Tweeting your *Destiny*: Profiling Users in the Twitter Landscape around an Online Game

Günter Wallner*, Simone Kriglstein† and Anders Drachen‡

*Institute of Art & Technology, University of Applied Arts Vienna, Vienna Austria
Email: guenter.wallner@uni-ak.ac.at
†Center for Technology Experience, Austrian Institute of Technology, Vienna, Austria
Email: simone.kriglstein@ait.ac.at
‡DC Labs, University of York, York, United Kingdom
Email: anders.drachen@york.ac.uk

Abstract—Social media has become a major communication channel for communities centered around video games. Consequently, social media offers a rich data source to study online communities and the discussions evolving around games. Towards this end, we explore a large-scale dataset consisting of over 1 million tweets related to the online multiplayer shooter *Destiny* and spanning a time period of about 14 months using unsupervised clustering and topic modelling. Furthermore, we correlate Twitter activity of over 3,000 players with their playtime. Our results contribute to the understanding of online player communities by identifying distinct player groups with respect to their Twitter characteristics, describing subgroups within the *Destiny* community, and uncovering broad topics of community interest.

Index Terms—Twitter, topic modelling, community analysis, Destiny, MMOG, profiling, game analytics

I. Introduction

Building healthy player communities has long been recognized as an important factor for the market success of online games (cf. [1]). Studying the social connections within online games has thus seen a growing interest over the years (see, e.g., [2]). However, social activity has long escaped the boundaries of the games themselves and game-related discourse is now taking place across all sorts of online discussion platforms. Indeed, gamers constitute an active subgroup of social media users (cf. [3]) making social media, including micro-blogging services such as Twitter, a valuable source for studying gamer communities. Despite being a rich data source, research using Twitter to understand the public discussion around games is still relatively sparse (see Section II for some notable exceptions). In our own previous work [4] we proposed a visual analytics system and demonstrated the usefulness of such an approach for qualitatively analyzing tweets in the context of games user research using a large-scale dataset of tweets related to the game *Destiny* [5]. *Destiny* is a multiplayer first-person shooter set in a science fiction world where players compete against each other in different player vs player (PvP) and player vs environment (PvE) game modes. It is a suitable use case for this kind of analysis because it exposes gameplay data through a public API and was supported by the developer over an extended time period. In this paper, relying on the same dataset, we continue our efforts in exploring community engagement on Twitter by taking a quantitative approach and utilizing unsupervised techniques, specifically clustering and topic modelling by means of Latent Dirichlet allocation (LDA), for the purpose of:

- Assessing the relationship between, and establishing player profiles with respect to, tweeting behaviour and playtime;
- Uncovering subgroups within the *Destiny* community based on how people identify themselves in their public Twitter profile descriptions;
- Identifying broad topics of interest to the community, their variation over time, and if and how these interests differ between subgroups.

In that sense, this paper contributes to creating a better understanding of online communities which form around games. This can help to maintain and shape a strong player community and, in turn, extend the longevity of a game. Insights gained through the proposed analysis methodologies can point to new directions for analyzing motivations and to specific preferences of players.

II. Related Work

Player communities have been studied not only within games (see [2] for a good overview) but also increasingly outside of games. Due to the focus of this paper we will restrict the discussion to the latter category and on work employing quantitative methods for analyzing Twitter datasets. For example, McDonald and Moffat [6] conducted a sentiment analysis to gain insights into the positive and negative reactions to the yearly changing theme of the Global Game Jam by capturing and analyzing Twitter data. Chatzakou et al. [7] undertook a quantitative analysis of tweets in relation to the Gamergate controversy. For this purpose, a dataset of 340,000 users and 1.6 million tweets was analyzed to understand users' properties and to investigate the content they posted as well if these were different from random Twitter users. Kalaitzis et al. [3] presented a method for predicting gaming-related properties from Twitter profiles by applying machine learning models on labelled accounts.

In our own previous work [4], we used visual analytics methods to identify which topics matter to *Destiny* players, to find influential members of the game's community, and to untangle the relationship between contextual information about the tweets and in-game activity. However, no automated topic modelling was performed. In contrast, in the present paper we make use of topic modelling to extract key themes from a large corpus of tweets. Topic modelling has been used extensively in other domains (e.g., [8]–[10]) but, to our best knowledge, there exist only a few works within in the context of games research. These are usually focused on the analysis of chat logs but not Twitter data. For example, Tinati et al. [11] analyzed over two years of gaming and chat data, comprised of over 4 million completed game sessions and 885,000 chat entries, from the citizen science game *EyeWire* to better understand how features of the game contribute to player communication. To examine the content of the chat messages topic modelling was performed. This helped them to identify the vocabulary as well as the discourse that occurred during the different stages of the game. Musabirov et al. [12] used text mining techniques, including topic modelling on game chat logs to improve tutorials in massively multiplayer online games. In order to investigate how the gender of streamers is associated with the nature of conversation, Nakandala et al. [13] analyzed chat messages from Twitch. Ryan et al. [14] presented GameNet, a tool for game discovery build upon a latent semantic analysis of Wikipedia articles for nearly 12,000 digital games to establish the semantic relatedness between them. The article also offers a comprehensive review of the use of natural language processing techniques in the context of game studies.

