Condensing Steam: Distilling the Diversity of Gamer Behavior

Mark O'Neill, Elham Vaziripour, Justin Wu, Daniel Zappala
Department of Computer Science
Brigham Young University
mto@byu.edu, elhamvaziripour@byu.edu, justinwu@byu.edu, zappala@cs.byu.edu

ABSTRACT

We present a comprehensive measurement study of the Steam gaming network, the first complete examination of any major gaming network, comprising all 108.7 million user accounts and 384.3 million owned games. We examine gamer behavior across the dimensions of social connectivity, playtime, game ownership, genre affinity, and monetary expenditure. As a whole, gamer behavior is highly diverse and characterized by heavy-tailed distributions. Most players exhibit modest behaviors in terms of the number of minutes played per day and the amount of money spent on games, though there is a long tail with outliers, such as gamers who maximize achievements or playtime stats, or gamers who collect games they don't play. We find some strong correlations that show that players tend to befriend those who are similar in terms of popularity, playtime, money spent, and games owned. We collect a second snapshot of the Steam network and show that our findings are robust across both measurements. We conclude by relating these findings to other relevant studies, including gamer stereotypes, game addiction, and social networking.

1. INTRODUCTION

It is estimated that nearly 60 percent of Americans play video games[1], bringing in annual revenue of over $25 billion for PC gaming alone[2]. There is a notable body of academic work focused on characterizing the gamers that make up this market, examining factors such as age, gender, and economic status [24, 11, 8]. That notwithstanding, the sheer size of this demographic has made comprehensive study of gamer behavior a difficult obstacle to tackle.

Research seeking to understand gamer behavior employs a variety of methods, including surveys, interviews, and large-scale data analysis. Surveys and interviews have the advantages of being able to discern motives and weigh demographic factors, such as gender or income [26, 25]. Data analysis complements this work by enabling researchers to accurately and directly characterize players' gaming behaviors based on statistics such as the amount of time spent playing. This is of particular value since data drawn from surveys can be skewed by participants who mischaracterize their behavior [10]. Moreover, the large datasets now available due to the online nature of game distribution and play allow us to study the behavior of millions of players at once, a scale that is simply not possible with surveys and interviews.

In this paper, we supplement earlier work performed on smaller scales [2, 3] by analyzing the gaming behavior of over 100 million users in the first comprehensive measurement of the Steam gaming network. Steam is one of the largest gaming networks in the world, and since our measurements has grown to 125 million active users at the time of publication.[3]

Steam provides a unique opportunity for analysis in several ways. First, because Steam is a platform for game distribution, our findings cover a broad spectrum of gamers. Previous measurements have often focused on a single game, such as World of Warcraft, Everquest, or Halo [24, 10, 6, 16]. Second, because Steam offers an open API for their platform, we are able to collect an exhaustive measurement of all users and games. Our measurements cover 108.7 million user accounts, 196.4 million friendships, 3.0 million groups, 384.3 million owned games, and a collective 1.11 million years of playtime. The comprehensive nature of this data enables us to accurately characterize the distributions of a variety of gamer behaviors across the entire Steam user base, bypassing potential sampling issues that may have skewed

[1] http://www.theesa.com/facts/pdfs/ESA_EF_2014.pdf/
[2] http://www.gamesindustry.biz/articles/2014-01-28-pc-gaming-market-to-exceed-USD25-billion-this-year-dfc/

[3] https://www.vg247.com/2015/02/24/steam-has-over-125-million-active-users-8-9m-concurrent-peak/

IMC 2016, November 14 - 16, 2016, Santa Monica, CA, USA

ISBN 978-1-4503-4526-2/16/11...$15.00
DOI: http://dx.doi.org/10.1145/2987443.2987489

previous results. Steam provides data on friendships, game ownership, and playtime data, but not access to more detailed game statistics. We thus necessarily focus on a general characterization of aggregate behaviors among Steam users, as opposed to an in-depth study of specific behaviors. However, even at this broad level we are able to better understand numerous aspects of gamer behavior.

So what are gamers really like? Using this dataset, we investigate questions concerning gamer behavior, such as how many friends they have, what kinds of groups they join, how much gamers play, what types of games they play, and how much money they spend. Our study indicates that as a whole gamer behavior is highly diverse and characterized by heavy-tailed distributions, meaning gamers are not easily characterized by summary statistics such as means. A variety of interesting behaviors manifest when examining this data in detail. In particular, our most salient findings are as follows:

- The number of friendships is low relative to other social networks, but most of the playtime is spent on multiplayer games.
- Most players exhibit modest behaviors in terms of the number of minutes played per day, and the amount of money spent on games. The 90th percentile for two-week playtime is 8.9 hours, the 90th percentile for games owned is 21, and the 90th percentile for lifetime market value is $317.64. Because these are 90th percentiles, the majority of users exhibit behaviors far below these values, which runs counter to stereotypes about gamers [12].
- There is a long tail of behaviors in all distributions, caused by a variety of different motivations. Some gamers focus on maximizing playtime stats or achievements, and others collect games they don't play.
- We find some strong correlations that show that players tend to befriend those who are similar in terms of popularity, playtime, money spent, and games owned.
- We demonstrate that playtime varies significantly over the course of a week for individual players; their playtime is not consistent from day to day.
- We examine achievements and find that there is a moderate correlation between how much a game is played and how many achievements it offers, and that players tend to get more achievements in adventure games.
- We collect a second snapshot of the Steam network and demonstrate that our findings are robust across both measurements. Despite large increases in game ownership (and money) in the heavy tail of the second study, the 80th percentile for these distributions doesn't increase nearly as much.

We conclude by relating these findings to other relevant studies, including gamer stereotypes, game addiction, and social networking. The large dataset we have collected frames the discussion of gamers as a diverse group and opens the door to study individual behaviors in more detail.

2. BACKGROUND

The Steam digital distribution service was started in 2003 and is owned and operated by the Valve Corporation, a video game developer. In this section we describe the Steam platform and related work that has used Steam to measure gamers.

2.1 The Steam Platform

Users of the service interact with it via a local client, available for Windows, Mac, and Linux. From the Steam client, gamers can download and launch games that have been associated with their accounts by purchase through Steam. After launch, the Steam client continues to run alongside the current game, providing access to instant messaging with friends, achievement notifications, game walkthroughs, and other features with an easily-accessible overlay.

The Steam catalog, as of May 2014, included titles from some 1,201 publishers, has attained an estimated 75% market share of PC video game sales[4], and as of that time had achieved a peak of 7.8 million concurrent users.[5] Valve reported over 75 million active Steam users[6], and we found 108.7 million total registered accounts at the time of our data collection.

In addition to its digital store, Steam provides numerous social networking features such as profile pages, friends, groups, instant messaging, voice chat, and news feeds. Steam also provides some social networking features specific to its gaming emphasis such as screenshot and video sharing, game achievement showcases, leaderboard statistics, and more. Each group on Steam has an event calendar, a group chat room, and a group message board to allow their members to better communicate and organize. A user's profile provides an overview of a user's social and gaming statistics. Additionally, the user's list of friends, joined groups, games owned, and recently played games are presented.

In addition, each user has a "Steam level" that is prominently displayed along with "badges" that denote various achievements on Steam. Badges increase a user's Steam level, which in turn unlocks additional widgets for users to enhance their profiles and increases the maximum number of friends they can have. Most badges are awarded by collecting complete sets of virtual trading cards, which can be obtained by playing corresponding games and trading with other users on the network.

[4]http://www.bloomberg.com/news/2013-11-04/valve-lines-up-console-partners-in-challenge-to-microsoft-sony.html/
[5]http://store.steampowered.com/stats/
[6]http://www.joystiq.com/2014/01/15/steam-has-75-million-active-users-valve-announces-at-dev-days/

Rank	Country	Percent
1	United States	20.21%
2	Russia	10.18%
3	Germany	7.56%
4	Britain	5.22%
5	France	5.19%
6	Brazil	3.95%
7	Canada	3.81%
8	Poland	3.20%
9	Australia	2.90%
10	Sweden	2.34%
	Other (226)	35.44%

Table 1: Breakdown of users' reported country. Shown for only the 10.7% of users that self-report a country of residence.

