate something inbetween. I wish I had four options, like "similar" "slightly similar" "not very relevant" "non-relevant" [...]"*	4
C8	**D:** Developers raised a concern that some matched games are not AAA although our goal is to match with top-selling games. **E:** *"some of these games are not AAA games"*	4
C9	**D:** Developers suggested a more extensive way to query for game recommendations. **E:** *"[...] I also think it would be nice to have system that can translate description supplied by players in a natural language to queries that can filter games by features more elaborate than tags such as game mechanics, art style, etc. It could rely to gameplay videos (or trailers) to collect the needed information to index games. I usually find myself having a hard time searching for games that fit my desires even using the usual tag systems or text queries [...]"*	3
C10	**D:** This recommendation approach should be tied to an existing site and made available to the public. **E:** *"[...] The idea as a whole is phenomenal. Especially if used in either of the platforms (itch or steam) [...]"*	3
C11	**D:** The "similarity" terminology should be avoided. Games should be recommended based on what the gamers might like. **E:** *"[...] The idea of improving discoverability sounds like a good one, but I think it shouldn't necessarily be correlated to big blockbuster games. It would be a bit more fair to suggest something "that the player doesn't know yet if they might like"."*	3
C12	**D:** Developers will improve the description, tag(s), and genre(s) of their indie games to help improve recommendation matching. **E:** *"[...] The issue is, many developers, including me at the time of release, try to somewhat "game" the tag system by really expanding what the "action" tag means, for example. This will properly cause issues for your fantastic site and so, I will updating the tags of my future games to reflect honestly on my games."*	1
C13	**D:** Not-Safe-For-Work (NSFW) games should not be matched with SFW games. **E:** *"[...] dont associate a sfw game with a nsfw one [...]"*	1
C14	**D:** Our recommendation should be compared to Steam recommendation. **E:** *"[...] Perhaps interesting to compare the Steam recommendations to yours."*	1
C15	**D:** There should be a limit range of popularity so that less popular games can get discovered. **E:** *"[...] If you want people to discover indie games, it would be better if games could only link to games within a certain range of popularity or lower than them. That would prevent users funneling upward to the biggest games, which they would probably already know about."*	1
C16	**D:** Besides showing similar games, games that are entirely different should be highlighted as well. **E:** *"[...] it would be great if you could display also the opposite results, i.e. the indie games that do not match AAA games, those which shine for originality and creativity."*	1
C17	**D:** Developers gave no suggestion in their feedback responses. **E:** *"A great goal. Indie games deserve more love!"*	63

due to genre mismatch (note that a game can have multiple genres). The "Visual Novel" genre of the `itch.io` games received the largest number of downvoted `Steam` game matches due to genre mismatch. This could be because `itch.io` has a "Visual Novel" genre and no "Visual Novel" tag while `Steam` has the opposite: it has the genre but not the tag. A further check of our dataset shows that 464 (16.4%) of the studied `itch.io` games have the "Visual Novel" genre while only 6 (1.8%) of the studied top-selling `Steam` games have the "Visual Novel" tag.

The above example shows there is an inconsistency between the tags and genres used by `itch.io` and by `Steam`. In another example, one developer said that *"[...] these results were quite off, missing two of the main genres of the game like roguelike and deckbuilding, for this case [...]"*. "Roguelike" is a sub-genre of role-playing game and "Deckbuilding" is a sub-genre of card game or board game. We inspected our dataset and found that both `itch.io` and `Steam` do not have the "Roguelike" or "Deckbuilding" genre. Additionally, `itch.io` has "Roguelike", "Rogue-like", "Deck building", "Deck-builder" tags while `Steam` has "Roguelike" and "Deckbuilding" tags. Therefore, future studies should explore a more extensive tag and genre ontology system that describes the games more thoroughly and can bridge the two game stores through mapping.

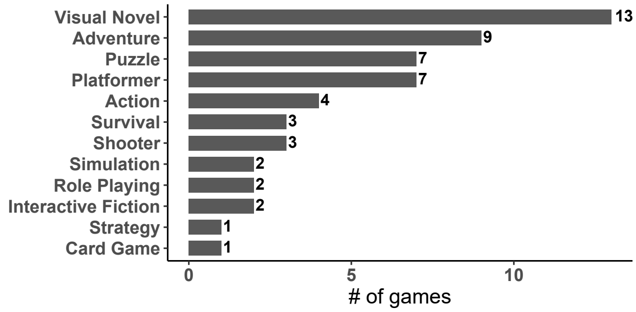

Figure 5: The distribution of game genres among those that were downvoted due to genre mismatch.

