Analysis of Video Game Network

Anuj Singh

Indian Institute of Technology, Bhilai, India
anujs@iitbhilai.ac.in

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

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.

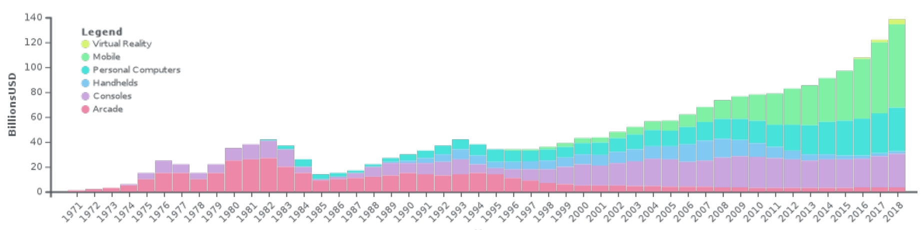

Fig. 1. Global revenues of the video game industry from 1971 to 2018, not adjusted for inflation[1].

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[3], movie recommendation systems[4], etc.. 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, etc.., interesting conclusions can be made from the network.

2 Related Works

Gaming networks have been analyzed in the context of social networks, such as in [2], and in other contexts such as the impact of video games on their audience [5]. Not much work has been done on game networks. Work has been done on movie networks [6], book networks [7], music networks [8] etc..

3 Data Collection and Network Formation

Steam[10] 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[9] to access some metrics related to games such as price, release date, description, genres, etc.. SteamSpy[11] 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. 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 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.

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

The clustering Coefficient Distribution is given in Figure 2. Since the average Clustering Coefficient is quite high (0.72) there are a large number of games on the tail.

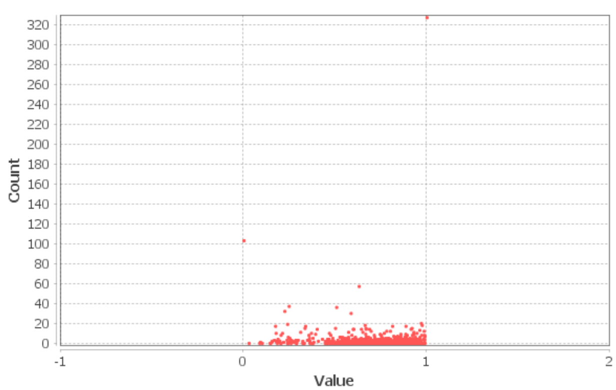

Fig. 2. Clustering Coefficient Distribution

The Weighted Degree Distribution is given in Figure 3. It is evident that the weighted degree is quite high further suggesting that this is a dense network.

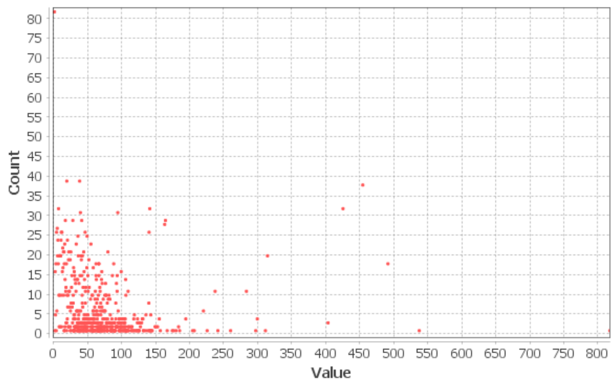

Fig. 3. Weighted Degree Distribution

The Betweenness Centrality Distribution is shown in Figure 4. Since the network is very dense, the average Betweenness Centrality is very low. The Harmonic Closeness Centrality Distribution is shown in Figure 5. The average Closeness Centrality is rather on the lower end as evident from the distribution.

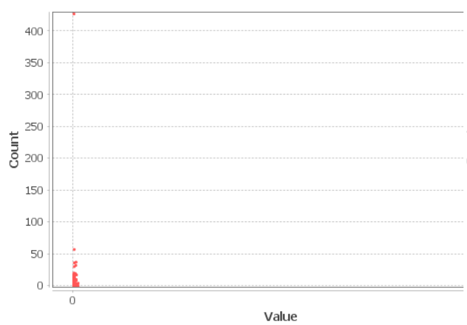

Fig. 4. Betweenness Centrality Distribution

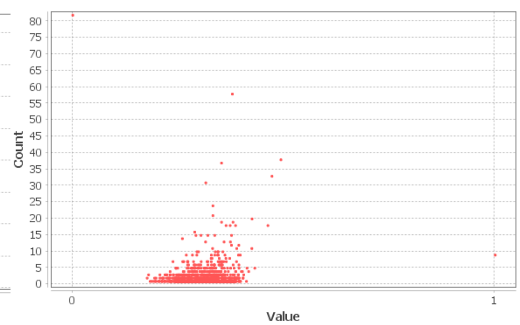

Fig. 5. Harmonic Closeness Centrality Distribution

5 Results

Running community classification on the game network, produces 11 major community classes as evident from Figure 6. These communities are based on tags, partitioning on which allows us to visualize various paramters by keeping the context of tag data, this 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

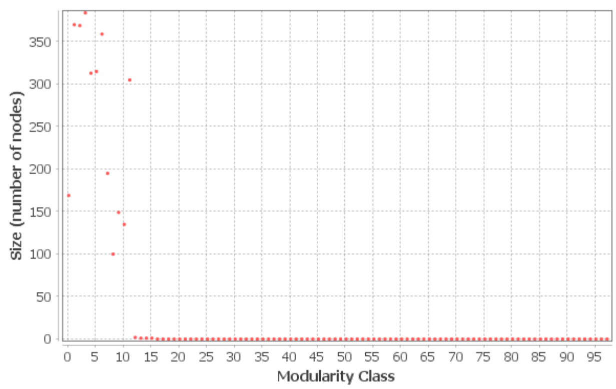

Fig. 6. Size of Communities

part of the metadata as shown in Figure 7. 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 8, 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 8 and Figure 7 we can conclude that FPS games, although being lesser in number as compared to other genres, enjoy disproportionately higher revenues and user base.

