Analysis of Popularity of Game Mods: A Case Study

Tapajit Dey
University of Tennessee
Knoxville, USA
tdey2@vols.utk.edu

Jacob Logan Massengill
University of Tennessee
Knoxville, USA
jacobmassengill10@gmail.com

Audris Mockus
University of Tennessee
Knoxville, USA
audris@utk.edu

Abstract

Video game mods have become an integral part of the gameplay experience for PC gamers and have drawn support of major game development companies. This area, nevertheless, is not well understood, especially in terms of what motivates mod creators and users. To explore this question we propose a data-driven approach that relies on data that can be obtained from online fora devoted for game mods. To illustrate our approach, we have collected data on deployment and popularity for the mods in six major PC games made by two companies and spanning more than a decade from different popular mod hosting websites. In particular, we investigate what features are present in the mods that are being developed and what features are popular among mod users. To accomplish that, we propose measures of popularity of the mods based on the number of unique downloads and define mod feature space by the tags associated with each mod. The preliminary investigation suggests, for example, that the features developed across the distinct games to be similar but the sets of features popular among users were only marginally similar to the sets of features being implemented by mod creators. We plan surveys of mod users and mod creators to determine causes for this discrepancy. We hope that our approach would allow answering important research and practical questions in the area of game mod development.

Author Keywords
Game Mods;Popularity;Tags;Mod features

ACM Classification Keywords
H.5.m [Information Interfaces and Presentation]: Miscellaneous; [Games User Research]

Introduction
Video Games are an undeniably important form of entertainment and a large business in today's world.

In a nutshell, game mods are the alteration of content from a video game that make it operate in a manner different from the original version. Mods are not standalone software, they operate on the already installed original game. Mods can modify some existing features of the game, add something entirely new, or may fix some bugs in the game. Creators of these mods include developers of the original game, users with enough technical expertise, or game development companies trying out their ideas to see if the users like them. In other words, modding is a "Do It Yourself" approach to technology personalization that can establish both socio-technical and distributed cognitions for how to innovate by taking control over technology design from their producers [6] [2].

Video Game Mods add replay value to the games [7]. Unlike mobile games, many of which are freemium, the traditional single-player games require a one-time purchase of the game. After the game is completed, the entertainment value is typically highly diminished. Newer games try to incorporate different storylines and different "achievements" in the game to add replay value. Mods, by nature, are an important tool for adding replay value and the role of the mod creators in the success of digital games cannot be denied, and it can be further harnessed [4]. Also, mods are almost always freely available. Many game companies appreciate the value of game mods [1] and provide tools and documentation to support their development.

Tens of thousands of mods exist for the more popular games and have millions of users. The data on mod creation and use may, therefore, provide a rich source of information to learn mod creator and user expectations, inform design of new features in future games or even suggest ideas for new games.

Here we propose a specific approach to leverage game mod data and use it to illustrate the potential for new insights such approach may bring. In this example we investigate the features that can make a mod popular. There is a common belief that mods are, in general, more attuned to the user expectations than a full game, since user-space for mods is much more restricted compared to the original games and as a result, the popularity of mods is determined by how well a particular mod can cater to the expectations of this rather niche community. This leads to following specific research questions:
RQ1. What type of mods are the creators making?
RQ2. What features make a mod popular?
RQ3. Do the creators create mods according to the demand of the community?

Approach
Our data driven approach starts from collecting the data on the mods from popular online fora. The focus on popularity of the mods and, in particular, popularity of the specific mod features, led to metrics (described below) we obtained from these websites.

Data Sources
The analysis is based on data collected from http://www.nexusmods.com, called Nexus hereafter. Similar

data could be obtained from the Steam Workshop and other sources. Nexus is one of the largest mod hosting websites with around 10 million registered users and over 100 thousand mod files in total. This website hosts and curates the mods for many games. This provides a uniform data collected over a long time period for a large and diverse collection of games. For each mod, Nexus has the data on the number of views, which is the number of people who viewed the individual mod's page; the Number of Downloads, the total number of times the mod was downloaded; the Number of Unique Downloads, the total number of times the mod was downloaded by unique users; and the Number of Endorsements, which is analogous to the number of users claiming to like the mod.

