Will my new videogame be *successful* ?

Group 21
Maxime Bonnesoeur
Michael Hodara
Tarik Kamel

EPFL

A Network Tour of Data Science, 2019-2020

Story

The video games market is expected to be worth over $**90B by 2020, from nearly $78b in 2017.** (wepc.com)

80% of the US video game industry 's $36B revenue in 2017 belongs **to software sales.**
(Entertainment Software Association, NPD Group, 2017)

There are more than **2.5 billion** gamers all over the world. (Newzoo)

What is STEAM ?

- Valve, a videogame developer created the largest PC games distribution platform
- Holds 75% of the PC games sales market in 2013 (Bloomberg)
- Created in 2003, has now more than 34'000 titles in their catalog (SteamSpy)
- $4.3B in revenue in 2017 (SteamSpy, not counting in-app purchases

Agenda

Data acquisition

Parsing of the data from API
First cleaning of the data

Exploration

Understanding the data
Regression
Graph Representation
Final cleaning of the data

Exploitation

What makes a game successful ?
Graph Analysis
Spectral Clustering

Conclusion and Q&A

Discussion of the results and reflections

Our Data

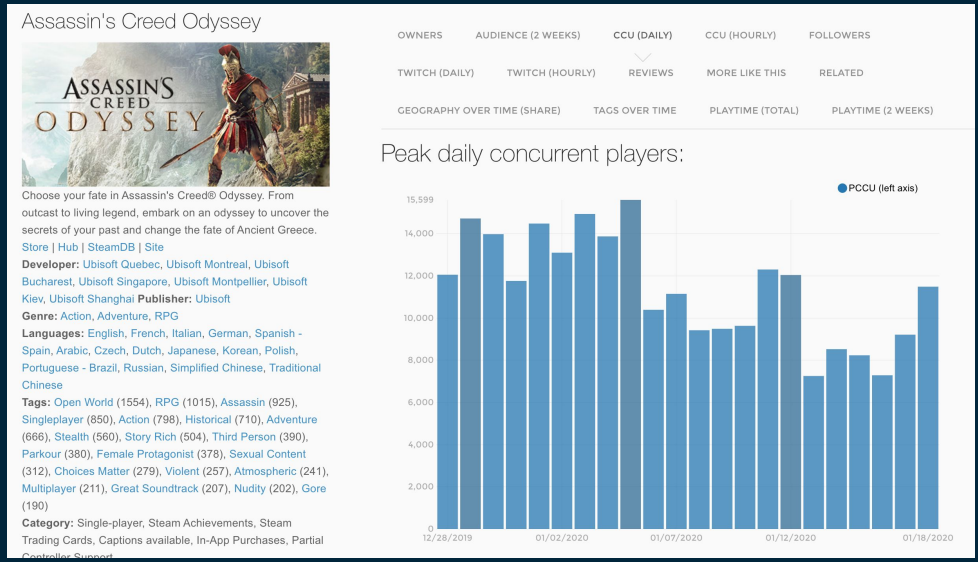
Steam API

- Information on the video game (support, language, genre, ...)
- Ratings of the games