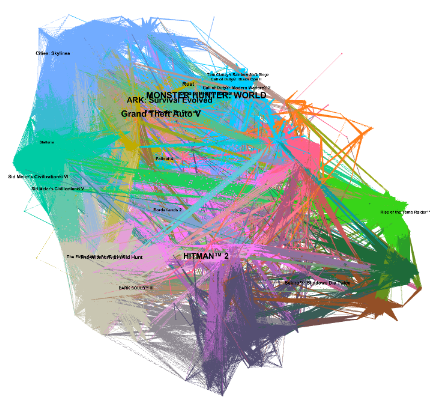

Fig. 7. Top Games By Revenue

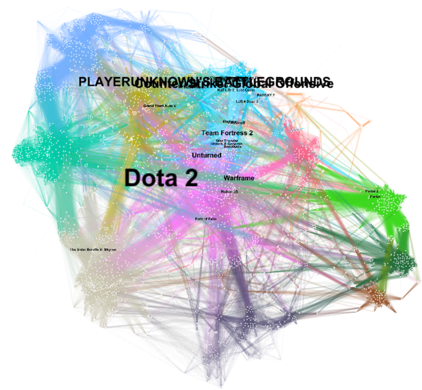

Fig. 8. Top Games By Number of Players

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. Looking at Figure 9 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. 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.

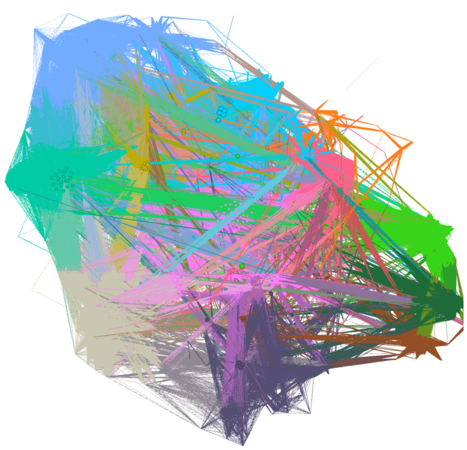

Fig. 9. Betweenness Mapped to Node Size

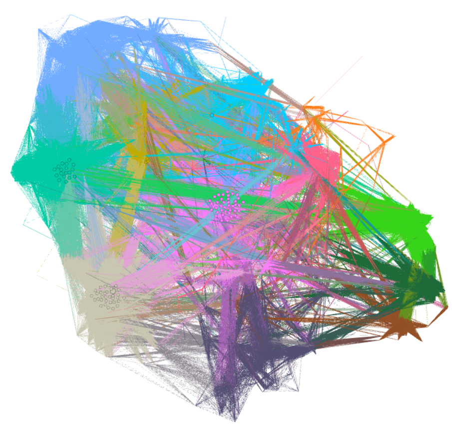

Fig. 10. Weighted Degree Mapped to Node Size

Finally considering the temporal aspects of the graph, we observe that some genres(tags) are more popular than others in different periods of time. As a general trend we saw the most popular genre(tag) transition from FPS to Strategy to RPG to Simulators to Open World over the period we have the data from (2004-2019).

6 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.
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. Naramura, Yuki (January 23, 2019). "Peak Video Game? Top Analyst Sees Industry Slumping in 2019". Bloomberg L.P.
2. Becker, Roi & Chernihov, Yifat & Shavitt, Yuval & Zilberman, Noa. (2012). An analysis of the Steam community network evolution. 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, IEEEI 2012. 1-5. 10.1109/EEEI.2012.6377133.
3. Bioglio, L., Pensa, R. Identification of key films and personalities in the history of cinema from a Western perspective. Appl Netw Sci 3, 50 (2018). https://doi.org/https://doi.org/10.1007/s41109-018-0105-0
4. Grujić, Jelena. (2008). Movies Recommendation Networks as Bipartite Graphs. 576-583. 10.1007/978-3-540-69387-1_66.
5. Kühn, S., Kugler, D., Schmalen, K. et al. Does playing violent video games cause aggression? A longitudinal intervention study. Mol Psychiatry 24, 1220–1234 (2019). https://doi.org/https://doi.org/10.1038/s41380-018-0031-7
6. Tatnall, Arthur & Gilding, Anthony. (1999). Actor-Network Theory and Information Systems Research.
7. Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In Proc. Int. World Wide Web Conf., pages 22–32, 2005.
8. Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. The Yahoo! Music dataset and KDD-Cup'11. In JMLR Workshop and Conf. Proc., volume 18, pages 3–18, 2012.
9. https://steamcommunity.com/dev
10. https://store.steampowered.com/
11. https://steamspy.com/