The website also has an extensive tagging system for every mod. Files can be tagged by both file authors and logged in members; if an author tags a file it becomes instantly tagged where as 3 normal members need to approve a tag for a mod before it becomes confirmed. A confirmed tag allows that mod to be found using the tag search (Ref: http://wiki.tesnexus.com/index.php/Tag).

To download a mod exceeding 2MB in size, one has to be a registered Nexus member. This helps Nexus to track the number of downloads and number of unique downloads for each mod. However, the website does not provide any API or any policy regarding data retrieval. We were able to extract the required data by collecting around 3000 pages in total and scraping data from the HTML source code of these pages.

Experimental Design: the Choice of Games
For our study we first chose **The Elder Scrolls V: Skyrim**, called Skyrim hereafter, developed by Bethesda Softworks, which is one of the most popular games on Nexus and had around 40,000 mods at the time of data collection, making it

a rich resource. From the experimental design perspective, we wanted to have both, a group of similar games (with similar player base) in our sample and a group of different games. For the first goal we selected other games in the same series (TES: The Elder Scrolls): **The Elder Scrolls IV: Oblivion**, called Oblivion hereafter, as a second game and **The Elder Scrolls III: Morrowind**, called Morrowind hereafter, as a third game. These two also have a large number of mods and the tags used in these games are almost the same.

Finally, we tried to generalize our findings by examining the mods for the games of the Dragon Age (DA) series: **Dragon Age Origins, Dragon Age 2** and **Dragon Age Inquisition**. Although the Fallout series games have more mods than the Dragon Age series games, we chose the Dragon Age series due to the fact that the parent company of both The Elder Scrolls series and Fallout series is same, whereas the Dragon Age series games are developed by Bioware. In this way we wanted to eliminate any unintended overlap in the types of mods.

Plans for a Survey
To augment the results found through the analysis surveys of mod users and mod creators will be conducted. In particular, the relationships among various metrics discovered in the analysis does not capture the motivations and causal relationships. Properly designed surveys would help in clarifying those points and, more generally, to validate these findings.

Theory and Metrics
We face several challenges oprationalizing mod feature space and popularity. We used tags to construct feature space, because Nexus provides an extensive and consistent tagging system and each tag is used to identify a

feature of a mod. We represent each mod i by a vector $f_{ij}, j = 1, \ldots, n$, where n is the number of tags available for that game and $f_{ij} = 1$ if the mod i contains tag j and zero otherwise. Four possible metrics could have been used as a measure of popularity: no. of views, downloads, unique downloads, and endorsements. All measures were found to be highly correlated for all games under consideration and we chose the number of unique downloads denoted as d_i because it captures only one download of a mod from each user independent of the number of times they may have downloaded the mod.

Findings and analysis

To investigate game features, we relied on the mod tags as noted above. We selected only the confirmed tags for our analysis. We found that the tags used for the mods for all games in the TES series and all games in the DA series were similar. The tags across the two series were found to be not all same, but conceptually similar. This helped us to formulate a generalized version of our findings using the TES series and DA series. For the mods that have no tags associated with them, we have added a tag "none".

RQ1: What type of mods are the creators making?
Generalizing the motivation of mod creators is a difficult task [8], so we decided to look at the features that are common among the mods to answer this question. As described above, the feature set of a mod is approximated by the set of tags that are specified by the mod developer. The distribution of counts per feature j is then simply $n_j = \sum_i f_{ij}$. We found that *English*, *New textures*, and *Gameplay Effects Change* are the most common three tags across all the games. The top three most commonly used tags in TES series apart from the former three are: *Lore Friendly, Fair and Balanced,* and *For Female Characters*. The top three popular tags in DA series other than the

Figure 1: TES Series Creator Tag Cloud.
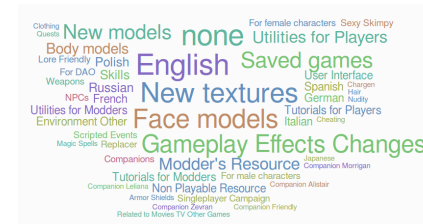
Figure 2: DA Series Creator Tag Cloud.

common three are: *Face Models, New Models,* and *Saved Games*. We also found that in all six games, a significant fraction (7% for Skyrim and DAO to 28% in DA2) of mods were untagged, thus appearing with the tag: none. Not surprisingly, percentage of untagged mods is more for relatively less popular games.