Smaller numbers of badges are awarded for participating in Steam community events and obtaining game collections of landmark sizes.

Steam users may opt to self-report a country of residence through their profiles. The distribution of countries for the 10.7% of users that have opted to do so is shown in Table 1. Prominent Western nations tend to have the most presence on Steam, but overall, users report to be from 236 distinct countries.

2.2 Measuring Steam

The Steam platform offers a unique opportunity to measure gamers at scale, due to its popularity and its open API. Though our work is the first to use the API directly, two prior studies have been conducted using partial data from the Steam network that is collected via a web crawler.

Becker et al. used a crawler to collect data on 9 million users (out of 30 million at the time), 82.2 million friendships, 1.98 million groups, and 1,824 games [2]. Becker primarily studied the structure of the gaming network, as defined by friendships among users, and found evidence of several small world graph characteristics. He also showed that the growth of users is exponential and the growth in degree distribution is linear over the period studied.

Blackburn et al. crawled the Steam Community web site to collect data on 12 million users (again out of 30 million), 88.5 million friendships, and 1.5 million groups [3]. Blackburn's focus was not on gamer behavior, but rather focused primarily on cheater behavior, finding that cheaters are not distinguished by their network characteristics.

Our data set and analysis are distinguished from these previous studies in both magnitude and focus. The exhaustive collection of data allows us to assess characteristics of gamer behavior at scale and with high confidence. For example, when previous studies collect a sample of Steam users with a crawl of the network, the data is biased since users with fewer friends are less likely to be crawled. Our results corroborate Becker's analysis of the evolution of the friend network, but their results were limited to a crawl of the large, connected component of the Steam network. As such, the scale of our results differ although the general behavior is similar. Moreover, our focus extends beyond a study of the social structure of the network to further analyze a wide variety of gamer behaviors.

3. METHODOLOGY

In this section we describe how we collect a full list of all the users, friendships, and games from the Steam system. We also explain how we validate and analyze the data.

3.1 Data Collection

Valve Corporation, the company that owns and operates Steam, provides a REST API, called the Steam Web API [7], for gathering information about users' profiles, friendships, game ownerships and playtimes, group memberships, and more. We use this API to crawl 108.7 million Steam accounts, along with the friendship list, games owned, play times, and group memberships corresponding to each account. Our dataset comprises all Steam accounts available at the time of collection. We also obtained information on 6,156 products available on Steam using the Steam desktop client.

While collecting data we acted within the terms and conditions of the public API and were careful to avoid negatively impacting the Steam service. To reduce strain on the Steam infrastructure, we limited our calls to the API to be roughly 85% of the maximum allowed by the API's terms of service. We did not collect any personally identifiable information and were limited only to data that Valve makes public regarding user accounts and games. We communicated with Valve regarding this work and provided them with drafts of our findings.

When a user registers with the Steam network, they are assigned a unique ID known as a Steam ID. Steam IDs have both 32-bit (e.g. STEAM_0:1:849986) and 64-bit representations (e.g. 76561197961965701), with a bijection between the two forms. These 64-bit Steam IDs are assigned in a sequential manner, beginning with a specified base value (76561197960265728). Dedicated game servers and game clients tend to use the 32-bit forms of Steam IDs, perhaps for legacy reasons. The Steam APIs and community web pages, however, use the 64-bit versions of Steam IDs, and we correspondingly deal exclusively with this representation in our study.

To obtain user account information, we exhaustively queried the Steam API for profile data for the entire population of Steam users, starting from the base ID and continuing until the API returned Steam accounts created just seconds before the moment of collection.We

[7]Steam Web API documentation, accessed 9 Nov 2014, http://steamcommunity.com/dev

crawled the ID space from February 28th, 2013 through March 18th, 2013, finding a total of 108.7 million user accounts as of the ending date of our crawl. This phase of data collection progressed relatively quickly because the corresponding API endpoint allowed us to collect up to 100 user profiles at once. Not all queried Steam IDs are associated with a valid Steam account, creating a variation of valid Steam account density in the queried ID space. Density was often below 50% in the beginning of the range until about 21.5% through, after which point density was consistently above 90%. We hypothesize that the non-uniform distribution of Steam accounts across the ID range is a consequence of changes in how IDs were assigned, deletion of old accounts, or both.

We consider all account IDs that are linked to account profiles to be valid Steam IDs. Using this list of valid Steam IDs, we made additional API requests for the friendships, game ownerships and playtimes, and group memberships corresponding to each user. This process required many more API calls than profile collection, because each query could only return data for a single Steam account at a time. Due to the vastly increased number of calls made to the API, in conjunction with API rate limits, this second data collection phase spanned several months, from May 5th, 2013 to November 5th, 2013.

In addition to user data, we also obtained information about the various products available on Steam. Such data is not available via the Steam Web API, but appears in the storefront of the Steam client itself. We analyzed the network traffic generated by the Steam client in its "Big Picture Mode" as it requested game data, and then emulated these requests for every product in the Steam catalog using each product's corresponding "App ID." The full list of App IDs used in these queries was obtained via an unpublicized endpoint of the Steam REST API. This data was collected on April 9th, 2014. Though the amount of data retrieved by these final queries was small, collection was performed over the span of several hours as we opted to rate-limit our queries to one every two seconds. The information collected during this final phase of collection included the associated genre(s), type (e.g. game, trailer, demo), developer name, current price, Metacritic rating, release date, and other such information corresponding to every product currently available on Steam.

Note that the data Valve provides do not include detailed play data for movement or actions in a game. For example, we can't tell whether two friends play the same game at the same time, or what kinds of interactions they have with a game. Rather, the data provided affords a high level view of overall user behavior regarding friendships, games owned, time and money spent playing games, achievements reached, and the correlations among these variables.

Our full datasest, including all source data mentioned in this work, is available at `steam.internet.byu.edu`.

3.2 Data Validation

All of the data collected concerning user accounts is publicly accessible from player profiles, through both the Steam website and client, which provides some avenues for further investigation and validation. We spent nearly a week extensively sampling our data manually to assure that accounts were associated with real users, both by randomly sampling hundreds of accounts and also by examining all accounts that exhibited extreme behaviors. To ensure that our data represented real players, we examined their name, friends, and posts on their public profile. A small number of accounts are marked by Valve to indicate whether they belong to Valve employees or developers, but even these are used as personal accounts to play games and marked only for cosmetic reasons. In cases where we observe anomalous behaviors (such as owning large numbers of games and not playing any of them), we manually checked all such accounts and confirmed that these appear to be real people who have different motivations from a typical player (e.g. collecting instead of playing), rather than test accounts.

3.3 Data Analysis

One contribution of our work demonstrates that the Steam social network and various gaming behaviors are characterized by heavy-tailed distributions. Some previous work measuring social networks have claimed networks follow power-law distributions, without properly distinguishing between types of heavy-tailed distributions [4, 18]. Properly identifying the precise nature of these distributions plays an important role in modeling because this provides important clues about the underlying generative mechanisms involved. Here we briefly describe the distributions mentioned in the paper as well as the technique employed in classifying our data as one of these distributions.

Heavy-tailed distributions are a class of probability distributions whose tails are not exponentially bounded. These distributions are of particular interest because they describe behaviors wherein the probability of observing an extreme event is more likely than would be suggested by a Gaussian (normal) distribution. As might be expected, however, not all distributions classified as "heavy-tailed" describe identical behaviors. Instead, they are further broken down by the specific likelihoods of these extreme events into broader classes such as long- and fat-tailed distributions as well as specific distributions such as power-laws and lognormal distributions.