Finally, there should be verification of tags assigned to a game to prevent developers from adding irrelevant tags (similar to "keyword stuffing" [1]). This is because one developer (C12) admitted that many developers "game" the tag system by adding irrelevant tags to their games which would impact the similarity matching.

R2: Manage a gamer's expectations. Table 5 shows that 10 developers (C5) suggested gamer's expectations should be managed appropriately so that they will not feel disappointed when they come to play an indie game after playing a top-selling AAA game. We also observed that 27 developers (C3) think that indie games are too different from games in the top-selling `Steam` list. For example, one developer wrote that *"[...] Your software has the capacity to work well for indie games which are roughly analogous to AAA genres (e.g. indie fighting games to AAA fighting games, indie puzzle games to AAA puzzle games), but the independent sphere has a large number of games which simply do not fit in the AAA genre paradigm [...]"*. Additionally, seven developers (C6) suggested that indie games should only be matched with other indie games (e.g., *"The only similar experiences I've seen were coming from the indie circles [...]"*). Four developers (C8) said that some of the matched `Steam` games are not AAA games, although this could be because 132 of the 326 top-selling games in our dataset have the "indie" genre assigned. S

These suggestions are consistent with the finding in Section 5 that a small number (2.5%) of indie game developers do not support our approach because of the scope difference between indie games and top-selling games. However, Table 4 shows only three reasons for downvoted matches are due to the difference in scope or length of the games. This is interesting because although indie games are usually smaller in scope than a top-selling title [32] due to a limited budget, the gap between them is not the main reason for downvoted recommendations, which underlines the feasibility of our idea. Therefore, future indie game recommendation systems should manage gamer's expectations well, e.g., by showing the recommended indie game's scope and length to gamers.

R3: Integrate a gamer's preferences when recommending similar indie games. Three developers (C11) suggested we should not use the "similarity" terminology but rather recommend games that gamers might like regardless of similarity. For example, one developer gave an example that gamers might like both `Horizon Zero Dawn` and `Stardew Valley` despite both games having little similarity. This requirement supports the previous finding that it is difficult to match indie games with top-selling games based on similarity alone and thus, the gamer's historical preference should be leveraged for the game recommendation. We discuss several prior works in Section 2 that used gamer's information for game recommendation [5, 12, 13, 34, 47, 49, 56]. However, these studies used gamer's data and make game recommendations within the same platform (`Steam`). Additionally, it is also worth noting that data pertaining to what gamers have added to their favorite list are restricted by `itch.io` and `Steam` themselves due to privacy concerns. Past studies [27, 29] used historical game owner data from Steam Spy [18]. However, the fact that a gamer owns a game does not necessarily mean they like that game. Also, recommendation approaches that rely solely on gamer preference would face the cold-start problem: new users without any preference would not be able to receive any recommendation [19]. Therefore, future studies should explore other ways to collect preference data of gamers who play games from both `itch.io` and `Steam` platforms and combine with our similarity content-based matching approach to provide higher quality and personalized recommendations.

R4: Implement standardized age-based restriction on explicit contents. One developer (C13) suggested filtering for adult-content games from matches to a general-audience game. Both `itch.io` and `Steam` have a "NSFW" (Not-Safe-For-Work) tag for adult-content games. Therefore, an additional check is needed to verify the gamer's age before recommending games tagged with "NSFW". Additionally, while `Steam` has a standard age rating (i.e., ESRB), `itch.io` does not [14]. Because the list of tags for adult- and graphic-content can be non-exhaustive (e.g., tags such as "Gore" denote graphic violent content), there is a need to compose a common and comprehensive list of tags for such purpose.

R5: Allow developers to add recommendations. In Table 5, 29 developers (C2) wrote down the specific `Steam` games they think should be recommended for their indie games. For example, developers of the `Hardware Tycoon` game recommended the `Game Dev Tycoon` game due to similar gameplay and theme. It should be noted that in our study, we used a dataset of 326 top-selling `Steam` games. At the time of our data collection, the games mentioned by the developers were not on the top-selling list. However, our goal in this paper is to show that our indie game recommendation approach can be done automatically and that it receives the majority of positive feedback. Therefore, future indie game recommendation studies should allow developers to manually add games they feel relevant to the recommendation lists (i.e., by showing a separate *staff-pick* list) or use a larger `Steam` game dataset.