The Tag Clouds in Figures 1 and 2 are used to illustrate this. Here we only show the aggregated tag clouds for the two series. To quantify the similarity in the distribution of features we measured the cosine similarity between the distribution of n_j, counts per feature j, for each game and found the values ranging between 0.65 to 0.88 within the DA series, 0.77 to 0.95 in TES series and 0.65 to 0.89 when

comparing across the two series. This indicates a fairly close similarity not only within the same series of games, but also across the series. In particular, the large font corresponds to tags commonly used in mods and could be compared to popularity-related tags in the results of RQ2 with *Lore Friendly* tag being influential in mod popularity and commonly used in TES, but not in DA and *Gameplay Effects Change* being popular and common in both TES and DA series. The reason for using Cosine similarity here is that cosine similarity considers angle between the vectors and tends to be more stable than, for example, euclidean distance in high dimensions (common in text and feature analysis contexts). Here we are trying to report our subjective perception of similarity among groups together with numbers obtained for these groups.

RQ2. What features make a mod popular?
For each mod i we have the number of unique downloads d_i. We created two measures of high popularity: $Top_p(d_i) = 1$ if d_i is in the top p-percentile of all download values and zero otherwise. We fitted two generalized linear model $Top_1(d) \sim c + \sum_j c_j t_j$ and $Top_{10}(d) \sim c + \sum_j c_j t_j$, one indicating top 1% and another indicating top 10% of the mods where $t_j = \sum_i d_i f_{ij}$, is the sum of the numbers of unique downloads for each mod i having the tag j and c, c_j are model coefficients that explain how much the presence of tag j increases (or decreases if c_j is negative) the predicted probability that the mod is popular. The total number of mods varied from 39804 in Skyrim, to 954 in DA2 and the percentage of deviance explained varied from 64% to 20%. The highest correlation among the predictors was 0.69 between *Clothing* and *Clothing:Female* with other correlations being below 0.43. We first excluded tags that

were not found to explain significant amount of deviance [1] and were left with approximately 10 tags for each game. We picked the significance values based on the number of mods in each game, with larger games requiring higher significance (lower p-values). For example, for Skyrim with 39804 mods we used p-value< $2e^{-16}$ and for DA2 with 954 mods we used p-value < 0.17 .

Interpretation of Results
Looking at the most popular mods for all six games we can observe the following trends: Tags like *Gameplay Effects Change, New Models, New Textures*, and other visual effect enhancing mods are more popular across the games, so are the *Lore Friendly* mods. Mods with no tags, i.e. *None* tags tend to be unpopular. Perhaps such mods may reflect the lack of marketing effort/skill from the creator's perspective and the difficulty of finding such mod by interested users because a search by any tag will not return the mod as result.

Popular tags in TES series games are almost entirely focused on visual effect and performance enhancement while *Quest* and *User Interface* related mods have a positive effect on popularity in the DA series games. This might be caused by the different nature of the games in the two series. TES series games feature a wide open world to loot and conquer, with numerous quests and small stories, but the interactions with NPCs and the stories are not very deep. Since the player roams around this vast open world, visual effects tend to attract more attention. DA series games have a much more involved storyline and complex interactions with NPCs, but the world is more restricted and number of quests is also much less compared to the TES series games. Also, DA series games feature a number of

---

[1]Ref: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/add1.html

active skills players can use. So, having mods to improve upon these features make more sense. In general, our findings suggest that mods that improve the Look-and-Feel of a game, improve the ease of interaction with the game, and overall provide a more immersive gameplay experience tend to be more popular. This finding coincides with two of the five-factor model of user motivation described by Yee (2006) [9]: Immersion and Escapism. The other factors (Relationship, Achievement, and Manipulation) cannot be verified by our study due to the difference of context and type of games studied. Similar findings have been reported in [5], [3].