Our Data

SteamSpy API

- Gather statistical data on Steam
- Sales data and popularity



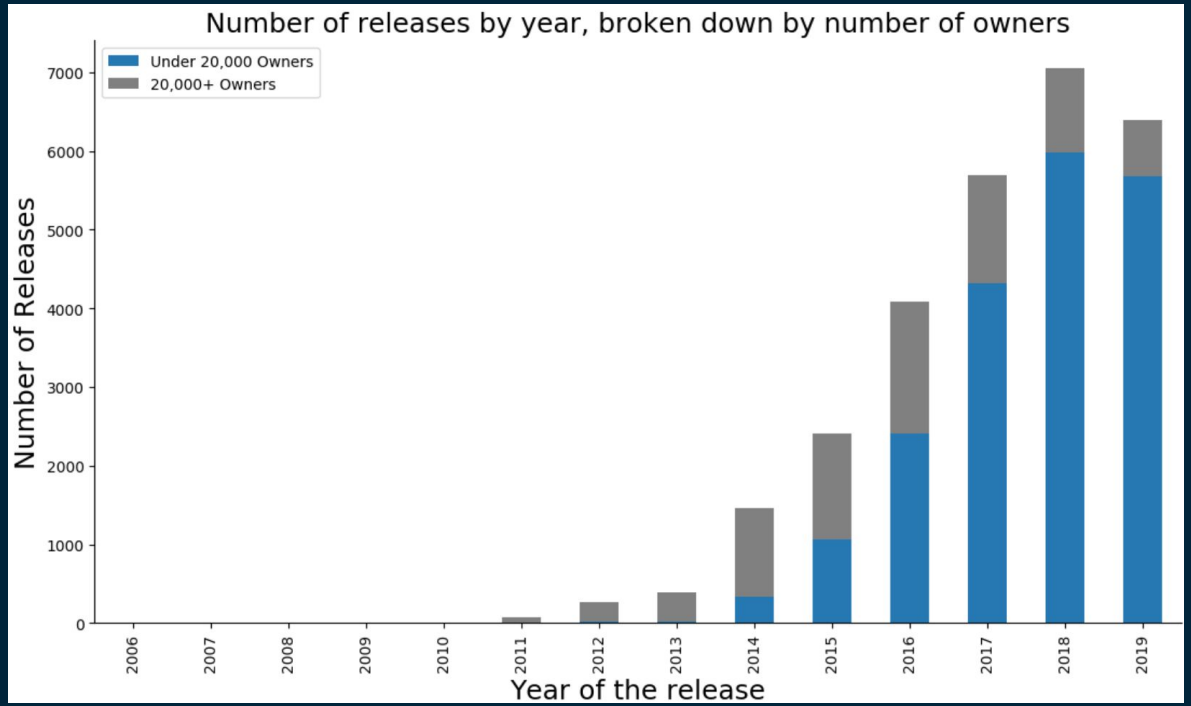
A Network Tour of Data Science, 2019-2020

Data collection

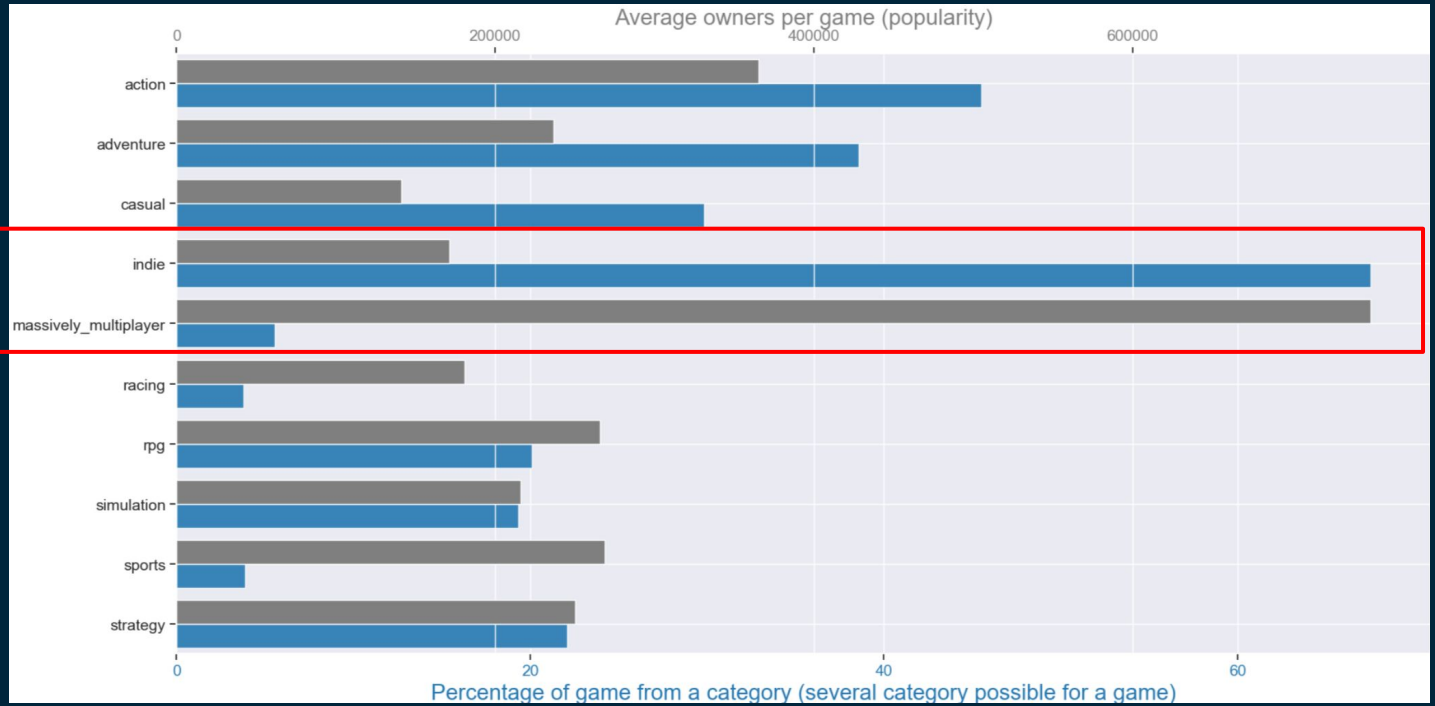
Why parsing the data by ourselves ?