III. DATA SET

The analysis presented in this paper is based on a dataset used in previous work [4] and which contains 1,062,230 tweets with relation to *Destiny* within a time period of about 14 months (from June 9th, 2016 to July 17th, 2017) and details of 381,076 accounts tweeting about *Destiny*. It was obtained using the Twitter Search API [15] and the following query:

```
(destiny AND (e3 OR activision OR bungie OR crucible OR crota
OR guardians OR ironbanner OR psn OR xbl OR gamertag OR
gameplay)) OR riseofiron OR destinygame OR destinythegame OR
gjallarhorn
```

The search term Destiny was combined with different game-related key words to reduce the odds of collecting tweets unrelated to the game. Besides the corpus of tweets, the dataset includes *Destiny* in-game behavioral data over a six-month period (June 10th, 2016 to December 1st, 2016) from 3,547 players who shared their PSN ID or Xbox Gamertag in their Twitter profile description. These IDs were extracted from the profile description using various regular expressions (cf. [4]). While the dataset contains a variety of in-game behavioral features, we will focus on a subset of those which are related to the overall time played during the above period, that is, time spent playing in the various PvP and PvE modes that *Destiny* offers. This information was gathered through the Bungie API [16] using the gamertags and PSN IDs provided in the profile descriptions. This allows us to assess if gameplay duration is connected to tweeting behaviour.

IV. TWEET BEHAVIOUR

To asses if and how activity on Twitter corresponds with in-game activity (i.e., time spend playing) we first focus on the 3,547 Twitter accounts for which in-game data is available. Specifically, we extracted the total time spend playing (PvP and PvE combined) and measures of Twitter activity (number of tweets, retweets, and replies as well as the total number of people a user is following) and indicators of popularity (number of followers, average number of retweets a tweet receives, average number of likes a tweet receives) in order to correlate the measures and to develop player profiles through k-means clustering.

A. Data Preprocessing

Of the 3,547 accounts for which gameplay data is available, 114 did not tweet during the six-month period and were thus excluded from analysis. Since k-means clustering is known to be sensitive to outliers (cf. [17]) and because some players exhibited very large values along one of the features (e.g., comparatively very high number of followers or play duration), we also removed entries with a z-score greater than ± 3 (i.e. removing the outlying 0.15% of both tails of the frequency distribution) in at least one column.

Of the 314 accounts removed, 253 formed an outlier among one feature and 61 formed an outlier on two or three features. The main contributor to outliers was the average number of retweets with 74 outliers, followed by the number of tweets with 69 outliers. However, all measures had outliers removed. The outliers thus mainly represent Twitter users who were exceptionally active or inactive along one dimension. This process left 3,119 accounts for analysis. Finally, given the different scales of the various metrics, we used min-max normalization to scale each metric to the range $[0...1]$.

B. Analysis and Results

A Spearman rank correlation did only yield weak correlations between the Twitter related measures and in-game activity with correlation coefficients between -0.15 and 0.18. Using scikit-learn [18], k-means clustering was performed with $k = 2...15$ and the resulting clusterings were assessed using the silhouette coefficient, pointing to a seven-cluster solution as described in the following. Figure 1 shows how the included measures vary between the seven identified profiles.

Profile 1 - Minimally Active: This cluster constitutes the largest group covering 43% of the whole sample ($N = 1,343$) and is comprised of people who show low activity on Twitter, have a small number of followers, and have (together with Profile 7) the lowest duration of play.

Profile 2 - Highly Active Gamers: This cluster includes Twitter users ($N = 644$) who spend considerable time playing *Destiny*, in fact their play duration is the highest across all profiles. These can be considered highly active gamers but
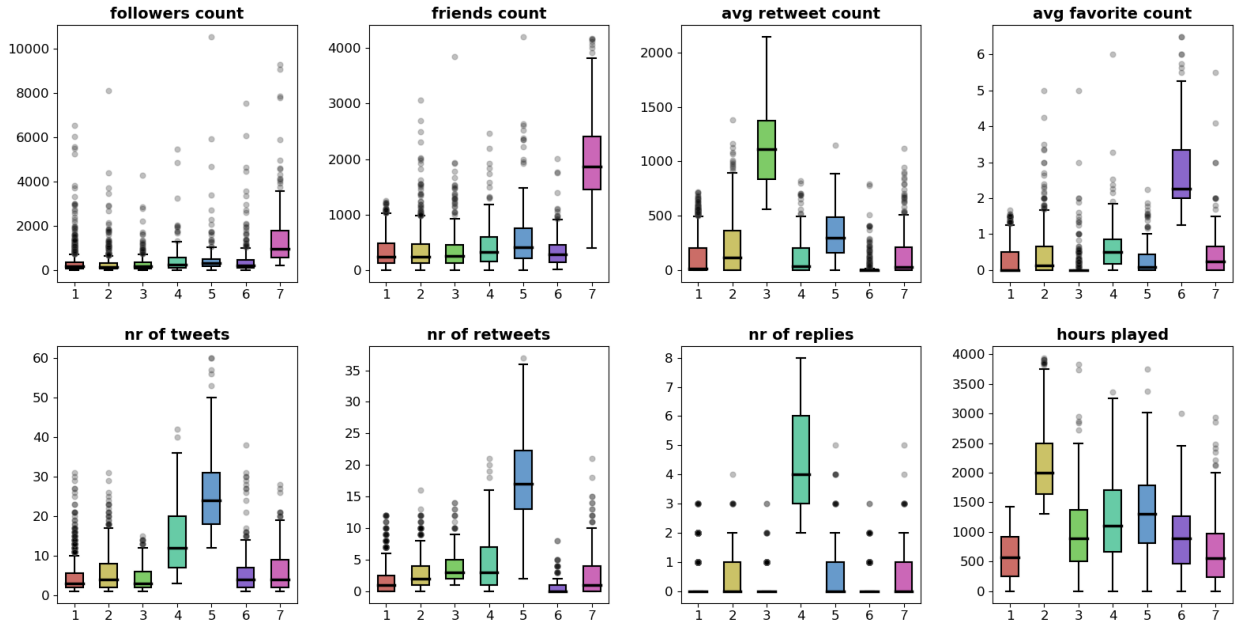
Fig. 1. Variation of various Twitter related features and play duration across the seven identified clusters.

at the same time they do not engage much on Twitter, with tweeting related measures being similar to the average user of Profile 1.

