Analysis of trends in Video Games

Anuj Singh
anujs@iitbhilai.ac.in
Indian Institute of Technology, Bhilai
Raipur, Chhattisgarh

ABSTRACT

We have created a dynamic network where the nodes are games and the weighted edges are the inverse of the number of common tags between games. The data has been collected from the Steam API and SteamSpy API and the data consists of games released from 2004 to 2019. We analyze the network and draw conclusions regarding what makes games popular, and other general trends as observed from the network.

KEYWORDS

Game Network, Complex Networks

ACM Reference Format:
Anuj Singh. 2021. Analysis of trends in Video Games. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The late 20th century saw the exponential rise of computers and by the the 21st century, they have become a part of our daily lives. Video games, piggy backing on the advance of computers, have risen in popularity and are quickly becoming the preferred mode of entertainment for many. This is evident from the rise in revenue the gaming industry has seen in the past half decade. From revenues of under 1 billion dollars in 1971 to 140 billion dollars in 2018 [9]. The driving force for this unprecedented growth has been shifting over the decades. From the rise of the arcade in 1980's, to personal computers and consoles in the 1990's to the boom due to mobile gaming in 2010's. Lately Virtual Reality games have entered the market to push the envelope further.

Movie recommendation systems have been around a long time and have moved from network based predictions to Machine Learning based predictions. Movie networks have been constructed for various reasons such as, identification of key films by Bioglio et al. [5] in which they attempted to key films and personalities of western films to find milestones in western cinema. Grujić, Jelena [7], aimed to use biparitite graphs for recommending movies. In the context of video games, similar analysis is just at its nascent stage, as the largest distributor of PC games, Steam only started recommending games in 2014, and pushed the second update on 2016. Thus analysis of game networks for various purposes that exist in their counterparts in the entertainment industry, such as movies and books, is yet to be done.

We have created a dynamic network where the nodes are games and the weighted edges are the inverse of the number of common tags between games. The nodes (games) appear according to their release date. This can have a variety of applications, from game recommendations to studying trends in the gaming industry such as dominant genres. Along with rich metadata such as game ratings, price, approximate users and other details, interesting conclusions can be made from the network.

2 RELATED WORKS

Gaming networks have been analyzed in the context of social networks, such as by Becker et al. [4] where they study the Steam Community network, which consists of users in the registered on Steam. They analysed basic characteristics and development over time of this network. Games have been studied in other contexts such as the impact of video games on their audience by Kühn, Stefan and Milasi, Santo [8]. Not much work has been done on game networks. Work has been done on book networks by Ziegler et al.[11] for book recommendation lists, Weng et al. [10] anlysed an actor role network to identify leading roles and hidden communities. The Brazillian Popular Music network was studied by Silva et al [6] to show that it was an example of a small world network.

3 DATA COLLECTION AND NETWORK FORMATION

Steam[1] is a video game digital distribution service by Valve. It has a catalogue of thousands of games for the PC platform. Steam provides an API[2] to access some metrics related to games such as price, release date, description, genres and several others. SteamSpy[3] adds additional user tags to separate games into different smaller genres. The data was collected using these two APIs over a period of a week. The data consists of the games on the Steam Platform from 2004 to 2019. A small number of games made by Steam's parent company, Valve already existed on Steam from before 2004, but there release data has been set to 2004 for normalization purposes. There were multiple hurdles in collecting data. Steam does not provide an easy way to distinguish between videos, demos and other products on its site other than games. In total there are about 90000 products on its platform, which includes games and other items previously mentioned. To overcome this hurdle, data had to be collected from the SteamSpy API first, which does allow only to query for games. This reduces the number of data points to be around 33000 to be fetched from the Steam API. Each data point included the name of the game, its release date, genre, description and many other things. User tags needed to be

Conference'17, July 2017, Washington, DC, USA

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

Table 1: Basic network paramters

Nodes	3326
Edges	115086
Avg. Clustering Coefficient	0.72
Diameter	9
Average Degree	70.475
Average Weighted Degree	65
Density	0.022
Modularity Index	0.675
Connected Components	87
Average Path length	3.42