As all the distributions described in this work are heavy-tailed distributions, we utilized the *powerlaw 1.3* package [1] - a Python tool designed to classify empirical, heavy-tailed data - to identify with greater specificity the precise nature of these distributions. This tool synthesizes the various types of heavy-tailed distributions and performs a goodness-of-fit test, the Kolmogorov-Smirnov statistic, against the empirical data, providing

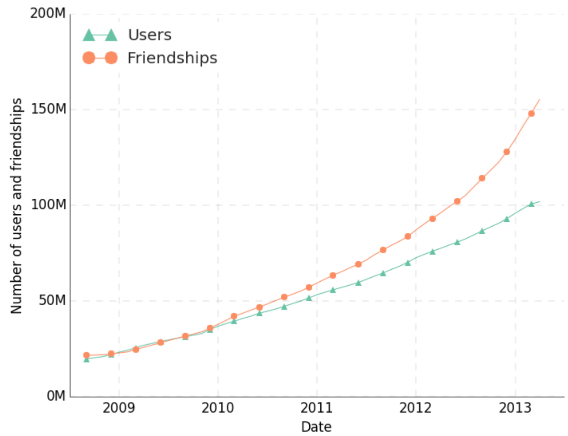

Figure 1: Evolution of Steam friendship graph. The number of users and the number of friendships shows a steady increase since 2008.

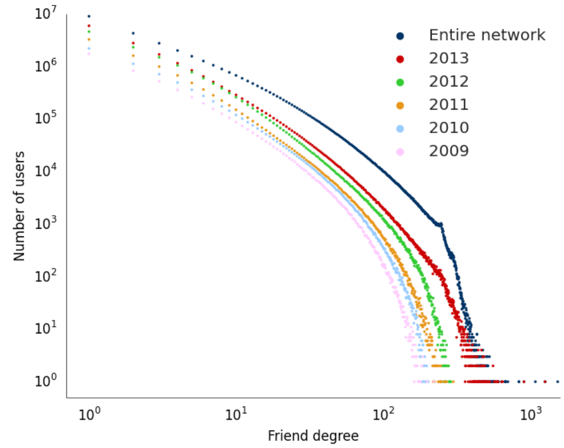

Figure 2: Evolution of Steam friend degree distribution. There is a long-tailed distribution of friendships per use, both each year and for the entire network.

a likelihood ratio that the data fits the given distribution type, a methodology described by Clauset et al. in [4].

Within this paper, we label the distributions captured in our data as one of the following types: heavy-tailed, long-tailed, lognormal, and truncated power law distributions (we do not observe any true power law distributions). Throughout the body of this work, we label the discussed distributions in accordance with the greatest specificity we are able to assign with confidence. In particular, we label a distribution as heavy-tailed when it passes the power law vs. exponential distribution test, but cannot with any degree of certainty clarify the label further than that. The label of long-tailed distribution is assigned to distributions that we were able to narrow to being either a lognormal or a truncated power law distribution, but for which the statistical comparison is unable to classify between the two. The detailed parameters for the fitting process and statistical tests, as well as the classifications for each distribution, are found in the Appendix.

4. SOCIAL STRUCTURE

Steam provides its users with the ability to establish social links with other users through friend relationships and group affiliations. In total, we find 196.37 million bidirectional friendships and 81.30 million group memberships on Steam.

4.1 Friendships

We study the graph of Steam friendships, made by considering each user to be a node and each friendship to be a link. Friendships in Steam are reciprocal, and thus modeled here as bidirectional links. Figure 1 shows the evolution of the Steam friendship graph since friendship timestamps were first recorded in September of 2008 until the time of our data collection. Note, the graph does not include friendships established from 2003 to 2008, so it does not reach the total 196.4 million friendships we collected. This graph reveals a steady increase of Steam users since 2008 as well as that friendships form at a faster rate than users join the network (demonstrated by the steeper, increasing slope of the latter curve).

To understand the connectivity of gamers in the Steam friend graph, we plot the distribution of the number of friends users have in Figure 2. Both yearly breakdowns and a distribution representing the full network to-date are provided. All years exhibit a long-tailed distribution of friendships per user: on average, 9 million gamers (88.06%) add ten or less friends per year and only 2,500 gamers (0.02%) add more than two hundred friends per year. The shape of each year's curve is very similar despite the steadily increasing size of Steam's user base over the years, indicating that overall gamer friendship behavior is independent of the size of the network.

We also observe two deviations from the otherwise smooth nature of the distribution for the entire network. The first occurs at a position corresponding to 250 friends and the second at a position corresponding to 300 friends. The number of users with friend counts above 250 are much less than expected given the prior part of the curve. Both of these are due to the default limitations placed on friend counts. Without special action, users may not have more than 250 friends. This limit is raised to 300 if a user links her Steam account with her Facebook account. Steam also offers five additional "friend slots" as a reward for each "level" obtained by a user in a meta-game involving virtual trading cards acquired by playing games and trading with other users. These hurdles account for the lower-than-expected number of people with friend counts above 250.

Since these distributions are long-tailed, rather than normal, the mean does not describe a "typical" user

Group Type	Count	Percent
Game Server	114	45.6%
Single Game	51	20.4%
Gaming Community	43	17.2%
Special Interest	35	14.0%
Steam	4	1.6%
Publisher	3	1.2%
Total	250	100.0%

Table 2: Breakdown of 250 largest groups by type. Game server groups comprise the largest portion.

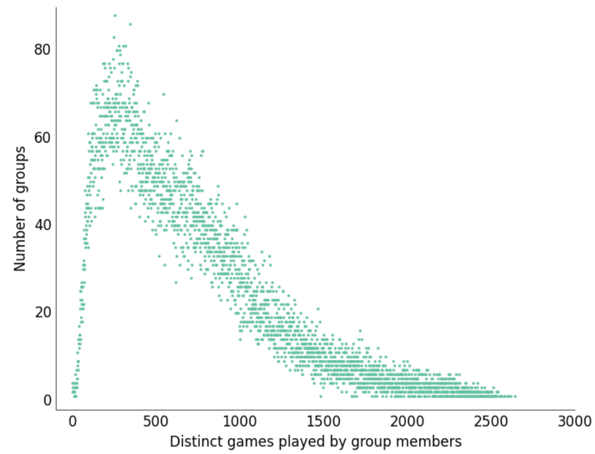

Figure 3: Distribution of the number of distinct games played by members of a group. Shows only groups with more than 100 members. Many groups play hundreds of games.

or behavior. For example, in the entire network, the average number of friends a user has is four, but only 1.85% of Steam users have four friends. However, with the full distribution visible, we can clearly see that the overwhelming majority of gamers have a few friends, and an increasingly smaller amount have larger numbers of friends. Such distributions are common throughout our results. We will repeatedly see that more modest measures of behavior describe the overwhelming majority of gamers while vast ranges of more extreme behavior can be attributed to a small minority.

As previously described, Steam profiles also allow their owners to report their location, with 10.7% and 4.0% of users reporting their country and city of residence, respectively. Using these values we explore the extent to which physical locality exists in users' friend lists. We find that 30.34% of friendships between users who both declare a country location are international. We find also that, for users who report a city, friends in 79.84% of friendships come from different cities. We thus see that gamers form friendships with more people outside their neighborhoods than inside.

4.2 Groups

Plotting groups by their membership counts reveals that group size distribution is heavy-tailed while the number of groups users join follows a long-tailed pattern. Likewise, plotting users by the number of groups they belong to reveals another heavy-tailed distribution, with a substantial amount of users belonging to just one or no group and a long tail of some users belonging to a large number of groups.

Unfortunately, Steam does not currently offer a standard way for a group to identify its nature to viewers. However, we manually investigated the pages of the top 250 largest groups and categorized each into one of six categories: "Single Game", "Game Server", "Gaming Community", "Publisher", "Special Interest", and "Steam". Single Game groups are those created for fans of an individual game. Game Server groups are those that host dedicated servers for one or more distinct games. Gaming Community groups focus on community identity and play multiple games. If a Gaming Community group also hosts dedicated servers it is labeled as a Game Server group. Publisher groups are formed for fans of a particular game publisher. Special Interest groups are those created for fans of topics unrelated to specific games (e.g. trading, activism, beer). Finally, Steam groups are official groups created by Valve for updates concerning Steam features.