Profile 3 - Influencers: Tweets of players in this group ($N = 320$) are often shared as they receive on average a large number of retweets. While they show less tweeting activity than players in Profile 5 they contribute to information diffusion to a much larger extent.

Profile 4 - Respondants: Players in this group ($N = 149$) are also quite active on Twitter in terms of the number of tweets and retweets (but less so than players belonging to Profile 5). However, they are primarily characterized by making a large number of replies compared to other profiles.

Profile 5 - Socially Active: This group of players ($N = 192$) is not only very active in the game (but considerable less than players belonging to Profile 1) but are also highly active on Twitter: they tweet a lot about *Destiny* and make many retweets. At the same time and despite being highly active on Twitter they do not attract as many retweets on average as players belonging to Profile 3.

Profile 6 - Favourites: Tweets from players in this group receive on average the most likes (avg. favorite count) but at the same time receive the lowest average number of retweets across all profiles ($N = 265$).

Profile 7 - Followers: Players in this group ($N = 206$) show – together with players of Profile 1 – the least in-game activity across the identified clusters. Similarly, they are also among the less active users on Twitter. What sets them apart from other profiles is that they follow a considerable larger number of people (friend count) and also have a larger number of followers than players belonging to the other profiles. These seem to be players that act as 'information distributors' which follow other people to gather information and share it with

there followers. These tweets are, however, considered less worthy of retweeting.

V. TOPIC MODELLING

Topic modelling is the process of finding topics (i.e., groups of words) which frequently occur in a collection of documents. For our analysis we have opted for Latent Dirichlet allocation (LDA) [19] as it is widely used in natural language processing and produces topics that are well interpretable by humans (cf. [20]). LDA is a probabilistic model, modelling each document as a mixture of latent topics, with the topics being represented by a distribution over words. In other words, LDA takes as input a collection of documents assumed to contain k topics. Each document belongs to the k identified topics with differing probabilities.

A. Data Preprocessing

We first filtered out all accounts with empty or non-English profile descriptions. For the latter, we followed the recommendation of Lui and Baldwin [21] and used a majority vote across the three language detection algorithms `langid` [22], `langDetect` [23], and `CLD2` [24]. That is, only accounts where at least two of the three algorithms classified the description as English were retained. Next, we removed fake accounts such as multiple accounts with exactly the same profile description or where to description was clearly suspicious (e.g., several accounts claiming to give away smartphones). In case of tweets, we dropped non-English tweets (following the procedure above), duplicate tweets, tweets with exactly the same text but including different URLs, tweets from the fake accounts identified before, and retweets. This resulted in 352,088 tweets and 223,459 accounts for topic modelling.

Label	c_v	Top Keywords[†]	No. and % of Profiles	
Streamers	0.579	twitch_streamer, youtube, youtube_channel, twitch, year_old, streamer, channel	25,245	11.3%
Gaming Enthusiasts	0.516	xbox_one, destiny, call_duty, xbox, psn, ps4, follow, follow_back, one, player	19,179	8.6%
Hobby Streamers	0.442	full_time, part_time, time, come_check, let_play, full, rainbow_six, part	16,228	7.3%
Gamers	0.430	video_game, game, video, love, like, play, life, thing, make, gaming	44,842	20.1%
News Outlets	0.420	gaming_news, gaming, news, subscribe_youtube, social_medium, gaming_community	22,991	10.3%
Students & SciFi Fans	0.390	web_chat, star_war, college_student, cod_player, add_xbox, come_join, star	14,696	6.6%
Casual Gamers	0.382	gamer, video_game, fan, lover, music, game, husband_father, sport, gaming	32,361	14.5%
Creative People	0.347	destiny_player, content_creator, graphic_designer, like_body, live_chat, coming_soon	15,621	7.0%
Official Accounts	0.345	official_twitter, twitter, official, business_inquiry, twitter_account, playing_video	15,409	6.9%
Clans & Members	0.280	proud_member, check_channel, destiny_clan, one_day, make_sure, member	16,887	7.6%

[†]Keywords are sorted by decreasing probability.   Words connected by underscores represent bigrams.

Following established procedures, tweet texts and profile descriptions were tokenized. Subsequently, all tokens were converted to lower case and punctuation characters were removed. Tokens representing numbers, mentions of other accounts (i.e., tokens starting with an @), and URLs were removed as were tokens with less than three characters and tokens representing stopwords. All remaining tokens were lemmatized to obtain their base form and to reduce data sparseness using the WordNet lemmatizer provided by NLTK [25]. Finally, tokens appearing in more than 80% of the documents as well as tokens only occurring in less than 3 documents were removed.

B. Analysis

We performed LDA on both the corpus of tokenized and lemmatized profile descriptions and tweet texts (extended with bigrams and trigrams of the tokens) using the Python library `gensim` [26]. A TF-IDF [27] term-weighting bag-of-words model was used to adjust for the fact that some words occur more frequently across documents (e.g., *Destiny* in our case). LDA models where calculated from 2 to 20 topics for profile descriptions and 2 to 200 topics for tweet texts. Lastly, each profile description and tweet was assigned to the most probably topic.