- Data from the swiss platform
- Data as fresh as possible

Game releases



Game repartition on the platform



Defining success

- Economic Success

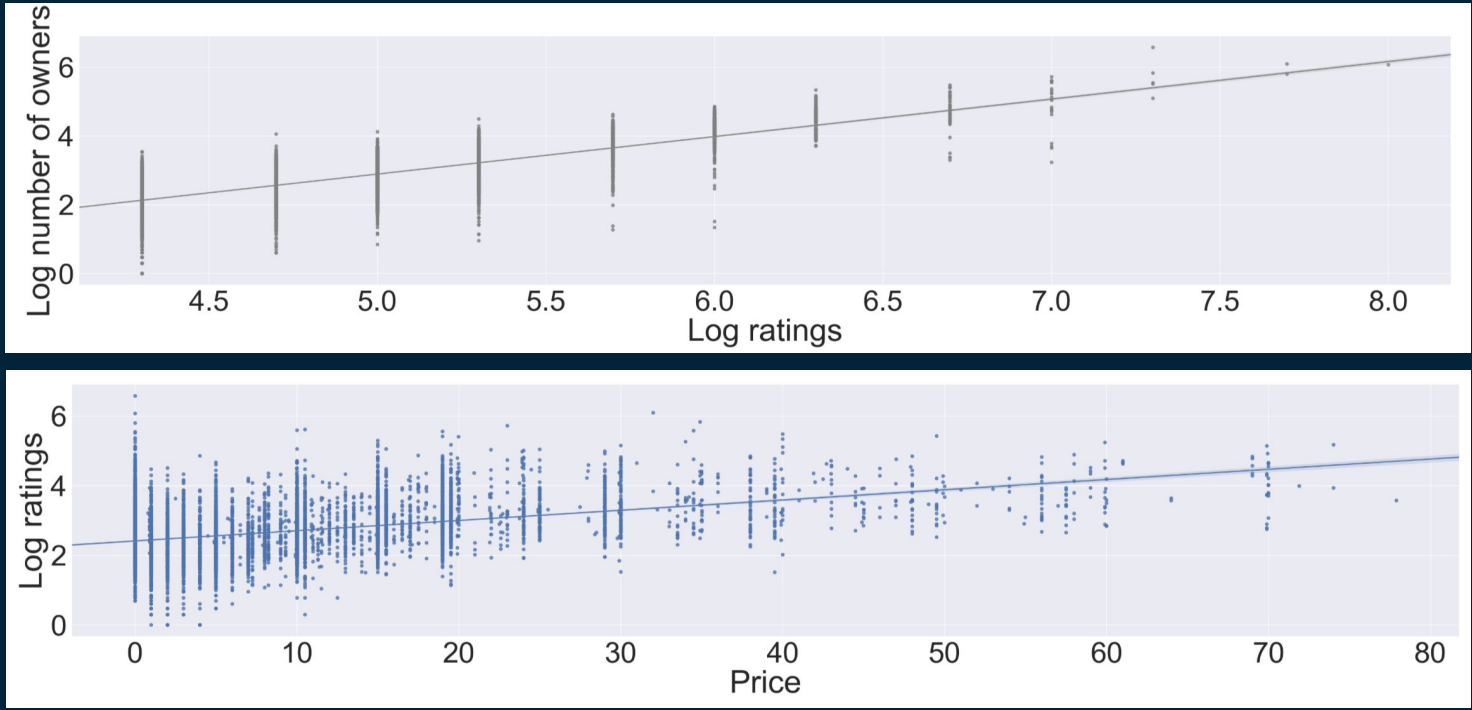
$$EconomicSuccess = Number_{Owners} \times Price_{Game}$$