The breakdown of the types of the top 250 most populous Steam groups is provided in Table 2. Game Server groups account for the largest portion of the top 250 groups, followed by Single Game and Gaming Community groups.

The popularity of groups centered on games prompts the question of what number of distinct games are played by the members of a single group. The answer to this question is depicted in Figure 3, which shows the number of groups distributed by the number of distinct games played by their members. We consider only large groups with 100 members or more, constraining the number of groups to 58,986. Small numbers of distinct games played by group members indicate a very focused group. Large numbers of distinct played games indicate groups with less emphasis on specific games.

We find that a relatively small number of groups have memberships that are dedicated to very small sets of games. Further investigation reveals that only 2,933 of these groups (4.97%) have members who dedicate 90%-100% of their playtime to a single game. We find that large amounts of groups have members with playtime for larger numbers of games (game counts of 100 - 1000). However, as the number of distinct games continues to increase the number of corresponding groups decreases. This decline occurs because most users who own games (89.78%) own less than 20 games, making large numbers of distinct games across the group a rarity.

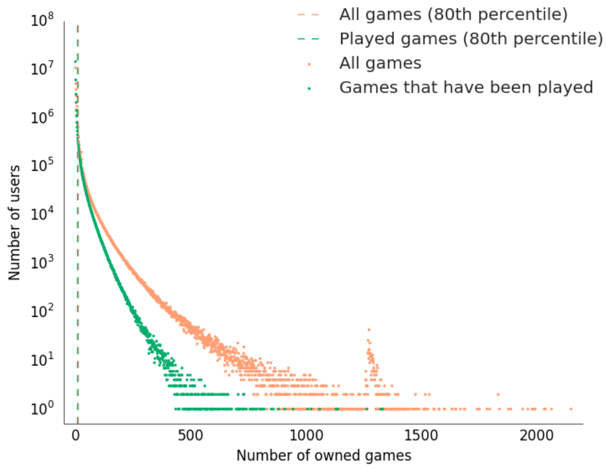

Figure 4: Distribution of game ownership. The 80th percentile is 10 games owned and 7 games played.

5. GAME OWNERSHIP

Figure 4 shows the long-tailed distribution of game ownership on Steam, including all games in a user's library as well as only those that have been played at some point. To put this data into perspective, we show also the 80th percentile values for these two types alongside the distributions as vertical lines. This means that 80% of Steam users who own at least one game fall to the left of the vertical lines on the graph. The relative placement of these values on the graph, of 10 and 7 games respectively, clearly demonstrates the sheer length and span of the heavy tail.

Two additional points are of note in Figure 4: the uptick of owners who own between 1268 and 1290 games (note also a similar uptick between $14,710 and $15,250 in Figure 8), as well as the distance between the two depicted distributions. We believe that this phenomenon can be attributed to both the tendency of players at large to acquire games they do not play as well as a specific "collector" subset of the Steam population for whom the acquisition of games is a goal in and of itself.

The existence of unplayed games within the library of a typical user can be explained in two ways. The first is the prevalence of game bundles sold on Steam (publisher collections, series packs, Humble Bundle [9] offers, etc.). One might purchase a bundle of games desiring only a single game in the bundle, and thus never play the rest. Additionally, there are limitations on time and energy; one only has so much free time, and the larger one's game collection, the more difficult it becomes to try them all, regardless of intent. For example, one author has only played 58% of the games associated with his account due to these reasons.

The other contributing factor to the effects illustrated in Figure 4 is the behavior of a specific set of Steam users we label as collectors, and is disparate from the propensity of normal Steam users to acquire unplayed games as previously described. These are users who acquire many games but never play most of them. As far as we have been able to tell through our investigations of this issue, this is not done as a financial investment because Steam does not allow users to trade or resell games that have already been redeemed into a user's game library. Furthermore, as a platform for digital distribution, there are no limits on the number of copies of a game sold with the exception of games that have been discontinued (which users are unable to forecast). Instead, users seemingly do this for social prestige or vanity, e.g., there are invite-only Steam groups whose membership is predicated upon a user owning at least a certain number of games.

As an example, our queries found 29 users with game libraries of at least 500 games who had never played a single one. The user with the largest game collection in our dataset owned 90.3% of the games currently available in the Steam store at the time we retrieved his data. This same user, however, had only accumulated playtime on 34.5% of those games in his collection[8]. This desire to "collect" is further reinforced by profile badge awards that directly correspond to game collection. Circumstantial evidence confirms that some users do prefer to collect – on forums users brag about their game ownership, others trade virtual commodities for game copies, and there are online leaderboards boasting account value. One user is well known for collecting spare copies of Barbie Dreamhouse Party, owning hundreds of copies of it.

We also investigate the genre dimension of game ownership. Steam assigns one or more genre labels to each game it offers. Most genre labels describe the principal gameplay mechanics innate to a game such as "Action" and "Strategy." There are, however, some exceptions: "Free to Play" indicates that a game is free, but is likely to have in-game "microtransaction" offers for additional features and "Indie" indicates that the title has been developed or published by a small, independent firm.

Figure 5 shows the ownership of games on Steam broken down by genre. The lighter bars represent the total number of games of that genre owned across all Steam accounts. The darker bars (the bottom segments) show how many of those games that are owned have not been played. The "Action" genre is by far the most popular in terms of ownership, followed distantly by the "Strategy" and "Indie" genres. Action games comprise 38.1% of Steam's current catalog. Secondly, we also note a large number of unplayed games in user's libraries. 41.49% of Action games, for example, are unplayed. We also see large percentages of unplayed games for all other genres (28.86% for Strategy, 32.3% for Indie, 24.26% for RPG, etc.). This further illustrates the gap seen in Figure 4 between played and unplayed games and shows that this behavior seems to be genre-agnostic.

[8]As discussed in Section 3.2, we did manually confirm that this was a real player.

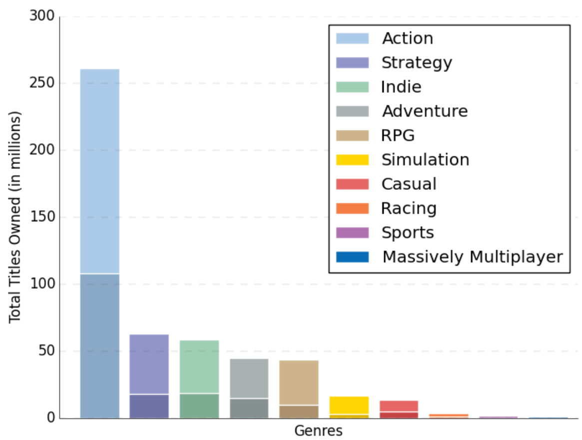

Figure 5: Game ownership by genre. Light bars represent total copies owned, and dark bars represent owned but unplayed games. The top-down ordering of the legend corresponds to a left-right order of genres. The Action genre is by far the most popular in terms of ownership. There is a significant gap between owned but unplayed games and played games.

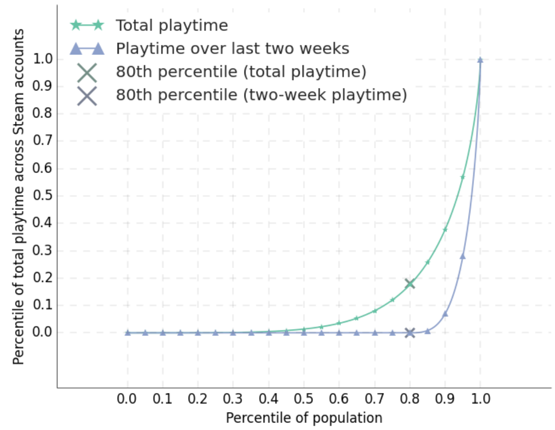

Figure 6: CDF of total and two-week playtime. The top 20% of Steam users account for 82.4% of the total playtime of all games played on Steam.

6. TIME AND MONEY

Given the numerous studies on game addiction [13], the resource expenditure devoted to this activity is of particular interest. The Steam API provides an opportunity to answer the questions of how much time and money is spent gaming.