To rate the quality of the topics, we assessed the coherence of each model using the c_V coherence score proposed by Roder et al. [28] and which has been shown by the authors to align well with human judgment. After inspection of the resulting plots and inspection of the topics for interpretability we opted for a 10 topic solution for the profile descriptions. In case of tweets, coherence did not show a distinctive elbow but increased slowly until around 175 topics after which it flattened out. For that reason, we also estimated the perplexity of each topic model showing a steep decline between 76 and 84 topics, thus either pointing to a 84 or 175 topic solution. While such a large number of topics can be expected (assuming that a broad range of issues is discussed by the community) it is also known that – in such a case – LDA tends to produce many 'junk' topics (cf. [29], [30]), that is, topics which do not capture the underlying semantics well and are thus hard to interpret by humans. After manual inspection of both models we found that the 175 solutions does not add much information

to the 84 solution which itself already contains topics that are difficult to interpret. To keep the following analysis focused on the more common and interpretable topics we thus only retain those with a topic coherence score greater than the average ($>$ 0.304) and which encompass a higher than average number of tweets ($> 4,191$).

C. Results

Table I lists the profile topics together with the topic coherence score c_v, a list of most likely keywords, as well as the number and percentage of profiles belonging to each topic. Bi-grams are represented using an underscore. Similarly, Table II lists the same information for topics inferred from the tweets. Focusing first on the profile descriptions we could derive the following topics on how people identify themselves on Twitter:

Gaming Enthusiasts: This group frequently and predominantly uses gaming-related terms such as Xbox and Xbox One, PSN, PS4, and different games such as *Call of Duty* to describe themselves on Twitter. For this reason we have coined them gaming enthusiasts.

Gamers: These constitute the largest group within the *Destiny* audience, expressing their love for video games or how they like playing video games in their profile descriptions. These are people who we consider avid gamers but who use less specific terms or games than *Gaming Enthusiasts* to express their interest.

Casual Gamers: These are people who do not identify themselves only as gamers in their profiles but for which gaming is one hobby next to, for example, music or sports.

Clans & Members: Given the multiplayer focus of *Destiny* and the clan culture surrounding the game it is interesting to explicitly observe a number of Twitter accounts which, for example, belong to clans or to people who identify themselves as members of a clan.

Students & SciFi Fans: People identifying first and foremost as students (but with a certain gaming affinity) and/or science fiction fans (e.g., *Star Wars*) constitute another group within the *Destiny* community. Given that *Destiny* takes place in a science-fiction setting it is fitting to see that the game resonates with the latter group.

TABLE II
PREVALENT TOPICS IN TWEETS AS IDENTIFIED BY LDA ANALYSIS.

Label	c_v	Top Keywords†	No. and % of Tweets	
Iron Banner	0.463	iron_banner, banner, iron, ironbanner, supremacy, control, surprised, destiny, gamestop	5,785	1.6%
Rise of Iron	0.438	rise_iron, rise, iron, bro, e32017, dlc, destiny, via, riseofiron, gameplay	7,782	2.2%
Trials of Osiris	0.503	trial_osiris, trialsofosiris, last_night, trial, add, night, flawless, osiris, add_psn	7,385	2.1%
Live Streaming	0.470	live, supportsmallstreamers, twitch, destinythegame, playing, stream, morning, destiny	8,019	2.3%
Announcing Videos	0.510	added_video, added, playlist, video, gameplay, basically, episode, story_mission, podcast	4,413	1.3%
Liking Videos	0.394	liked_video, liked, video, gameplay, really_hope, fam, agree, destiny, crucible, hope	4,906	1.4%
Announcements	0.380	release_date, gameplay_trailer, release, date, trailer, sorry, seems, gameplay, called	4,684	1.3%
Gameplay	0.320	gameplay_reveal, reveal, gameplay, cayde, via, destiny, grinding, idk, watch, red_bar	5,492	1.6%
Infrastructure	0.320	server, destiny, wtf, riseofiron, ps4live, queue, happened, tapir, destinyriseofiron	6,225	1.8%
(Dis)Likes/Observations	0.381	look_like, look, feel_like, like, gary, next_week, feel, holy_shit, birthday, holy	6,718	1.9%
Various	0.327	life, boy, reason, bought, seen, mine, hoping, crucible_highlight, gave, good_luck	4,208	1.2%

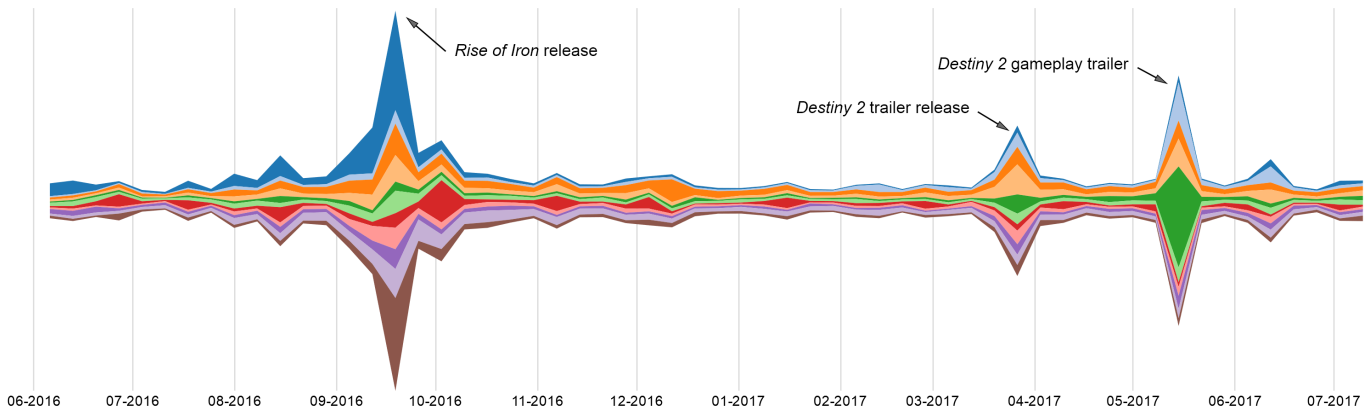†Keywords are sorted by decreasing probability.   Words connected by underscores represent bigrams.