R6: Allow gamers to input a description of the types of games they want to be recommended. Three developers (C9) suggested that a more flexible way to query for games should be implemented. For example, gamers should be able to give a requirement in text form, and games that match it would be recommended. Prior research by Ryan et al. [41] allows game-design students to search for related games by textual description of an idea. Other prior studies investigated several approaches to search for relevant in-game moments using natural language queries [57] or visual queries [58]. These techniques could be adapted to work on gamer-supplied descriptions to identify the required matched game elements in the indie games they might like.

R7: Integrate the approach into an existing popular game distribution platform. Three developers (C10) suggested that our approach should be integrated into an existing site such as itch.io or Steam to attract a large audience.

R8: Also recommend the indie games that are the least similar to showcase their uniqueness. One developer (C16) suggested that the recommendation tool should also showcase indie games that do not match any top-selling games. As we discussed earlier, many developers mentioned the uniqueness of indie games, making these games difficult to be matched. Highlighting such unique games would still allow these games to benefit from the approach in this paper. One approach is to gather the itch.io games that are the least similar to all other Steam games and display them as a separate list.

R9: Limit the range of the recommended indie games' popularity. One developer (C15) suggested that the tool should not always recommend highly popular indie games to prevent gamers from being always drawn towards more popular games (similar to a suggestion found by Kholodylo and Strauss [22]). Therefore, future indie game recommendation tools could use metrics such as a game's rating or the number of ratings as an indicator of its popularity to make recommendations accordingly.

R10: Improve the data collection setup of the study and compare the results with Steam's recommendation. Four developers (C7) suggested improving the setup for our study. In particular, they suggested the rating of recommendation should be on the scale of 1 to 10 instead of just thumbs-up and thumbs-down so that different degrees of similarity can be recorded. Another suggestion is that they prefer to be contacted via email instead of other social media means. However, the majority of itch.io indie game developers in our dataset do not provide public emails on their page while most have Twitter accounts. Therefore, we chose to contact them via Twitter when emails were not available. Finally, one developer (C14) suggested comparing our approach with Steam recommendations, although this would be difficult because the majority of itch.io games are not on Steam.

7 THREATS TO VALIDITY

Internal validity: Threats to *internal validity* relate to the bias of the experimenter. We relied on manual effort to categorize developer responses. These tasks are difficult to automate. Topic modeling techniques such as the latent Dirichlet allocation (LDA) require a large amount of data and rely on the use of similar game terminology across all responses while developers tend to use different terminology in their responses. A similar software engineering study [55] also analyzed survey responses manually. To mitigate this threat, the two authors categorized the responses independently and compared the results. We achieved a high level of average agreement rate and any disagreement was resolved through discussion.

External validity: We only studied games from the itch.io and Steam platforms. These two platforms are the largest of their kinds by the number of games hosted (see Table 1). Therefore, the games in our dataset should be representative of a large number of games which helps us identify the core improvement points for our approach. Further studies could investigate how our recommendation approach generalizes to other game distribution platforms.

Additionally, we evaluated the itch.io to Steam game matches by asking the indie game developers to rate for their games while the target implementation of the recommendation system is the other way around (i.e., allowing gamers to select their preferred Steam game to get recommended a list of indie games). The indie game developers may be biased about their own games. Future studies should also collect feedback from players of the Steam games.

Construct validity: We used open-ended questions when we asked for developer responses. Close-ended questions with a pre-defined set of answers (e.g., a pre-defined set of reasons for bad recommendation matches) could have made it easier to aggregate the responses. However, close-ended questions are restrictive and do not provide rich answers. Since our study is the first to explore cross-store game recommendation, we felt that we needed substantial feedback from the developers to improve our approach.

Another threat is that we used the *thumbs-up vs thumbs-down* scale to measure whether a Steam game matches an itch.io game. We used this binary scale for ease of calculating the precision metrics (which need a binary input) and for separating good and bad recommendations. We found in Section 6 that a few developers suggested we use a scale of 1–10 instead. Future studies should explore using such a scale to evaluate recommendation approaches.

Finally, our recommendation results have fairly low average precisions. However, the goal of our study is to obtain developers' feedback on our approach and to lay out the requirements for such a recommendation system, and collect an initial set of ground truth of relevant recommendations. Therefore, we could not optimize the recommendation results before contacting the developers.