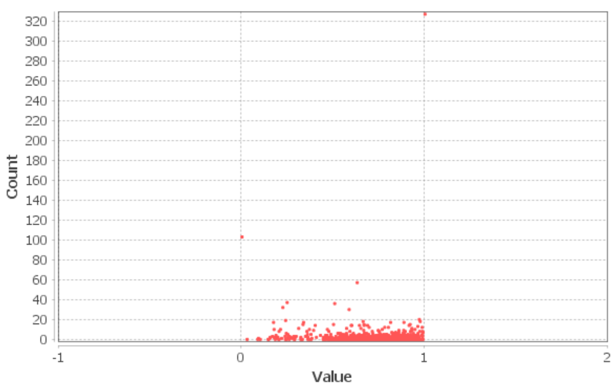

Figure 1: Clustering Coefficient Distribution

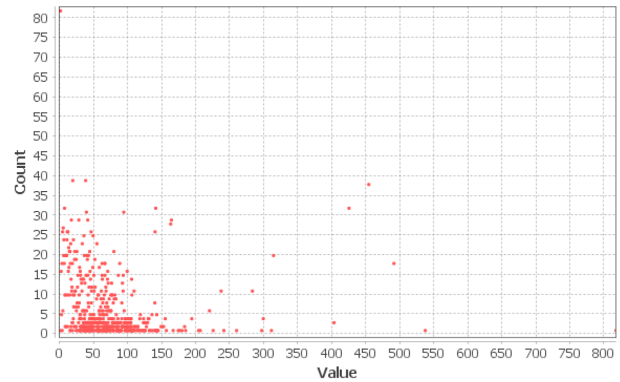

Figure 2: Weighted Degree Distribution

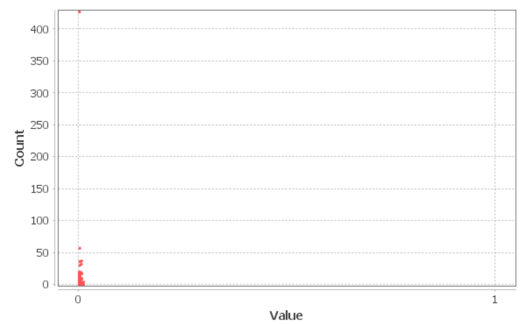

Figure 3: Betweenness Centrality Distribution

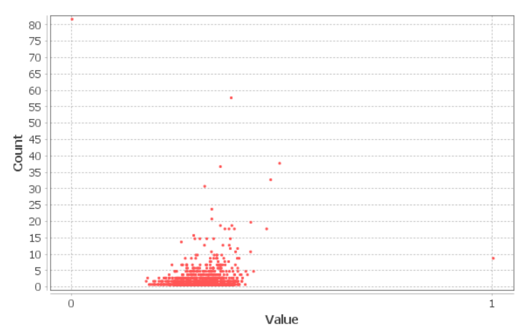

Figure 4: Harmonic Closeness Centrality Distribution

collected from the SteamSpy API. This created another hurdle as SteamSpy contains hundreds of tags, which would make it impossible to link any two games based on similar tags. Thus the number of tags had to be cut down, to the top 20 tags. Weighted Edges were created between two games by counting the number of similar tags between them and using the inverse of this count as the weight of the edge. Unfortunately, this created a very dense network of over 33000 nodes and over 3 million edges, which would be near impossible to analyze. To remedy this issue, the number of games were reduced using the total count of ratings recieved by a game. This filters out less popular games. The threshold was set to 500 total ratings. This reduced the number of nodes to around 3000. Finally another approach had to be used to reduce the number of edges in the graph. The tag data was limited to only 3 per game, and games must have at least 2 tags in common to form an edge between them. This reduced the number of edges to around 100000. Finally the dataset was managable.

4 NETWORK DESCRIPTION

The basic network characteristics are summarized in Table 1.

The clustering Coefficient Distribution is given in Figure 1. Since the average Clustering Coefficient is quite high (0.72) there are a large number of games on the tail.

The Weighted Degree Distribution is given in Figure 2. It is evident that the weighted degree is quite high further suggesting that this is a dense network.

The Betweenness Centrality Distribution is shown in Figure 3. Since the network is very dense, the average Betweenness Centrality is very low. The Harmonic Closeness Centrality Distribution is shown in Figure 4. The average Closeness Centrality is rather on the lower end as evident from the distribution.

5 RESULTS

Running community classification on the game network, produces 11 major community classes as evident from Figure 5. These communities are based on tags, partitioning on which allows us to visualize various paramters by keeping the context of tag data, this

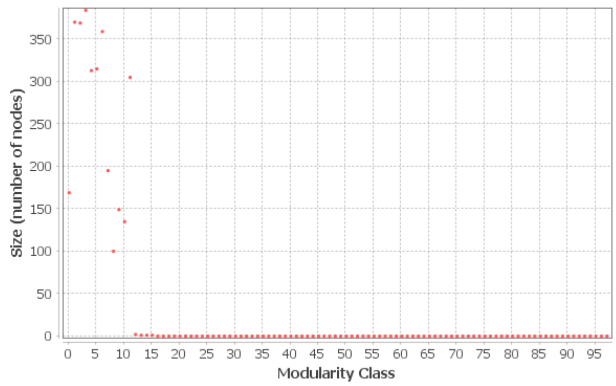

Figure 5: Size of Communities

is represented by the different colors present in the network figures. The major communities are Free To Play, FPS, RPG, Open World, Strategy, Simulation, Horror, Fighting, Platformer, Point and Click and 2D. Using this context we can also find the top games by revenue, which is a part of the metadata as shown in Figure 6. The revenue is distributed across genres, but it is evident that FPS games have a larger share of the total revenue, even after the low number of games in the genre (FPS is represented by blue)

If we look at the games by the total number of players, as shown in Figure 7, we see apart from the strategic RPG Dota 2, it is again dominated by FPS games such as Player Unknown Battlegrounds and CS:GO. Based on the Figure 7 and Figure 6 we can conclude that FPS games, although being lesser in number as compared to other genres, enjoy disproportionately higher revenues and user base.

