

Thesis draft

Joshua Megnauth

September 21, 2020

1 Introduction: Reddit, gamers, and social networks

Gaming is an omnipresent artistic medium enjoyed by a plurality of Americans. Research from the *Entertainment Software Association* (E.S.A.) has found that about 65% of American adults play video games. About 46% of said gamers are female (Association, 2019). The E.S.A.'s research counters the tired trope of gamers as young, immature boys. Despite the prevalence and diversity of gamers academia is woefully behind on ludology, the study of gaming. Gaming lacks the prestige of other artistic entertainment media such as books, music, and film. An interested party may find texts on the innovations of Allen Ginsberg's poem *Howl* or Sonic Youth's *Daydream Nation* while scarcely finding an academic article on the contributions of the video game *Doom*.

Ludology is as exciting as fields such as A.I. despite lacking the selfsame effervescence. In other words, ludology may be studied from many different angles. Political sociologists would find much to mine from the philosophical schisms in the gaming community. Recent grassroots movements to push for more representation in video games spawned a revanchist and sexist counter movement. Besides the political, sociologists may write ethnographies of the online communities or interactions that occur within gaming. Mark Chen's *Leet Noob: The Life and Death of an Expert Player Group in 'World of Warcraft'* is a study of a group of gamers who tackle dungeons in *World of Warcraft*. Chen writes from the perspective of both a gamer and a social scientist to apply theories such as social and cultural capital, actor-network theory, social construction, et cetera to the video game (Chen, 2011). The list above is clearly not exhaustive. However, I'd like to mention, as an aside, that online communities are transient; ignoring ludology risks permanently missing out on an aspect of culture.

My thesis focusses on the intersection between network science via the social network Reddit and video games. I am particularly interested in the distribution of gamers across subreddits. Reddit and the related terminology will be explained below. I hypothesize that gamers are more likely to post on multiple gaming subreddits (that is, gamers are more likely to have a community that spans across subreddits). Analyzing the network dispersion of gamers is useful for both the social sciences as well as marketing. Network science is concerned with the spread of information by virtue of its focus on ties. Users who post on multiple subs may pass information across subreddits.

My guiding principle is thus to use my statistical and computing skill coupled with my domain knowledge of video games in order to contribute to ludology. Previous studies on gaming lack domain expertise while exhibiting logical fallacies—both of which seem to be caused by the hermetic world of academia. More proper ludology needs to seed the field until the study of video games resembles writing on music or films.

An anecdote engendered my thesis. I've observed; anecdotally, like many others; that nerds of a feather tend to flock together. Gamers tend to share a sample from a common set of interests such as anime, wrestling, technology, certain tastes in music, et cetera. This observation is far from holistic—meaning, gamers do not adopt all of the same traits, and clearly not every person who watches wrestling, for example, is a gamer. However, the basic observation led to my thesis. My research is not seeking to answer that question specifically, but my thesis may provide the groundwork toward studying that larger topic. Gamers may preferentially attach to other gamers or to other ostensibly related subjects such as anime or eSports. Either conclusion—that gamers preferentially attach or not—is useful for ludology.

2 Background: Social network analysis

2.1 A digital approach

My goal is to contribute to both ludology as well as studies of Reddit. Gaming by virtue is a heavily technological affair. Gamers are often only lightly insulated from technology. For example, anyone who follows gaming news would likely run into computer terms. Even gamers who do not follow gaming news would run into patches, lag, bugs, or have to know how to connect their consoles to the internet, et cetera. Gaming is unlike television or cable in that video games have a strongly "digital native" feel—especially for P.C. (computer) gamers. Gamers discuss video games online on platforms such as Reddit or news sites with forums and comment systems. The culmination of

these "techy" qualities does not imply that all gamers are computer scientists but that they are generally reasonably comfortable with technology. All of this is to say that a network analysis focussing on gamers and Reddit is sensible because of the focus on internet media.

2.2 Why video games?

Gamers and ludology, as discussed earlier, are legitimate areas of study that are largely ignored by academia as well as having an image problem regardless of gaming's prevalence. Gaming as a medium is sometimes strangely maligned as coercing individuals toward violence. In 2019, President Donald Trump as well as Kevin McCarthy and Dan Patrick, both Republican politicians, scapegoated video games as reported by an article in the New York *Times*. The article explains that games are often blamed for shootings despite the lack of causal evidence. The American Psychological Association, according to the *Times*, is that video games and violence are not linked. Dr. Chris Ferguson of the A.P.A. humorously mentioned that the data linking bananas to suicide are about the same as those data connecting video games and other media to violence (Draper, 2019).