Fig. 2. Streamgraph of the 11 most coherent and frequent topics over time (June 9th, 2016 to July 17th, 2017, ● Rise of Iron, ● Announcements, ● Live Streaming, ● (Dis)Likes/Observations, ● Gameplay, ● Liking Videos, ● Iron Banner, ● Various, ● Announcing Videos, ● Trials of Osiris, ● Infrastructure).

Streamers: Within this topic fall players who describe themselves as streamers (e.g., Twitch) or who are maintaining a gaming-related channel, for example, on YouTube.

Hobby Streamers: This group is related to the above as people also express their interest in streaming, but are not as engaged in it judging from the keywords (e.g., part-time). We thus labelled this group as hobbyists.

News Outlets: These accounts are related to some sort of news outlets such as, for example, gaming websites, magazines, or YouTube channels that cover gaming-related news.

Creative People: *Destiny* seems to also attract – what we coined – creative people, that is, people such as graphic designers or content creators which also highlight their interest in the game in their public profile description.

Official Accounts: Profile descriptions within this topic point to various official accounts or accounts of businesses which tweet about *Destiny* or where, for example, the business is related to gaming.

Shifting the focus to topics identified in the tweets themselves (cf. Table II) we witness a number of topics pertaining to in-game events, including the weekly *Trials of Osiris* event and the once-per-month held *Iron Banner* as well as to game extensions such as *Rise of Iron* (released on September 20th, 2016). For example:

> Time for some iron banner. Let's play!!! #xbox #xboxone #bungie #destiny #destinythegame #RiseofIron...

Other topics evolve around news or announcements regarding the game such as trailers, release dates, or gameplay related discussions such as the reveal of new gameplay material, e.g.:

> Only 14 days until the Destiny 2 gameplay reveal. I hope they show the Crucible, and any changes Bungie made!

Streaming and the creation and watching of videos are also prevalent topics, with people announcing that they are currently live streaming, as exemplified by the following tweet

> I'm live on Twitch - Watch me at #SupportSmallStreamers #destiny #riseofiron Clan raid and helps

or that they have uploaded new videos with others, in turn, expressing that they liked certain videos. Issues related to what we called *Infrastructure* including server related issues, queuing for a match, or online platforms (e.g., PlayStation Live) form another common topic.

People also frequently express what they like or do not like about the game or make observations, for instance:

> I was so hyped for destiny2, but all the gameplay just looks like another expansion of the old destiny. So might not even get it

Lastly, we find a topic (entitled *Various* in Table II) which is hard to grasp based on the keywords and assigned tweets and which covers quite a broad range of different issues such as people tweeting about how *Destiny* is affecting their lives, such as:

> This is my life now. Good thing I don't have plans tomorrow. #RiseOfIron #destinythegame #Destiny #PS4share

Figure 2 shows how these topics vary over time. As evident from the figure, tweets surrounding the *Rise of Iron* extension increase steadily until the release day but after about two to three weeks interest was low again. That is, while the extension seems to have been highly anticipated, interest seems to have faded quickly which might be in line with the short duration of the extension, ranging from only a couple of hours to about 20 hours.[1] At the same time we can see a large increase of tweets related to *Infrastructure* at the day of the *Rise of Iron* release. Inspecting tweets falling within this topic on this day shows that the release was accompanied by server issues with people complaining about servers being down, for instance:

> @DestinyTheGame Servers are down :^(

Returning to topics related to in-game events, the graph also shows how conversation around the *Iron Banner* is peaking once per month (at the time it is held). In terms of streaming we can see increased activity around the *Rise of Iron* release but even more so in mid-December, 2016 around the time when *The Dawning* went live on December, 13th.

> Happy dawning guardians video and stream will be going on after I get off of work today!!! #Destiny #thedawning

While *The Dawning* in-game event lasted for about 3 weeks, tweets about streaming *Destiny* where soon back to 'normal' a few days after it started. In contrast, the Halloween event *Festival of the Lost* held in the last week of October and first week of November did not attract noticeable attention from streamers.

In relation to the *Gameplay* topic, we can observe the largest activity in mid-May, 2017 (when a *Destiny 2* trailer with actual gameplay footage was released) with some increased activity since the announcement and release of the first *Destiny 2* trailer about one and half months earlier (end of March, 2017). Announcements (a topic closely related to the previous one) are also increased during that time, peaking at the aforementioned two time periods. Expressing ones opinions (liking or disliking something) also strongly coincides with new releases (such as *Rise of Iron*) or announcements (e.g., *Destiny 2* trailer). That is, the community immediately expresses its thoughts and opinions with respect to new announcements.