For each game in a Steam user's library, playtime is directly recorded in two ways: the total playtime for that game since it was added to the user's library, and a rolling playtime value spanning the two weeks leading up to the moment the query was issued. These values are recorded at the granularity of a minute. Games generally cannot be played without recording this data – games must be run via the Steam client and even if disconnected from the Internet the data is reported after reconnection.

The precise amount of money spent by users on the games in their library is not directly revealed through the API. Despite this limitation, we provide an approximation of money spent by calculating the market value of each user's account using the 2014 price of Steam games as shown on the Steam storefront. We assume an averaging effect in the current prices of games in the Steam store. While many games are purchased on sale, there are also permanent price reductions for games that are not listed as discounts; both are reflected in the current market value approximation that we use. Thus, while there are occurrences where our estimates are likely too high, because a game in a player's library was purchased on sale, there will similarly be occurrences where our estimate is too low, because a game was purchased prior to a price reduction. Given that there is no way to get accurate pricing information without receipt information that Steam keeps private, this approximation method is, to our knowledge, no worse than any other.

Using these methods across the entire population of Steam users, we find a cumulative 1.11 million years of playtime as well as a market value of $5,326,471,034.78 of Steam games.

6.1 Resource Expenditure Across the Population

Figure 6 portrays the distribution of the cumulative values of total and two-week playtime across the entirety of Steam. The proximity of the two curves to the x-axis again demonstrates the unevenness of behavior between the vast majority of Steam users and their more extreme counterparts. For reference, the top 20% of Steam users account for 82.4% of the total playtime of all games played on Steam. This number parallels the well-known "80-20 rule" or Pareto principle, used frequently to describe heavy-tailed behavior. Similar behavior is seen in YouTube, where one measurement indicates the top 20% of users contribute 72.5% of the videos [5]. The two-week playtime data reveals that the top 10% of gamers contribute 93.0% of two-week playtime totals.

As stated earlier, the two-week playtime statistics are reported in rolling, two-week windows leading up to the time of the query. It is worth noting that as the phase of our data collection process which included these values spanned a period of several months, it is fairly likely that the reported numbers are representative of "normal" playing habits, i.e. these values actually represent a variety of distinct two-week periods falling within that time span.

As can be seen in Figure 6, over 80% of Steam gamers *had not played any of their games at all* in the two weeks

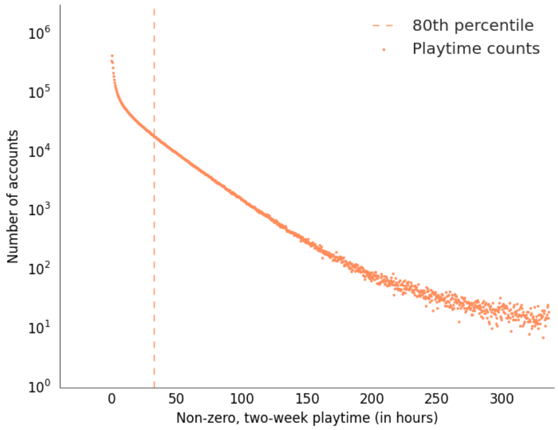

Figure 7: Distribution of non-zero, two-week playtimes. The 80th percentile is 32.05 hours, or about 16 hours a week on average.

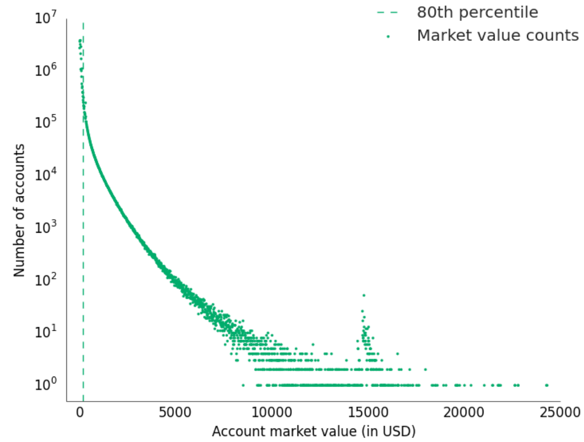

Figure 8: Distribution of account market values across Steam. The 80th percentile is a value of $150.88.

prior to our requesting their gaming data. Thus, not only are playing habits observed during a small window of time not representative across the population as a whole, for more than 80% of Steam gamers, they may not even be representative of the amount of time that particular user plays generally.

Figure 7 shows the distribution of two-week playtime when these Steam users are removed from the graph, i.e. when only non-zero two-week playtimes are shown. This illustrates the playtime trends of the 20% of Steam users who *did* play during their two-week period. To give a sense of the scale of the length of the tail shown here, the 80th percentile value - of 32.05 hours, or just over 16 hours a week on average - is also plotted. Note also that this 80^{th} percentile value of non-zero two-week playtimes corresponds to the 95^{th} percentile of overall two-week playtimes.

The very end of the tail reaches the maximum possible value of 336 hours, or the total number of hours in two weeks. We do not mean to suggest that there are players who spend literally every second of every day playing games. The Steam client tracks playtime for as long as a game is left open, and not necessarily for the amount of time a user may be active in said game. As such, users who leave their game open will contribute to these recorded values. Some players may be purposely racking up high playtime statistics as a bragging right, since the most recent two-week value appears in a player's profile. Our experience with Steam leads us to believe this practice is not common and does not affect the bulk of our playtime results, since this can consume substantial system resources and does not lead to significant benefits. However, there are indeed some users who have extreme playtimes on the order of 80-90% of the maximum two-week playtime. These values constitute a very minor portion of our data (0.01% of users), so we note these caveats and leave further exploration of this behavior to future work.

The monetary analogue of our playtime graphs can be found in Figure 8, which displays the distribution of the current account market value of Steam users' game libraries. The anomaly in the graph is likely due to game collectors, as discussed with Figure 4. Again demonstrating the extreme nature of values in the tail, the 80th percentile value, which appears in the graph to be nearly zero, actually carries a value of $150.88. Consider, furthermore, that the single largest current account value ($24,315.40) is over 160 times larger than this 80th percentile number. Additionally, though not displayed in this figure, we note that "80-20 rule" behavior is again present: the top 20% of Steam users account for 73% of the total current market value of owned games on Steam.

In short, the vast majority of Steam gamers behave drastically differently from a select few who both play and spend orders of magnitude more than their peers.

6.2 Resource Expenditure by Game Type

Figure 9 details the breakdown per-genre of both playtime and account market values on Steam (foreground bars represent playtime and background bars represent account market value). As games can fall within multiple genres, there exists a certain degree of overlap between the values displayed, though values can still be directly compared against the total number of games owned by Steam users. Perhaps the most interesting data point with respect to resource expenditure per-genre concerns the action genre. The action genre constitutes 49.24% of total playtime on Steam as well as 51.88% of the total current market value of Steam users' game libraries. Although it must be said that Steam users do own more action games than other genres on Steam, as depicted in Figure 5, it must also be noted that these numbers are still over-representative, as only 38.3% of offered Steam

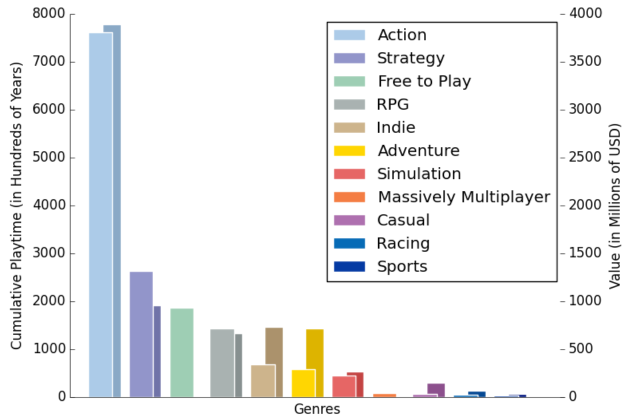

Figure 9: Cumulative playtime by genre. Foreground bars correspond to left axis, background bars correspond to right axis. The top-down ordering of legend corresponds to a left-right order of genres. The Action genre is the most popular.