Coming to the Centrality Measures, we first need to understand what the Centrality Measures represent in these networks. Since the nodes are joined using genres(tags), a high Betweenness node will connect diverse genres(tags). As such the tags associated with high Betweenness Nodes are diverse and consist of a rich variety. There are a very few games which meet this criteria. Additionally, the high Betweenness games were some of the least popular and lowest grossing games. The Spearman's rank correlation for Betweenness and Revenue was -0.2. Total ratings are a measure to gauge user engagement with games. The Spearman's rank correlation for Betweenness and Total Ratings was -0.09. This leads us to conclude that players like games which refine and pertain to a specific genre, rather than combining elements of diverse genres. The network is quite dense, and the average path length is very small (3.42) which indicates that there are at least some common elements associated with most games, which allow easy transition from one genre to another. Real world example exist to show how easy it is to transition from one genre to a completely different genres jumping only through a few games. Finally looking at the weighted degree, we find that most top genres have hubs inside them, i.e., nodes which have a large number of links within there own community. This tend to service the exact genre it belongs to. Unfortunately we could not find any correlation between the weighted degree (how strongly it services a particular genre) and revenue.

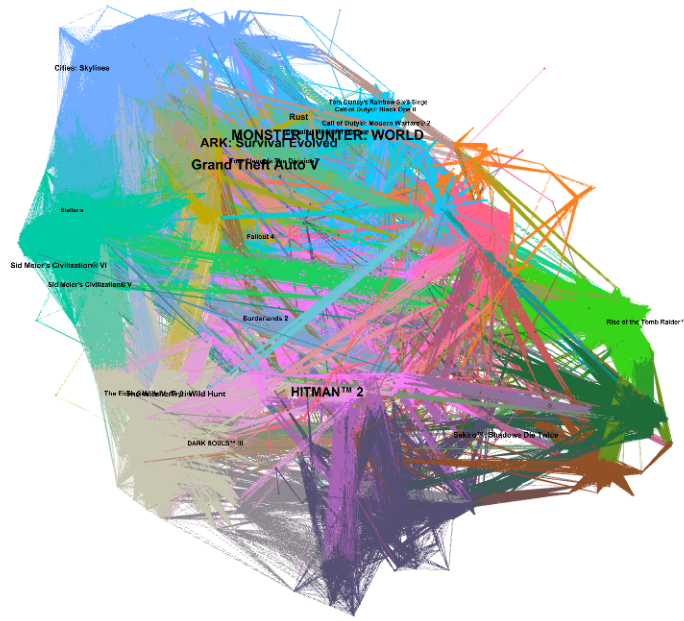

Figure 6: Top Games By Revenue. Most high grossing games lie in the light blue area of the graph, representing FPS games

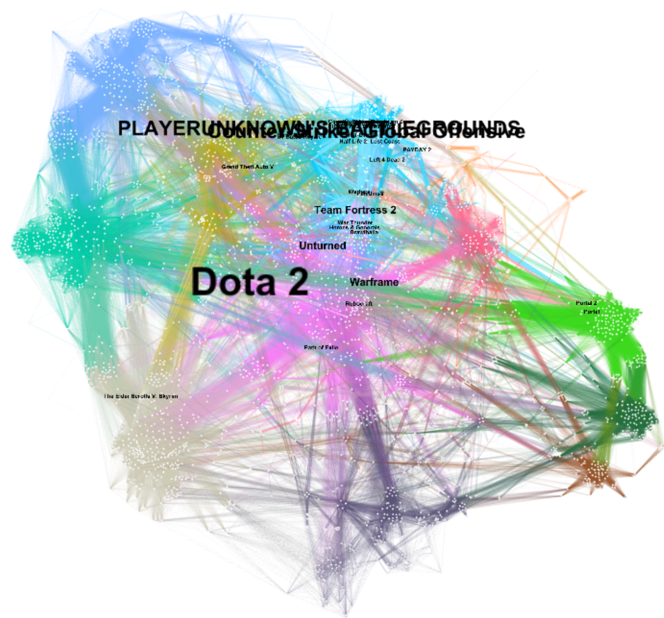

Figure 7: Top Games By Number of Players. The top games with the largest user base are Dota 2, Counter Strike:Global Offensive and Playerunkown Battlegrounds