Most research published on gaming is cursory while also lacking domain knowledge on the subject. A search through databases that compile articles, such as JSTOR, on video games shows that research is decidedly lacking in imagination as they mostly focus on the antiquated violence question. One such article published in *Frontiers of Psychology* exhibits a very flawed methodology in studying gamers. The researchers use an imbalanced and relatively small random sample that skews young and male. Females consisted of a scant 15% of their sample. The writers justified the skewed sample by referencing a two decade old paper despite publishing their article in 2019 and using recent sources otherwise (von der Heiden, Braun, Muller, and Egloff, 2019). Research from the NPD Group corroborates what the E.S.A. found and gamers already know: about half of gamers are women and more females own the 3DS, Wii U, and Nintendo Switch than males (Valentine, 2019). Beyond the poor sampling the research heavily relies on p-values without associated effect sizes. Furthermore, the basis for the study is intrinsically flawed. Gamers are a wide social unit. We would not sample around 3000 music buffs in order to see if "problematic music listening" is associated with "poor psychological functioning."

2.3 What is Reddit?

Reddit is a social network reminiscent of the forums and bulletin board systems (B.B.S.) of the past. Reddit is designed for lengthy deliberation, or at least facilitates it, which is like forums and unlike Twitter or Facebook. Subreddits is Reddit parlance for a community which is similar to a forum. Pseudonymous users may post or reply to discussion topics. Individual subreddits are moderated by screened volunteers from the community; rules are enforced and users may not be able to post if their reputation—known as karma—is too low. Needless to say, Reddit’s structure engenders higher quality content—generally speaking at least as Reddit is also home to lengthier discussions on puerile matters. Reddit users, who are pseudonymous, are known by the demonym Redditors.

Reddit reflects gaming’s preeminence via *many* the many subreddits on the topic; individual consoles, series, genres, abstruse memes, as well as technological concepts such as emulation all have incredibly active subs. Reddit is the seventeenth most popular site in the world according to *Alexa*, a subsidiary of Amazon that collects internet data. The social network’s position varies from day to day but remains consistently high. Reddit is ahead of several prominent sites such as Netflix, Hulu, Crunchyroll, Instagram, Microsoft’s site, Live (Outlook), Twitch, Bing, eBay, AliExpress, and others. Reddit is mainly behind sites such as Chinese social networks such as Weibo as well as Google and YouTube Alexa, 2020. The largest gaming subreddit on Reddit, */r/gaming*, is the fourth most popular total which places *gaming* above the subreddits for music, sports, relationships, movies, news, gifs, cats, cute pictures, books, programming, et cetera. In fact, the only subreddit more popular than *gaming* that is *not* a default subscription is */r/funny*. The subreddit for *League of Legends*, a single video game, is the sixtieth most popular subreddit Sizz, 2020.

Redditors may post on as many or as little subreddits as they wish. For example, a user may post on the subreddits for PlayStation 4, P.C. gaming, Linux, and the systems programming language, Rust. Another user may frequent the subs for K-Pop, anime, and the Nintendo Switch.

2.4 Big data—Volume, Variety, and Velocity

2.5 Weak ties in social networks

To doGranovetter, 1973

2.6 Co-posting networks via Reddit

To do

3 Methodology

```
2020-09-19T09:59:19.002Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T09:59:32.722Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=PS3&size=1000&sort=desc&before=1600382930.
2020-09-19T09:59:32.724Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T09:59:46.253Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=PS2&size=1000&sort=desc&before=1600286005.
2020-09-19T09:59:46.254Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T09:59:58.502Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=PSX&size=1000&sort=desc&before=1600189647.
2020-09-19T09:59:58.504Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:00:13.376Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=emulation&size=1000&sort=desc&before=1600285827.
2020-09-19T10:00:13.377Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:00:24.729Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=otomegames&size=1000&sort=desc&before=1600380344.
2020-09-19T10:00:24.731Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:00:35.936Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=demonssouls&size=1000&sort=desc&before=1600428089.
2020-09-19T10:00:35.937Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:00:47.549Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=metalgearsolid&size=1000&sort=desc&before=160038852
4.
2020-09-19T10:00:47.555Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:01:11.330Z ERROR thesis_gamer_scraper::scraperclient::scraperclient > Request: error decoding response b
ody: EOF while parsing a value at line 1 column 0 @ https://api.pushshift.io/reddit/comment/search?sort_type=created_ut
c&subreddit=gaming&size=1000&sort=desc&before=1600445951
2020-09-19T10:01:11.330Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Sleeping: 10 seconds
2020-09-19T10:01:24.788Z INFO thesis_gamer_scraper::scraperclient::scraperclient > Scraped 100 nodes from https://api
.pushshift.io/reddit/comment/search?sort_type=created_utc&subreddit=LeagueOfLegends&size=1000&sort=desc&before=16004452
10.
```