- Rating (SteamDb)

$$ReviewScore = \frac{PositiveReviews}{TotalReviews}$$

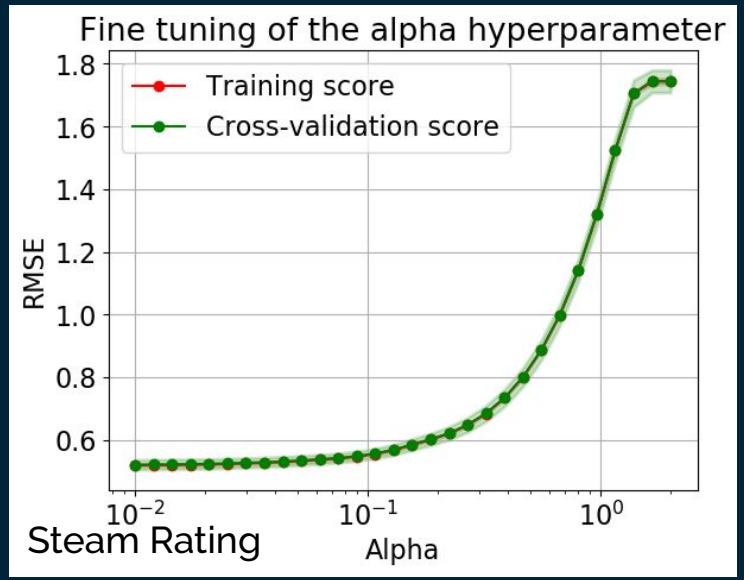
$$Rating = ReviewScore - (ReviewScore - 0.5) \times 2^{-log_{10}(TotalReviews+1)}$$

What influences the ratings



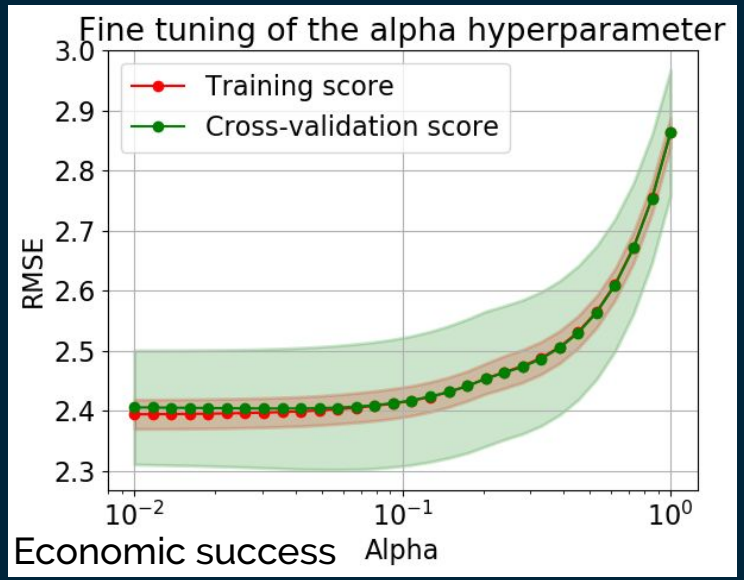
Exploration

Which features are the most important for success ?