In terms of announcing and liking videos there is generally no strong variation over time but a somewhat increased activity can be witnessed after the *Rise of Iron* release with people immediately starting to share new gameplay footage, among others walkthrough videos of the new content:

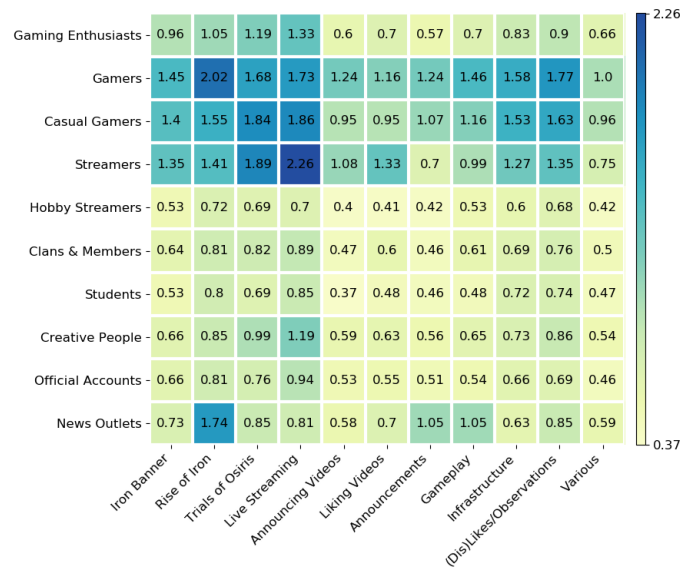|	Iron Banner	Rise of Iron	Trials of Osiris	Live Streaming	Announcing Videos	Liking Videos	Announcements	Gameplay	Infrastructure	(Dis)Likes/Observations	Various
Gaming Enthusiasts	0.96	1.05	1.19	1.33	0.6	0.7	0.57	0.7	0.83	0.9	0.66
Gamers	1.45	2.02	1.68	1.73	1.24	1.16	1.24	1.46	1.58	1.77	1.0
Casual Gamers	1.4	1.55	1.84	1.86	0.95	0.95	1.07	1.16	1.53	1.63	0.96
Streamers	1.35	1.41	1.89	2.26	1.08	1.33	0.7	0.99	1.27	1.35	0.75
Hobby Streamers	0.53	0.72	0.69	0.7	0.4	0.41	0.42	0.53	0.6	0.68	0.42
Clans & Members	0.64	0.81	0.82	0.89	0.47	0.6	0.46	0.61	0.69	0.76	0.5
Students	0.53	0.8	0.69	0.85	0.37	0.48	0.46	0.48	0.72	0.74	0.47
Creative People	0.66	0.85	0.99	1.19	0.59	0.63	0.56	0.65	0.73	0.86	0.54
Official Accounts	0.66	0.81	0.76	0.94	0.53	0.55	0.51	0.54	0.66	0.69	0.46
News Outlets	0.73	1.74	0.85	0.81	0.58	0.7	1.05	1.05	0.63	0.85	0.59

Fig. 3. Tweet topic distribution (x-axis) across the 10 identified profile groups (y-axis) in percentages (60,469 tweets in total).

> I added a video to a @YouTube playlist https://t.co/3iXEL8uu8Z Destiny Rise of Iron Download Complete Walkthrough Gameplay Story

The other topics show no notable variations during the analyzed time period.

Lastly, we investigated if and how the different identified groups differ in terms of which topics they are tweeting about. Figure 3 shows the percentage of tweets falling within each topic across the ten identified user groups. The most active groups with respect to the above tweet topics are *Gaming Enthusiasts*, *Casual Gamers*, and *Streamers*. Streamers unsurprisingly tweet the most about streaming followed by different game-related events, with the *Trials of Osiris* (a weekly event) garnering particular attention among streamers, more than *Rise of Iron* or the *Iron Banner*. The higher interest of streamers in the *Trials of Osiris* could be a result of the more frequent conduction of the event, compared to the *Iron Banner* which is only held once per month and *Rise of Iron* for which – as discussed above – interest seemed to have faded quickly after its release. *News Outlets* showed the highest interest in the *Rise of Iron* extension (the major addition to the game in the observed time period), *Announcements*, and *Gameplay* material. *Infrastructure* related issues are most important for *Gamers* and *Casual Gamers* but interestingly not so much for *Gaming enthusiasts*. Otherwise we could not observe strong variations: *Gamers* and *Casual Gamers* show a similar pattern, as do *Hobby Streamers*, *Clans & Members*, *Students & SciFi Fans*, *Creative People*, and *Official Accounts*.

VI. DISCUSSION

Focusing on a subset of Twitter users which shared their Xbox Gamertag or PSN ID we developed seven profiles of players based on their Twitter characteristics and in-game

activity. While the largest fraction (Profile 1) shows only limited Twitter activity there is a much smaller group of very active players with respect to retweeting (Profile 5). These can be considered to mainly contribute to information spreading (perhaps because they are highly enthusiastic about the game judging from their duration of play) but they do not necessarily receive a lot of retweets themselves. While these are very active, there is another group of players (Profile 3) which can be considered more influential as their tweets receive more attention and are shared more widely. These are, however, not necessarily the ones with the most followers (which would be players belonging to Profile 7). This reflects the findings of Cha et al. [31] who found that the number of followers is not related to the number of retweets, which is largely driven be the attributed content value. Besides there is also a group of players (Profile 6) who do not receive many retweets at all but at the same time garner a lot of likes. People appear to like the content these players post but at the same time deem it less worthy to distribute the information further. Others (Profile 4), in turn, seem to like to engage in discussions and replying to others. Particularly noticeable is that players spending the most time in *Destiny* show comparable low Twitter activity (Profile 2). In a recent study by Canossa et al. [32], influencers in in-game social networks (in terms of network centrality) showed only about one-fourth of the playtime of power players (i.e. the most engaged players). A similar phenomenon can be found in our study: the most engaged players (in terms of playtime) are not necessarily the most active or influential on Twitter.