8 CONCLUSION

In this study, we proposed an approach to improve the discoverability of indie games on itch.io by recommending similar indie games to players of top-selling Steam games. First, we showed that our approach can be automated by calculating the similarity score between an indie game and a top-selling game. Then, we collected feedback from 195 indie game developers to help us evaluate our approach. The most important findings of our study are:

- A majority (67.9%) of indie game developers that offered verbose responses support our recommendation idea.
- It is feasible to automate the recommendation by cross-store content-based similarity matching but further work is required to optimize the results.
- A labelled dataset showing the relevant matches between itch.io indie games and top-selling Steam games [3].

We also lay out the 10 requirements for future iterations of our approach. The most important ones are:

1. An extensive standardized tag and genre ontology system.
2. Manage a gamer's expectations when recommending indie games that are more narrow in scope.
3. Integrate a gamer's preferences.
4. Implement standardized age-based restriction.
5. Also recommend indie games that are the least similar to showcase their uniqueness and less popular indie games.

Developers of indie game recommendation systems can use our findings to improve the recommendation quality. Researchers in indie game discoverability can use our findings to further extend the research direction in this field.

REFERENCES

[1] [n.d.]. Irrelevant keywords - Search Console Help. https://support.google.com/webmasters/answer/66358 (Last accessed: 2020-07-05).

[2] [n.d.]. Steam: Game and Player Statistics. https://store.steampowered.com/stats/Steam-Game-and-Player-Statistics (Last accessed: 2020-07-05).

[3] [n.d.]. Supplementary material for our paper: Labelled Dataset and Sample Screenshot. https://figshare.com/s/bed46a84774015f89b7b (Last accessed: 2020-08-13).

[4] 2017. Video: How replays helped Poly Bridge succeed via social media. https://www.gamasutra.com/view/news/308236/Video_How_replays_helped_Poly_Bridge_succeed_via_social_media.php (Last accessed: 2020-08-10).

[5] Syed Muhammad Anwar, Talha Shahzad, Zunaira Sattar, Rahma Khan, and Muhammad Majid. 2017. A game recommender system using collaborative filtering (GAMBIT). In *14th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*. IEEE, 328–332.

[6] Vladimir Araujo, Felipe Rios, and Denis Parra. 2019. Data mining for item recommendation in MOBA games. In *Proceedings of the 13th ACM Conference on Recommender Systems*. 393–397.

[7] Paul Bertens, Anna Guitart, Pei Pei Chen, and África Periáñez. 2018. A Machine-Learning Item Recommendation System for Video Games. In *IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–4.

[8] Jeremy Blackburn, Nicolas Kourtellis, John Skvoretz, Matei Ripeanu, and Adriana Iamnitchi. 2014. Cheating in Online Games: A Social Network Perspective. *Transactions on Internet Technology (TOIT)* 13, 3 (2014), 9.

[9] Laura Catalá, Vicente Julián, and José-Antonio Gil-Gómez. 2014. A CBR-Based Game Recommender for Rehabilitation Videogames in Social Networks. In *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 370–377.

[10] Chris Chambers, Wu-chang Feng, Sambit Sahu, and Debanjan Saha. 2005. Measurement-based Characterization of a Collection of On-line Games. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*. 1–1.

[11] Zhengxing Chen, Truong-Huy D Nguyen, Yuyu Xu, Christopher Amato, Seth Cooper, Yizhou Sun, and Magy Seif El-Nasr. 2018. The Art of Drafting: A Team-Oriented Hero Recommendation System for Multiplayer Online Battle Arena Games. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 200–208.

[12] Germán Cheuque, José Guzmán, and Denis Parra. 2019. Recommender Systems for Online Video Game Platforms: the Case of STEAM. In *Companion Proceedings of The 2019 World Wide Web Conference*. 763–771.

[13] Anthony Chow, Min-Hui Nicole Foo, and Giuseppe Manai. 2014. HybridRank: A Hybrid Content-Based Approach To Mobile Game Recommendations. In *CBRecSys@ RecSys*. 10–13.

[14] Leaf Corcoran. 2018. Lets talk about a content/age rating system for itch.io - Ideas and Feedback. https://itch.io/t/251186/lets-talk-about-a-contentage-rating-system-for-itchio (Last accessed: 2020-08-09).