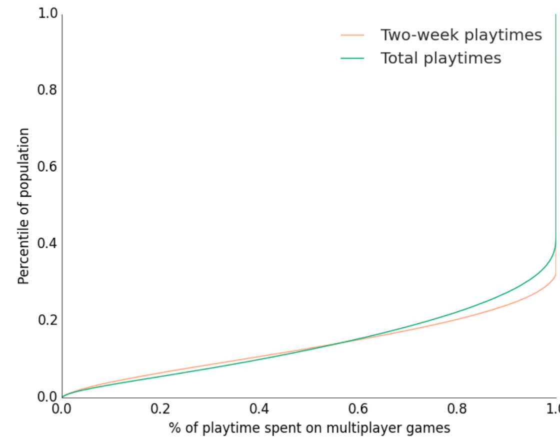

Figure 10: Distribution of multiplayer game playtime. 67.7% of two-week playtimes as well as 57.7% of total playtimes are devoted entirely to playing multiplayer games.

games fall within this genre (at the time these data were collected).

Figure 10 depicts the ratio of playtime on Steam devoted to single or multiplayer games. 67.7% of two-week playtimes as well as 57.7% of total playtimes are devoted entirely to playing multiplayer games, despite the fact that only 48.7% of games on Steam are multiplayer games (i.e. have a multiplayer component). We hypothesize that this over-representation is due to the tendency of single player-only games to present players with a very limited set of original gameplay options. This places limits on their replayability, and thus an upper limit on the amount of time a player may spend on them. In contrast, the natural dynamics of multiplayer features often allow sustained, original experiences with other persons and incentivize more time investment via competitive elements.

7. CORRELATIONS AND HOMOPHILY

One interesting area to explore is whether there are correlations between different aspects of gamer behavior. Do players who own more games tend to play more often or have more friends? Do players with more friends play more often?

We examine plots of correlations to see general trends and use Spearman's rank correlation to determine the significance of the correlation. Spearman's ρ varies from -1 (inversely correlated) to 1 (positively correlated). One interpretation of the absolute value of ρ is that from 0 to 0.19 the correlation is "very weak", 0.20 to 0.39 is "weak", 0.40 to 0.59 is "moderate", 0.60 to 0.79 is "strong" and 0.80 to 1.0 is "very strong".

We find only a weak correlation between the number of games owned and the number of friends a player has (ρ=0.34), between number of games owned and two-week playtime (ρ=0.28), and between number of games owned and the total playtime (ρ=0.21). We find no correlation between both the number of friends and two-week playtime (ρ=0.09) and between number of friends and total playtime (ρ=0.17).

We find stronger correlations when we consider homophily – the tendency of players to be friends with those that are similar to them. Homophily has previously been demonstrated with Twitter [23, 14], where friends tend to share topical interests, geographic location, and popularity.

We find a strong correlation between a player's account market value and their friends' average account market value (ρ=0.77), between a player's number of friends and the average number of their friends' friends (ρ=0.62), and between a player's total playtime and their friends' total playtime (ρ=0.61). We find a moderate correlation between the number of games a player owns and the average number of games the player's friends own (ρ=0.45). An example plot of these correlations is shown in Figure 11, which shows the correlation for market value. These correlations indicate that players tend to befriend those who are similar in terms of popularity, playtime, money spent, and games owned.

8. EVOLUTION

In this section we discuss the evolution of user behavior over time. While we have performed a comprehensive analysis of gamer behavior on Steam, our data provides but a single aggregate measure of each user's behavior in total, from the moment their account was created until the time we collected their data. This raises two questions: are our results truly representative of the behavior of Steam users or are they an artifact of the

Figure 11: Correlation of market value with friends' market value. (ρ=0.77)

time when we collected our data? Does user behavior change over time?

To answer these questions, we collected two additional data sets. Recall that we collected game ownership and playtime data for 108.7 million Steam accounts between May 5th, 2013 and November 5th, 2013. We created another snapshot of game data for the same users roughly a year later, between August 14th, 2014 and October 3rd, 2014. Additionally, to examine user behavior on a finer-grained timescale, we collected game data for a sample of 0.5% of our user set over a week's time, from Saturday, November 1st through Friday, November 7th, 2014. (Collecting this data every day for all users is not possible due to voluntary rate limiting on the API, as requested by Valve.) To sample users, we ordered them based on their total minutes played over the lifetime of their account. We then selected a uniform, random sample of 0.5% of all users across this space.

In examining our second, full snapshot, we find that, on the whole, while the magnitude of values in the tail change drastically, the distribution classifications remain unchanged from those presented earlier (see Table 4 in the Appendix), bolstering confidence in our analyses. Particularly fascinating, however, is the finding that despite these huge increases in the magnitude of resource expenditures by Steam users in the heavy tail, the magnitude of 80th percentile values did not change to nearly the same degree. For reference, the highest account market value in our first snapshot, $24,315.40, increased to $46,633.69 in our second, although the 80th percentile account market value only increased from $150.88 to $224.93. The game ownership numbers depict this behavior in an even more dramatic fashion: the top ownership number in the first snapshot of 2,148 games increased to 3,919 whereas the 80th percentile ownership number changed only from 10 to 15.

Figure 12 shows the playtime per-day over a one week period for our sample of users. In constructing this set, we first removed all users who did not play at all during this seven-day period. We then ordered users by the number of hours they played on the first day, from 0 to 24. Colors in the graph range from white, for 0 hours, to black, for 24 hours. On all subsequent days, users occupy the same position along the x-axis they occupied on the first day, with their corresponding playtime on that specific day colored in the same manner, illustrating any changes in their playing behavior over these seven days.

Viewing the graph over all seven days indicates that many users who did not play on the first day of our collection played significant amounts of time on subsequent days. In other words, game playtime is not a characteristic unique to a singular group of "heavy hitters". Instead, a variety of players have heavy usage on some days and lower usage on other days. Changes in individual behavior notwithstanding, there is still a general trend wherein the players who played the most on the first day generally play more than their counterparts on subsequent days (as evidenced by the overall lighter shading of the left half of the graph).

9. ACHIEVEMENTS

The Steam API was recently updated with an endpoint to obtain information about in-game achievements. Achievements are rewards given to players for completing various tasks while in-game, which are defined by the game developers. Achievement completion is permanent and users have the option of displaying a selection of their achievements on their profiles. Achievement lists and statistics are also available publicly as a means for users to compare their game performance to one another or to the community at large. On May 6th, 2016, we collected achievement data for each game on Steam. For each game, this included the name of each achievement and percentage of players who earned that achievement. For each game, only players who own that game are considered in the game's achievement completion records.

We analyzed the achievement records to find possible correlations between achievements and game-playing behaviors. First, we assessed whether the number of achievements a game offers has any correlation with amount of time players spend playing that game. The number of achievements offered by games ranges from 0 to 1629, with a mode of 12, an average of 33.1, and a median of 24. When evaluating games' cumulative playtime with respect to the number of achievements they offer, we found a very weak positive correlation (Spearman's correlation, $R = 0.16$). However, when we constrain the range of offered achievements to only games that offer 1-90 achievements, we find a moderate correlation ($R = 0.53$). Beyond the range of 90 achievements offered, there is no correlation with playtime ($R = -0.02$).

This suggests a possibility that achievements incentivize more playtime than would have otherwise occurred.

Figure 12: User playtime tracked over a week. We followed a sample of 0.5% of the Steam population over the period of a week. These users are represented by lines on the figure, with darker lines representing greater playtime. On day one, these users were sorted by their playtime on that day from lowest to highest. The lines representing these users occupy the same position on the x-axis across all subsequent days, allowing their behavior to be visually followed through the progression of this week-long period.