Figure 1: Scraper hard at work

My research project requires relatively recent data as well as observations relevant to video games. Researchers in the computational social sciences seem to trend toward gathering their own data. The internet is a dynamic space of constant data production. Structured data may be limited or flawed for many topics, including ludology. Scraping data allows computational social scientists to gather raw, unpasteurized data rather than relying on old, static sources. A survey, for example, is saliently flawed for gamers because conducting one properly would likely require a large sample across a wide span of ages as well as capture gamers who play on different platforms. Scraping data seems both more reliable as well as more direct in a case like that as well as for my project.

While large Reddit data sets exist via the *Stanford Network Analysis Project* as well as *Kaggle*, scraping data from Reddit is not too difficult due to the open Reddit API [Reddit, 2020](https://api.pushshift.io/). Bindings for Python exist via *PRAW* as well. However, the Reddit API imposes several restrictions and requires an account to use properly as well. The restrictions are reasonable due to Reddit's high traffic: rapacious scrapers may hammer the site and associated Content Delivery Networks. Caching services, such as the open

source Pushshift, aims to alleviate the stress caused by scrapers as well as provide a more convenient API Baumgartner, n.d. I opted for Pushshift to gather my data due to its flexibility.

My scraper gathers N Node-Edge pairs from a provided list of subreddits. I wrote the scraper in the systems programming language, Rust, and the source code is available on GitHub¹. My program's high level logic is as follows:

1. Collect 100 nodes with associated metadata from a subreddit. Each observation contains the Redditor name, epoch timestamps, and subreddit name.
2. Paginate by setting the *before* API parameter to the earliest timestamp from the recent scrape of 100 nodes.
3. Add each gathered node to a hash set which precludes duplicate observations.
4. Repeat steps one through three for the set of subreddits in a round robin manner.
5. Repeat steps one through four until N is reached.
6. Remove deleted users and automoderator replies.
7. Hash all of the Redditor names with SHA256 at an attempt at privacy².
8. Save everything to a CSV.

My data set³ contains about 78,000 observations gathered via the logic above.

References

Alexa. (2020). Alexa - Top Sites. Retrieved from <https://alexa.com/topsites>
Association, E. S. (2019). 2019 Essential Facts about the Computer and Video Game Industry. Retrieved from <https://theesa.com/esa-research/2019-essential-facts-about-the-computer-and-video-game-industry/>
Baumgartner, J. (n.d.). Pushshift API. Retrieved from <https://pushshift.io>

¹https://github.com/joshuamegnauth54/thesis_gamer_scraper

²Very little of the data sets found online actually hashed the user names as the information is all public. However, hashing at least adds a minimum layer of privacy.

³Available here: <https://github.com/joshuamegnauth54/GamerDistributionThesis2020/data>

- Chen, M. (2011). *Leet Noob: The Life and Death of an Expert Player Group in 'World of Warcraft'*. Peter Lang Inc., International Academic Publishers.
- Draper, K. (2019, August 5). Video Games Aren't Why Shootings Happen. Politicians Still Blame Them. *New York Times*.
- Granovetter, M. S. (1973). The Strength of Weak Ties. *The American Journal of Sociology*, 78, 1360–1380.
- Reddit. (2020). Reddit API. Retrieved from <https://reddit.com/dev/api>
- Sizz, M. (2020). Retrieved from <https://redditlist.com>
- Valentine, R. (2019). EEDAR: Nintendo Switch Attracting More Women, Wider Age Ranges Over Time. Retrieved from <https://www.gamesindustry.biz/articles/2019-02-11-eedar-nintendo-switch-attracting-more-women-wider-age-ranges-over-time>
- von der Heiden, J., Braun, B., Muller, K., & Egloff, B. (2019). The Association Between Video Gaming and Psychological Functioning. *Frontiers in Psychology*.