Using LDA we grouped the profile descriptions into ten reoccurring topics to get an impression of the demographics of the community. These topics can be grouped into larger categories: people identifying themselves as gamers to various degrees, people playing games but considering them first and foremost not as gamers, streamers, news outlets, and official accounts. It needs, however, to be acknowledged that the identified profiles are based on descriptions of how people see themselves, that is, someone using a lot of gaming related keywords may actually not be a professional gamer. That said, such a categorization can be helpful to profile the audience a game attracts and to offer content which matches the interest of specific subgroups.

In terms of tweet content we identified several common topics as listed in Table II. These topics are related to in-game content (e.g., events, extensions), streaming and video sharing, infrastructure such as server issues, expressing opinions, and news concerning the game. It has to be kept in mind that this list of topics is not exhaustive as we have focused on the most coherent and frequent topics. Comparing these topics with the themes identified by Drescher et al. [4] – who used a visual analytics approach and did not rely on unsupervised topic modelling but on the same dataset – we can observe some differences. In general, Drescher et al. identified more specific issues such as people looking for others to play with or people sharing walkthroughs. While LDA also produced a topic related to matches described by keywords such as

private_match, match, took, and private ($c_v = 0.326$, 4,179 tweets) the associated tweets covered all aspects related to matches such as players being excited about the introduction of private matches or posting about playing a match. Similarly, walkthroughs were not grouped into a single topic but were part of different topics. The topics produced by LDA should be viewed as broad themes, rather than encompassing specific aspects. In this sense, topic modelling allowed us to extract greater themes which can serve as a starting point to identify more specific issues. Further work may explore the use of (semi-)supervised methods, e.g., targeted topic modelling [30].

Plotting topics over time offers a means to study trends, to monitor audience interest and anticipation, and also to observe for how long the interest pertains. Quickly fading interest in, for example, events or extensions may provide an indication for when to release new content in order to sustain interest in a game. It is noticeable that interest vanishes quite quickly in many cases (e.g., with respect to new game content) which also affects streaming activity (as in case of *The Dawning*). In light of this, regular events (e.g., *Iron Banner*) may be a good way to keep the talk about a game going. Monitoring topics could also be an efficient means to detect issues and changes in audience attitudes and sentiment.

When interpreting the results of this study some limitations need to be highlighted. First, due to the query that was used to retrieve the tweets some topics (e.g., *Rise of Iron*) may be more prevalent than others. These additional constraints were added because if Twitter is only queried for `Destiny` it returns too many tweets totally unrelated to the game. If searched only for the official hashtag `#DestinyTheGame`, the returned tweets only cover a small subset of *Destiny* related tweets. That is, constructing an unbiased query which still ensures that the tweets relate to the game is a challenging task. In addition, the Twitter Search API used to construct the dataset is focused on relevance and not completeness.[2] However, as the sampling is not detailed it is unclear if and which bias is introduced by the API. Please note, that each tweet and profile description was associated with the most probably topic. If a document belongs to different topics with low but similar probabilities such a hard assignment may not be nuanced enough. Thus, data on quantity (e.g., number of tweets falling within each topic) should be interpreted as indicative rather than definite. While we opted for LDA because of the human-interpretability of topics (cf. [20]), others (e.g., [33], [34]) cautioned that LDA might not work well with Twitter and produce incoherent topics due to tweets being very short. This was, for example, evident with the topic labelled *Various* (cf. Table II) but by focusing on the topics with the highest coherence scores we could alleviate this issue in general.

VII. CONCLUSIONS

In this paper we explored a longitudinal dataset of tweets using unsupervised clustering and topic modelling to characterize the Twitter community formed around the multiplayer

---

[2]https://developer.twitter.com/en/docs/tweets/search/overview/standard.html (Accessed: March 2019)

online shooter *Destiny*. Our results indicate that there are distinct player groups with respect to their tweeting activity, but that Twitter-related features are not directly correlated with playtime. We identified several prevalent topics, including in-game content, streaming, infrastructure, news, and opinions. Further insights could be gained by looking into less common topics within the topic model. We also identified several subgroups of which the game's audience is composed of. However, apart from a few exceptions such as streamers and news outlets, these subgroups show a similar tweeting pattern with respect to the identified topics. While we feel that topic modelling can be an important tool for studying online communities, automatically extracted topics may not align with a specific framing of an issue an analyst may have in mind. In this sense, a combination of a human-driven visual analytics approach in combination with semi-supervised topic modelling techniques appears to be a promising direction for future work.

REFERENCES

[1] C. Ruggles, G. Wadley, and M. R. Gibbs, "Online community building techniques used by video game developers," in *Entertainment Computing - ICEC 2005*, F. Kishino, Y. Kitamura, H. Kato, and N. Nagata, Eds. Springer, 2005, pp. 114–125.

[2] H. Warmelink and M. Siitonen, "Player communities in multiplayer online games: A systematic review of empirical research," in *Proc. of DiGRA 2011 Conference: Think Design Play, Digital Games Research Association*, 2011.