[15] Dana. 2020. What Are AAA Games? A Guide To Unofficial Terminology. https://geekygamingstuff.com/what-are-aaa-games/ (Last accessed: 2020-08-12).

[16] Tapajit Dey, Jacob Logan Massengill, and Audris Mockus. 2016. Analysis of Popularity of Game Mods: A Case Study. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 133–139.

[17] Allegra Frank. 2019. Untitled Goose Game is a funny video game about an asshole goose. Its power is in its simplicity. https://www.vox.com/culture/2019/10/10/20897212/what-is-untitled-goose-game-explained-nintendo-switch-pc-mac (Last accessed: 2020-08-10).

[18] Sergey Galyonkin. [n.d.]. Steam Spy. http://steamspy.com/ (Last accessed: 2020-07-30).

[19] Huba Gaspar. 2018. The Cold Start Problem for Recommender Systems. https://www.yuspify.com/blog/cold-start-problem-recommender-systems/ (Last accessed: 2020-08-13).

[20] Lucas Hanke and Luiz Chaimowicz. 2017. A Recommender System for Hero Line-Ups in MOBA Games. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

[21] Itch.io. [n.d.]. Top Games. https://itch.io/games (Last accessed: 2020-08-12).

[22] Marta Kholodylo and Christine Strauss. 2019. What Independent Game Developers Expect from Recommender Systems: A Qualitative Study. In *Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services*. 563–570.

[23] Ray R Larson. 2010. Introduction to Information Retrieval. *Journal of the American Society for Information Science and Technology* 61, 4 (2010), 852–853.

[24] Daniel Lee, Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. 2020. Building the Perfect Game–An Empirical Study of Game Modifications. *Empirical Software Engineering* (2020), 1–23.

[25] Daniel Lee, Gopi Krishnan Rajbahadur, Dayi Lin, Mohammed Sayagh, Cor-Paul Bezemer, and Ahmed E Hassan. 2019. An Empirical Study of the Characteristics of Popular Minecraft Mods. *Empirical Software Engineering* (2019), 1–23.

[26] Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. 2017. Studying the urgent updates of popular games on the Steam platform. *Empirical Software Engineering* 22, 4 (2017), 2095–2126.

[27] Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. 2018. An empirical study of early access games on the Steam platform. *Empirical Software Engineering* 23, 2 (2018), 771–799.

[28] Dayi Lin, Cor-Paul Bezemer, and Ahmed E Hassan. 2019. Identifying gameplay videos that exhibit bugs in computer games. *Empirical Software Engineering* (2019), 1–28.

[29] Dayi Lin, Cor-Paul Bezemer, Ying Zou, and Ahmed E Hassan. 2019. An Empirical Study of Game Reviews on the Steam Platform. *Empirical Software Engineering* 24, 1 (2019), 170–207.

[30] Jeffrey D Long, Du Feng, and Norman Cliff. 2003. Ordinal Analysis of Behavioral Data. *Handbook of psychology* (2003).

[31] Wenli Looi, Manmeet Dhaliwal, Reda Alhajj, and Jon Rokne. 2018. Recommender System for Items in Dota 2. *IEEE Transactions on Games* (2018).

[32] Brendan Lowry. 2017. The major differences between 'indie' and 'AAA' video games. https://www.windowscentral.com/indie-vs-aaa-which-type-game-you (Last accessed: 2020-07-25).

[33] Maciej Pacula. 2009. A matrix factorization algorithm for music recommendation using implicit user feedback.

[34] Javier Pérez-Marcos, Lucía Martín-Gómez, Diego M Jiménez-Bravo, Vivian F López, and María N Moreno-García. 2020. Hybrid system for video game recommendation based on implicit ratings and social networks. *Journal of Ambient Intelligence and Humanized Computing* (2020), 1–11.

[35] Lev Poretski and Ofer Arazy. 2017. Placing Value on Community Co-creations: A Study of a Video Game 'Modding' Community. In *Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 480–491.

[36] Martin F Porter. 2001. Snowball: A language for stemming algorithms.

[37] Jason Della Rocca. 2020. Discoverability by Design: Rethinking How Video Games Get Noticed. https://trends.cmf-fmc.ca/embedded-discoverability-video-games (Last accessed: 2020-08-13).