However, we note that this is merely a moderate correlation as currently studied and the relationship may be due to other factors. For example, it could be that higher-quality games that would be played more anyway simply have developers who choose to take advantage of Steam's achievement features. If the relationship is causal, though, it is interesting to note that it does not endure beyond 90 achievements offered. We hypothesize that this bound could be due to user perception of the quality of those achievements and possible fatigue when faced with collecting so many. One Reddit user remarked on his pursuit of achievements by saying, *"when I first play through a game, I play it the way I want to play it. When I'm done, should I choose, I like to go for achievements just to elongate the game some more. Sure, there will be stupid achievements that will be way too hard or uninteresting for me to complete, so I don't do [those]"*.

Next, we evaluated achievement collection behavior with respect to game attributes. Intuition suggests that there may be a difference in achievement gathering behavior between single player and multiplayer games, as the latter can provide extended playtime behavior through the originality of interaction with others. However, we found that the mode of the average completion rate for both types of games was 5% (the mode number of achievements offered by both types of games is 12). In addition, the medians of the average completion rates were 11% for single player games and 12% for multiplayer games (where the median of achievements offered were 24 and 21, respectively). Inspecting the corresponding averages of these results yielded similar results, but both values were also much higher (14% average completion for multiplayer, 15% for single player), suggesting that there is a minority group of players who aggressively seek achievements and skew the average above both the median and the mode. Players in this group are known to the community as "achievement hunters". We note, however, that such data is also consistent with alternative hypotheses. For example, it could be that a portion of games offer achievements that are easily obtained relative to those of other games, offsetting the average from the median in a similar fashion.

Further assessment of the existence and nature of the achievement hunter group requires access to individual players' achievement statistics instead of aggregations collected. We also found that achievement completion varies based on the genre of the game being played. For example, Adventure games offer a fairly average amount of achievements but have the highest average completion rate (19%) and Strategy games have a very low average amount of achievements but their completion rate is also low (11%). These could be due to differences in the type and difficulty of achievements offered in differing genres, or differences in the goals of their respective players. We leave further analysis of achievement nature to future work.

10. DISCUSSION

To get an overall view of our data, we calculated several useful percentiles for the major distributions we reported earlier, as shown in Table 3. Most gamers spend modest amounts of time and money playing games, and half did not play at all during the two week period we measured and are not otherwise highly engaged in terms of friends, games owned, or money spent. The vast majority of gamers appear to be casual players. Significant future work will require identifying and interviewing individual players across the spectrum of behaviors to better understand gamers and their motivations. We reflect on how these numbers impact a number of related research areas.

10.1 Gamer Stereotypes

Stereotypes of gamers are prevalent in society, characterizing them as "isolated, pale-skinned teenage boys sit hunched forward on a sofa in some dark basement space, obsessively mashing buttons" [12]. Early work debunking this stereotype primarily relies on interviews and surveys, while more recent papers have collected data directly from game logs or via APIs. The seminal work in this area is a study by Williams, Yee, and Caplan

Attribute	50th percentile	80th percentile	90th percentile	95th percentile	99th percentile
Friends	4	15	29	50	122
Owned games	4	10	21	39	115
Group membership	2	7	13	22	62
Account market value	$49.97	$150.88	$317.64	$587.63	$1,593.78
Total playtime	34 hrs	336.4 hrs	739.8 hrs	1233.9 hrs	2660.1 hrs
Two-week playtime	0 hrs	0 hrs	8.7 hrs	25.5 hrs	70.8 hrs

Table 3: Percentiles computed on various game player attributes.

that surveys 7,000 players of *EverQuest II*, a massively multiplayer online (MMO) game, and compares this self-reported data to measurements collected by the game operator [24]. Their study debunks a number of stereotypes regarding gamers, finding that gamers are older than expected, and that women play more than men.

Our work adds additional evidence to this area, particularly related to whether gamers as a whole spend most of their time playing. Over the 108.7 million users we measure, the 90th percentile for two week playtime is 8.7 hours, or a little over a half hour per day. Even the 95th percentile is 25.5 hours over a two week period, or under 2 hours per day. The majority of users exhibit behaviors far below these values. This demonstrates the potential to ground studies of gamers in the massive quantities of data that are now available.

10.2 Game Addiction

Internet game addiction is a hotly debated topic. There has been a wide variability in reported rates of addiction, ranging anywhere from 0.2% in Germany [7] to 46% [22] in Taiwan, however these variations appear to be due to improper research methods – study samples tended to be non-representative, limited in scope, and relied on self-reported data [21, 7, 19]. Reported high rates of addiction result from studying subgroups of gamers in the online role-playing community, whereas studies using representative samples typically report addiction rates less than 5% [7]. A review of the literature by Kuss and Griffiths [13] finds that addiction prevalence among students ranges from 2.2% to 12%, in various different age groups, and that prevalence among all gamers ranges from 8% to 12%. However, the literature uses different assessments of what constitutes addiction as well as different participant groups.

While it is beyond the scope of this work to address the issue of online game addiction, our data provides a much broader picture of user playing time and money spent, covering many more games and users than previously studied. In addition, the data set we have compiled and analyzed is comprehensive, as opposed to sampled; includes measurements of playtime and game ownership that have been collected by the service itself, rather than self-reported; and covers a wide range of users spanning the entirety of the Steam user base as opposed to select groups such as those that are fans of a particular game or genre. Our data could help the community define a cutoff for what is considered addicting; for example, the top 1% play more than 5 hours a day, have hundreds of games, or have spent thousands of dollars. Because we collected so much data, this 1% represents over a million gamers, so this demographic should be studied in more depth. Future work with this data could follow up with individual players to study how their expenditures of time and money may qualify as addiction.

10.3 Social Networking Among Gamers

While Steam is only partly a social network, we can relate our measurements of Steam's friendships to the extensive measurements that have been performed of other social networks. The primary result we see is that Steam is best characterized as a network of friends, rather than a network where celebrity accounts may gather very large numbers of followers. All Steam friendships are reciprocal, most users have very few friendships, and the maximum number of friends is capped by various Steam policies. In addition, the network exhibits a strong homophily with respect to friends – as users have more friends, they tend to connect to those with more friends. This is similar to the correlations seen in a measurement of Flickr [17], indicating the Steam friendship network functions like other networks without celebrities. Finally, Steam's network follows a lognormal distribution, similar to studies of Google+ [15] and Facebook [20]. In a network with a lognormal distribution of friends, this means that probabilistically there will be more users with a small number of friends than in a power-law network. This is in contrast with measurements of Twitter [14] and YouTube [17], which indicate that these networks follow a power law distribution (though in some cases the analysis is only graphical).

11. CONCLUSION

The vast scale of Steam, both by number of games and by number of users, reveals the tremendous diversity of behavior among gamers. This diversity of behavior manifests itself as a continuous spectrum wherein every possible behavior is embodied. Most players exhibit modest behaviors associated with casual gaming, with outliers who have more extreme or unusual behavior. We find a variety of strong correlations that indicate social networking shapes how users play. A data set this comprehensive offers numerous possibilities for followup research to gather qualitative data from individual users and correlate this to their quantified playing habits.

12. REFERENCES

[1] J. Alstott, E. Bullmore, and D. Plenz. powerlaw: a Python package for analysis of heavy-tailed distributions. *PLOS ONE*, 9(1), 2014.

[2] R. Becker, Y. Chernihov, Y. Shavitt, and N. Zilberman. An analysis of the Steam community network evolution. In *IEEE Society of Electrical & Electronics Engineers in Israel (SEEEI)*, 2012.

[3] J. Blackburn, R. Simha, N. Kourtellis, X. Zuo, C. Long, M. Ripeanu, J. Skvoretz, and A. Iamnitchi. Cheaters in the Steam community gaming social network. *arXiv preprint arXiv:1112.4915*, 2011.

[4] A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *Society for Industrial and Applied Mathematics (SIAM) Review*, 51(4):661–703, 2009.

[5] Y. Ding, Y. Du, Y. Hu, Z. Liu, L. Wang, K. Ross, and A. Ghose. Broadcast yourself: Understanding YouTube uploaders. In *ACM Internet Measurement Conference (IMC)*, 2011.

[6] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore. Alone together?: Exploring the social dynamics of massively multiplayer online games. In *ACM Coference on Human Factors in Computing Systems (SIGCHI)*, 2006.