RQ3. What is the correspondence between the two sets? To measure the similarity between the set of features popular among users and the set of features commonly developed the creators, we measured the cosine similarity between $t_j = \frac{\sum_i d_{ij} f_{ij}}{\sum_i f_{ij}}$, or the average number of downloads for each feature j, and n_j or the counts of mods per feature j, for each game and found the values ranging from 0.2 for DAI and 0.19 for Oblivion to 0.38 for DAO and 0.36 for Skyrim. This suggests that there is some similarity between the type of mods the creators are making and the type of mods that are popular among users, but it is not very high. Notably, for the most popular games in the two series (DAO and Skyrim), there is a closer alignment between the two than for the less popular games. Is it because the mod creators can more accurately judge user's interest in popular games? Would it make sense for the mod creators to make mods with the most popular features if their objective is gaining more downloads? Or is it essential for the ecosystem to have a wide range of features, including features that are not widely used? If the number of mods for a popular feature is minimal, is it because the current mods are able to serve the purpose adequately? These and other questions are a subject of further research and could be answered if, for example, the time series of mod creation and popularity were obtained.

Summary

We propose a data driven approach to study creation and use of game mods. To illustrate the approach, we analyze data from the Nexus website on six games belonging to two series. We propose a way to project a mod into a feature space defined by its tags. We measure mod popularity via the number of unique downloads, the popularity of each feature among creators via the total number of mods per feature, and the popularity of each feature among the users via the sum of unique downloads of each mod having that feature. We have also propose the similarity measure between the features popular among users and features commonly incorporated by creators.

We have found a general trend among the feature spaces and highly uneven distribution of feature popularity. We have also discovered a similarity among the features popular among the users and the features often incorporated by mod creators, confirming that the mod market appears to be driven by and in general attuned to user expectations.

We hope to further validate the proposed approach of studying creation and use dynamics in game mods by conducting interviews with mod creators and users and list several questions of practical and theoretical importance that could be answered using our approach. More generally, we hope that study of game mods will lead to insights that would result in better gaming experience and more effective game production.

REFERENCES

1. Reina Y Arakji and Karl R Lang. 2007. Digital consumer

networks and producer-consumer collaboration: innovation and product development in the video game industry. *Journal of Management Information Systems* 24, 2 (2007), 195–219.

2. Caio Camargo. 2009. Modding: changing the game, changing the industry. *Crossroads* 15, 3 (2009), 18–19.

3. Henry Morlock, Todd Yando, and Karen Nigolean. 1985. Motivation of video game players. *Psychological reports* 57, 1 (1985), 247–250.

4. Hector Postigo. 2007. Of mods and modders chasing down the value of fan-based digital game modifications. *Games and Culture* 2, 4 (2007), 300–313.

5. Andrew K Przybylski, C Scott Rigby, and Richard M Ryan. 2010. A motivational model of video game engagement. *Review of general psychology* 14, 2 (2010), 154.

6. Walt Scacchi. 2010. Computer game mods, modders, modding, and the mod scene. *First Monday* 15, 3 (2010). http://firstmonday.org/article/view/2965/2526.

7. Walt Scacchi. 2011. Modding as an open source approach to extending computer game systems. In *Open Source Systems: Grounding Research*. Springer, 62–74.

8. Olli Sotamaa. 2010. When the game is not enough: Motivations and practices among computer game modding culture. *Games and Culture* (2010).

9. Nick Yee. 2006. The Demographics, Motivations and Derived Experiences of Users of Massively-Multiuser Online Graphical Environments. *PRESENCE: Teleoperators and Virtual Environments* 15 (2006), 309–329.