6 TEMPORAL ASPECTS

We analyzed the growth rates of various genres calculated on a per year basis. The results are shown in Figure 9 We see that the top

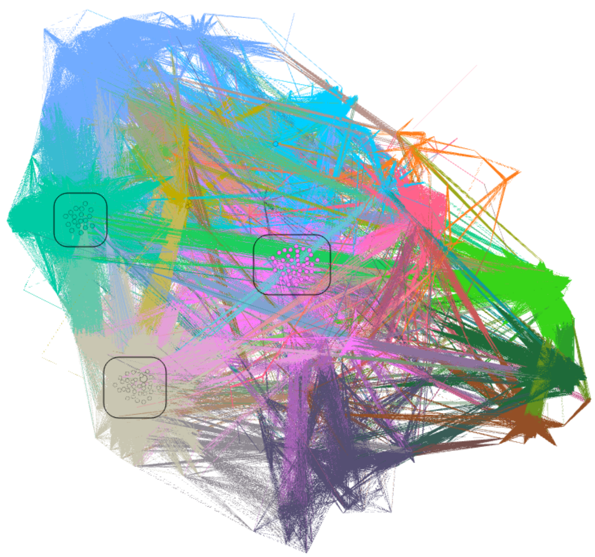

Figure 8: Weighted Degree. Encircled parts show a collection of High Weighted Degree nodes at the center of their genres

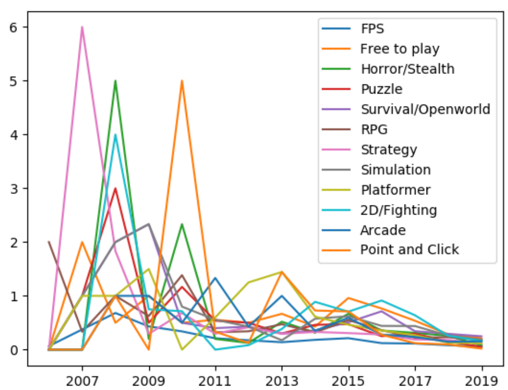

Figure 9: Growth rate in terms of number of games in genres

growing genre is different for different years. Growth rates of all genres decrease over time as the number of games start increasing leading to lower growth rates even if more games are released during the year.

7 CONCLUSIONS AND FUTURE WORK

We conclude that FPS games have a disproportionately larger user base and revenue compared to the number of games in the genre. This shows that the genre has massive growth opportunities. We find that players tend to not like games which are a mishmash of different genres, rather they prefer games which refine genres and pertain to the specific genre. We see that it is relatively easy for players to transition from one genre to another within a few games. Finally we observe a general trend in popular genres over time. In future we can analyse this network on a more advanced level, categorizing this graph and the impact of external factors such as recession on the graph. Steam provides rich metadata and along with 3rd party APIs such as SteamSpy, there are extensive attributes to analyze in the game network. This work briefly touches the tip of the analysis that can be done using the metadata. For example, game descriptions can be used to create better links between games. User ratings can be associated and trends can be observed between ratings, revenue and genre. Using these along with game descriptions, powerful recommendation systems can be made of which there is a lack of currently.

REFERENCES

[1] [n.d.]. https://store.steampowered.com/
[2] [n.d.]. https://steamcommunity.com/dev
[3] [n.d.]. https://steamspy.com/
[4] Roi Becker, Yifat Chernihov, Yuval Shavitt, and Noa Zilberman. 2012. An analysis of the steam community network evolution. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*. IEEE, 1–5.
[5] Livio Bioglio and Ruggero G Pensa. 2018. Identification of key films and personalities in the history of cinema from a Western perspective. *Applied Network Science* 3, 1 (2018), 50.
[6] D de Lima e Silva, M Medeiros Soares, MVC Henriques, MT Schivani Alves, SG De Aguiar, TP De Carvalho, G Corso, and LS Lucena. 2004. The complex network of the Brazilian Popular Music. *Physica A: Statistical Mechanics and its Applications* 332 (2004), 559–565.
[7] Jelena Grujić. 2008. Movies recommendation networks as bipartite graphs. In *International Conference on Computational Science*. Springer, 576–583.
[8] Stefan Kühn and Santo Milasi. 2018. Executive Summary. *World Employment and Social Outlook* 2018, 1 (2018), 1–3.
[9] Yuki Nakamura. [n.d.]. Peak Video Game? Top Analyst Sees Industry Slumping in 2019. https://www.bloomberg.com/news/articles/2019-01-23/peak-video-game-top-analyst-sees-industry-slumping-in-2019
[10] Chung-Yi Weng, Wei-Ta Chu, and Ja-Ling Wu. 2007. Movie analysis based on roles' social network. In *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 1403–1406.
[11] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*. 22–32.