[7] R. Festl, M. Scharkow, and T. Quandt. Problematic computer game use among adolescents, younger and older adults. *Addiction*, 108(3):592–599, 2013.

[8] M. D. Griffiths, M. N. Davies, and D. Chappell. Breaking the stereotype: The case of online gaming. *CyberPsychology, Behavior, and Social Networking*, 6(1):81–91, 2003.

[9] Humble Bundle. Humble Bundle: PC and Android 10. https://www.humblebundle.com/. Online; accessed 9-Nov-2014.

[10] A. S. Kahn, R. Ratan, and D. Williams. Why we distort in self-report: Predictors of self-report errors in video game play. *Journal of Computer-Mediated Communication*, 2014.

[11] R. Kowert, R. Festl, and T. Quandt. Unpopular, overweight, and socially inept: Reconsidering the stereotype of online gamers. *CyberPsychology, Behavior, and Social Networking*, 2013.

[12] R. Kowert, M. D. Griffiths, and J. A. Oldmeadow. Geek or chic? Emerging stereotypes of online gamers. *Bulletin of Science, Technology & Society*, 32(6):471–479, 2012.

[13] D. J. Kuss and M. D. Griffiths. Internet gaming addiction: A systematic review of empirical research. *International Journal of Mental Health and Addiction*, 10(2):278–296, 2012.

[14] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *ACM World Wide Web Conference (WWW)*, 2010.

[15] G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida. New kid on the block: Exploring the Google+ social graph. In *ACM Internet Measurement Conference (IMC)*, 2012.

[16] W. Mason and A. Clauset. Friends FTW! friendship and competition in Halo: Reach. In *ACM Conference on Computer-Supported Cooperative Work (CSCW)*, 2013.

[17] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *ACM Internet Measurement Conference (IMC)*, 2007.

[18] M. Mitzenmacher. Editorial: The future of power law research. *Internet Mathematics*, 2(4):525–534, 2005.

[19] N. M. Petry. Commentary on van rooij et al.: Gaming addiction–a psychiatric disorder or not? *Addiction*, 106(1):213–214, 2011.

[20] J. Ugander, B. Karrer, L. Backstrom, and C. Marlow. The anatomy of the Facebook social graph. *arXiv preprint arXiv:1111.4503*, 2011.

[21] A. J. Van Rooij, T. M. Schoenmakers, A. A. Vermulst, R. J. Van Den Eijnden, and D. Van De Mheen. Online video game addiction: identification of addicted adolescent gamers. *Addiction*, 106(1):205–212, 2011.

[22] C.-S. Wan and W.-B. Chiou. Psychological motives and online games addiction: Atest of flow theory and humanistic needs theory for taiwanese adolescents. *CyberPsychology, Behavior, and Social Networking*, 9(3):317–324, 2006.

[23] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *ACM International Conference on Web Search and Data Mining (WSDM)*, 2010.

[24] D. Williams, N. Yee, and S. E. Caplan. Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of Computer-Mediated Communication*, 13(4):993–1018, 2008.

[25] N. Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and virtual environments*, 15(3):309–329, 2006.

[26] N. Yee. Motivations for play in online games. *CyberPsychology, Behavior, and Social Networking*, 9(6):772–775, 2006.

APPENDIX

To classify distributions, we use the Python *power-law 1.3* package [1]. First, the tool fits the provided data to power law and lognormal parameters, which are included in our downloadable dataset. Next, because there are multiple classes of heavy-tailed distributions, the tool compares the data between successive pairs of heavy-tailed distribution classes in order to determine the best fit. The tested spectrum spans four classes, ranging from an exponentially bounded distribution at one end (meaning a non-heavy-tailed distribution) to a true power law distribution at the other end. The two intermediate classes are, with respect to the order given above, lognormal and truncated power law. The truncated power law describes a distribution which primarily follows a power law but is "truncated" at the end by an exponentially bounded tail.

Table 4 shows the likelihood values and final classifications resulting from the pair-wise comparisons made by the tool. At each step, the classification process involves three factors. First, for each comparison, the calculated likelihood ratio (R) is compared to zero. A value significantly different from zero allows classification to proceed to the next step. Next, the sign of R indicates which of the two compared distributions is more likely; a positive value prefers the first distribution, while a negative value favors the second. Finally, a p-value states the confidence with which the classification is made. For a small p-value, i.e. ($p < 0.05$), we can state with high confidence that the classification assigned by R is statistically significant. For higher values of p, neither class is a statistically better fit than its counterpart.

Distribution	Power law vs. exponential		Power law vs. lognormal		Truncated power law vs. power law		Truncated power law vs. lognormal		Classification
	R	P	R	P	R	P	R	P	
Account market values	7422.963	$9.15E-104$	-49.578	$4.74E-12$	50.486	0	0.907	0.861	Long-tailed
Account market values (second snapshot)	9685.442	$1.73E-95$	-43.527	$3.39E-10$	52.872	0	9.344	0.082	Long-tailed
Total playtime	455501.857	0	-22961.173	0	18402.092	0	-4559.081	$1.22E-68$	Lognormal
Total playtime (second snapshot)	544385.180	0	-37595.782	0	28263.279	0	-9332.504	$2.45E-31$	Lognormal
Two-week playtime	28049.688	0	-1678.422	0	2172.226	0	493.804	$1.22E-68$	Truncated power law
Two-week playtime (second snapshot)	23981.993	0	-1101.598	$4.83E-204$	1764.264	0	662.666	$5.06E-140$	Truncated power law
Game ownership	5247.774	$1.59E-47$	-31.021	$3.93E-07$	28.565	$4.09E-14$	-2.456	0.602	Long-tailed
Game ownership (second snapshot)	8271.094	$9.75E-56$	-19.209	$3.15E-05$	26.552	$3.16E-13$	7.343	0.069	Long-tailed
Played game ownership	3343.406	$8.39E-32$	-24.832	$2.81E-07$	27.651	$1.03E-13$	2.819	0.494	Long-tailed
Played game ownership (second snapshot)	5914.060	$1.34E-37$	-27.350	$1.85E-08$	33.604	$2.22E-16$	6.254	0.218	Long-tailed
Group size	3381.435	$5.15E-28$	-0.967	0.604	2.097	0.041	1.129	0.541	Heavy-tailed
Group membership per user	4812.540	$7.35E-37$	-13.006	$2.24E-05$	12.374	$6.54E-07$	-0.632	0.808	Long-tailed
Friendship (through 2009)	727.727	$3.43E-15$	-51.378	$5.62E-09$	53.920	0	2.542	0.635	Long-tailed
Friendship (through 2010)	609.157	$1.83E-11$	-80.858	$5.11E-12$	77.856	0	-3.002	0.680	Long-tailed
Friendship (through 2011)	685.893	$1.54E-11$	-97.851	$2.51E-14$	90.753	0	-7.097	0.392	Long-tailed
Friendship (through 2012)	460.516	$1.67E-07$	-138.067	$5.43E-32$	86.126	0	-51.941	$2.24E-09$	Lognormal
Friendship (through 2013)	1921.317	$1.48E-27$	-75.725	$2.88E-13$	82.802	0	7.0767	0.337	Long-tailed
Friendship (2009 only)	245.003	$4.65E-13$	-22.693	$4.03E-05$	20.752	$1.18E-10$	0.924	0.0683	Long-tailed
Friendship (2010 only)	402.606	$6.34E-11$	-30.243	$6.13E-07$	24.288	$3.18E-12$	-5.954	0.120	Long-tailed
Friendship (2011 only)	614.630	$5.51E-14$	-29.295	$4.35E-06$	24.177	$3.56E-12$	-5.118	0.212	Long-tailed
Friendship (2012 only)	653.744	$2.16E-14$	-38.515	$7.61E-08$	27.385	$1.35E-13$	-11.130	0.0219	Lognormal
Friendship (2013 only)	1864.431	$5.95E-25$	-61.474	$8.93E-12$	46.751	0	-14.724	0.027	Lognormal

Table 4: Distribution classification and likelihood tests