[3] A. Kalaitzis, M. I. Gorinova, Y. Lewenberg, Y. Bachrach, M. Fagan, D. Carignan, and N. Gautam, "Predicting gaming related properties from Twitter profiles," in *Proc. of the IEEE 2nd International Conference on Big Data Computing Service and Applications*. IEEE, 2016, pp. 28–35.

[4] C. Drescher, G. Wallner, S. Kriglstein, R. Sifa, A. Drachen, and M. Pohl, "What moves players?: Visual data exploration of Twitter and gameplay data," in *Proc. of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 2018, pp. 560:1–560:13.

[5] Bungie, "Destiny," Game [PlayStation 3, PlayStation 4, Xbox 360, Xbox One], September 2014, Activision, Santa Monica, CA, USA.

[6] B. McDonald and D. Moffat, "Using sentiment analysis to track reaction to the Global Game Jam theme," in *Proc. of the International Conference on Game Jams, Hackathons, and Game Creation Events*. ACM, 2016, pp. 50–53.

[7] D. Chatzakou, N. Kourtellis, J. Blackburn, E. De Cristofaro, G. Stringhini, and A. Vakali, "Measuring #gamergate: A tale of hate, sexism, and bullying," in *Proc. of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 1285–1290.

[8] A. Nigam, H. K. Dambanemuya, M. Joshi, and N. V. Chawla, "Harvesting social signals to inform peace processes implementation and monitoring," *Big Data*, vol. 5, no. 4, pp. 337–355, 2017.

[9] K. W. Prier, M. S. Smith, C. Giraud-Carrier, and C. L. Hanson, "Identifying health-related topics on Twitter," in *Proc. of the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Springer, 2011, pp. 18–25.

[10] D. Surian, D. Q. Nguyen, G. Kennedy, M. Johnson, E. Coiera, and A. G. Dunn, "Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection," *Journal of Medical Internet Research*, vol. 18, no. 8, 2016.

[11] R. Tinati, M. Luczak-Roesch, E. Simperl, N. Shadbolt, and W. Hall, "'/command' and conquer: Analysing discussion in a citizen science game," in *Proc. of the ACM Web Science Conference*. ACM, 2015, pp. 26:1–26:10.

[12] I. Musabirov, P. Okopny, and D. Bulygin, "Analyzing chat logs in online games for tutorial improvement," in *Proc. of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, 2015, pp. 661–666.

[13] S. C. Nakandala, G. L. Ciampaglia, N. M. Su, and Y.-Y. Ahn, "Gendered conversation in a social game-streaming platform," in *Proc. of 11th International AAAI Conference on Web and Social Media*. AAAI, 2017, pp. 162–171.

[14] J. O. Ryan, E. Kaltman, M. Mateas, and N. Wardrip-Fruin, "What we talk about when we talk about games: Bottom-up game studies using natural language processing," in *Proc. of the 10th International Conference on the Foundations of Digital Games*, 2015.

[15] Twitter, Inc., "Twitter search API," 2017, https://dev.twitter.com/rest/public/search (Accessed August, 2017).

[16] Bungie Inc., "Documentation for bungienet.platform.destinyservices," https://www.bungie.net/platform/destiny/help/ Accessed: March, 2019.

[17] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[20] J. Chang, J. Boyd-Graber, S. Gerrish, C. Wang, and D. M. Blei, "Reading tea leaves: How humans interpret topic models," in *Proc. of the 22nd International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2009, pp. 288–296.

[21] M. Lui and T. Baldwin, "Accurate language identification of Twitter messages," in *Proc. of the 5th Workshop on Language Analysis for Social Media*. Association for Computational Linguistics, 2014, pp. 17–25.

[22] ——, "Langid.py: An off-the-shelf language identification tool," in *Proc. of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 2012, pp. 25–30.

[23] S. Nakatani, "Language detection library," http://www.slideshare.net/shuyo/language-detection-library-for-java. (Accessed: March, 2019).

[24] R. Al-Rfou, "Pycld2 - Python bindings for compact language detector 2," 2015, https://github.com/aboSamoor/pycld2 (Accessed: January, 2019).

[25] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.", 2009.

[26] R. Řehůřek and P. Sojka, "Software framework for topic modelling with large corpora," in *Proc. of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, 2010, pp. 45–50.

[27] G. Salton and M. J. McGill, *Introduction to Modern Information Retrieval*. New York, NY, USA: McGraw-Hill, Inc., 1986.

[28] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proc. of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 399–408.

[29] S. I. Nikolenko, S. Koltcov, and O. Koltsova, "Topic modelling for qualitative studies," *J. Inf. Sci.*, vol. 43, no. 1, pp. 88–102, 2017.

[30] S. Wang, Z. Chen, G. Fei, B. Liu, and S. Emery, "Targeted topic modeling for focused analysis," in *Proc. of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 1235–1244.

[31] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in Twitter: The million follower fallacy," in *Proc. of the Fourth International AAAI Conference on Weblogs and Social Media*. AAAI Publications, 2010, pp. 10–17.

[32] A. Canossa, A. Azadvar, C. Harteveld, A. Drachen, and S. Deterding, "Influencers in multiplayer online shooters: Evidence of social contagion in playtime and social play," in *Proc. of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, pp. 259:1–259:12.

[33] R. Mehrotra, S. Sanner, W. Buntine, and L. Xie, "Improving LDA topic models for microblogs via tweet pooling and automatic labeling," in *Proc. of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2013, pp. 889–892.

[34] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li, "Comparing Twitter and traditional media using topic models," in *Advances in Information Retrieval*, P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Mudoch, Eds. Springer, 2011, pp. 338–349.