[38] Jeanine Romano, Jeffrey D Kromrey, Jesse Coraggio, Jeff Skowronek, and Linda Devine. 2006. Exploring methods for evaluating group differences on the NSSE and other surveys: Are the t-test and Cohen's d indices the most appropriate choices. In *Annual meeting of the Southern Association for Institutional Research*.

[39] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic Keyword Extraction from Individual Documents. *Text mining: applications and theory* 1 (2010), 1–20.

[40] James Owen Ryan, Eric Kaltman, Andrew Max Fisher, Taylor Owen-Milner, Michael Mateas, and Noah Wardrip-Fruin. 2017. GameSpace: An Explorable Visualization of the Videogame Medium. (2017).

[41] James Owen Ryan, Eric Kaltman, Timothy Hong, Katherine Isbister, Michael Mateas, and Noah Wardrip-Fruin. 2016. GameNet and GameSage: Videogame Discovery as Design Insight. In *Proceedings of 1st International Joint Conference of DiGRA and FDG*.

[42] James Owen Ryan, Eric Kaltman, Timothy Hong, Michael Mateas, and Noah Wardrip-Fruin. 2015. People Tend to Like Related Games. In *International Conference on the Foundations of Digital Games (FDG)*.

[43] James Owen Ryan, Eric Kaltman, Michael Mateas, and Noah Wardrip-Fruin. 2015. What we talk about when we talk about games: Bottom-up game studies using natural language processing. In *International Conference on the Foundations of Digital Games (FDG)*.

[44] Scikit-learn. [n.d.]. CountVectorizer. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html (Last accessed: 2020-07-25).

[45] Carolyn B. Seaman. 1999. Qualitative Methods in Empirical Studies of Software Engineering. *IEEE Transactions on Software Engineering* 25, 4 (1999), 557–572.

[46] Carolyn B Seaman, Forrest Shull, Myrna Regardie, Denis Elbert, Raimund L Feldmann, Yuepu Guo, and Sally Godfrey. 2008. Defect categorization: making use of a decade of widely varying historical data. In *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. 149–157.

[47] Rafet Sifa, Christian Bauckhage, and Anders Drachen. 2014. Archetypal Game Recommender Systems. In *Proceedings of the LWA 2014 Workshops: KDML, IR and FGWM*. 45–56.

[48] Rafet Sifa, Christian Bauckhage, and Anders Drachen. 2014. The Playtime Principle: Large-scale cross-games interest modeling. In *Conference on Computational Intelligence and Games*. IEEE, 1–8.

[49] Darren Smith. 2012. Recommendation engine for electronic game shopping channel. US Patent 8,298,087.

[50] Steam. [n.d.]. Top Sellers. https://store.steampowered.com/search/?filter=topsellers (Last accessed: 2020-08-12).

[51] Craig Stern. 2015. What makes a game indie: a universal definition. https://sinisterdesign.net/what-makes-a-game-indie-a-universal-definition (Last accessed: 2020-08-01).

[52] Zhaojie Tao, Ming Cheung, James She, and Ringo Lam. 2014. Item Recommendation Using Collaborative Filtering in Mobile Social Games: A Case Study. In *IEEE*

International Conference on Big Data and Cloud Computing. IEEE, 293–297.

[53] Quang N Vu and Cor-Paul Bezemer. 2020. An Empirical Study of the Characteristics of Popular Game Jams and Their High-ranking Submissions on itch.io. In *International Conference on the Foundations of Digital Games (FDG)*. 1–11.

[54] Frank Wilcoxon. 1992. Individual Comparisons by Ranking Methods. In *Breakthroughs in statistics*. Springer, 196–202.

[55] Yuhao Wu, Shaowei Wang, Cor-Paul Bezemer, and Katsuro Inoue. 2019. How do developers utilize source code from Stack Overflow? *Empirical Software Engineering* 24, 2 (2019), 637–673.

[56] Hsin-Chang Yang and Zi-Rui Huang. 2019. Mining personality traits from social messages for game recommender systems. *Knowledge-Based Systems* 165 (2019), 157–168.

[57] Xiaoxuan Zhang and Adam M Smith. 2019. Retrieving Videogame Moments with Natural Language Queries. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–7.

[58] Xiaoxuan Zhang, Zeping Zhan, Misha Holtz, and Adam M Smith. 2018. Crawling, Indexing, and Retrieving Moments in Videogames. In *Proceedings of the 13th International Conference on the Foundations of Digital Games*. 1–10.