Steam Rating

alpha range = [-2;0.5]
Optimal alpha = 0.0329

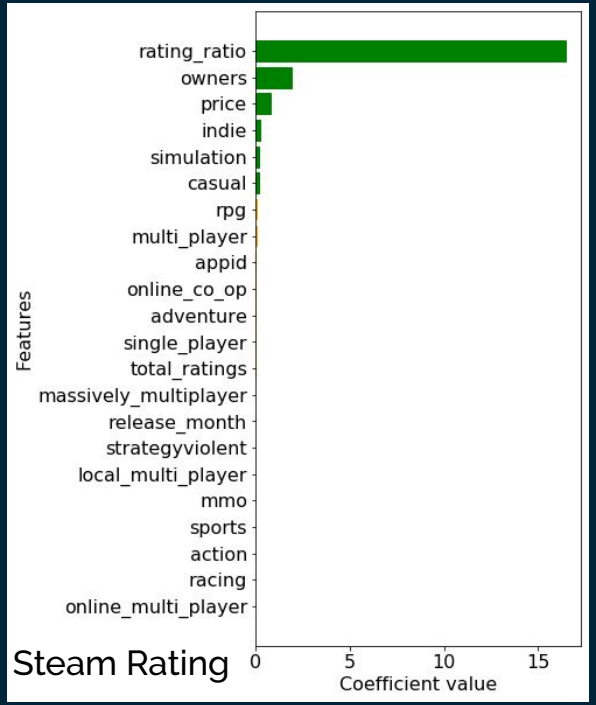


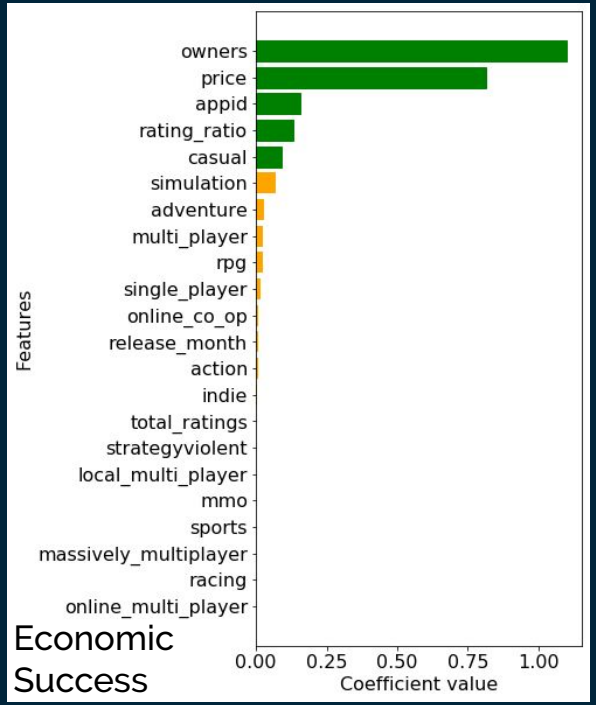
Economic success

alpha range = [-2;0.5]
Optimal alpha = 0.01

Exploration

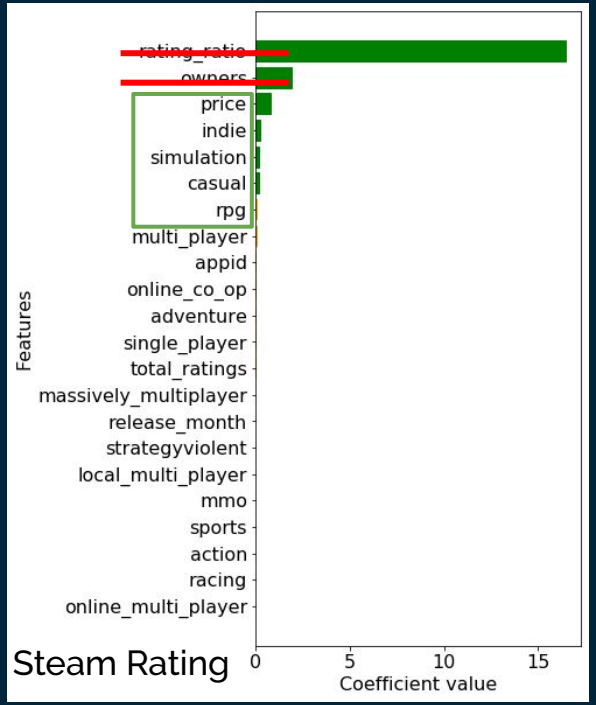
Which features are the most important for success ?

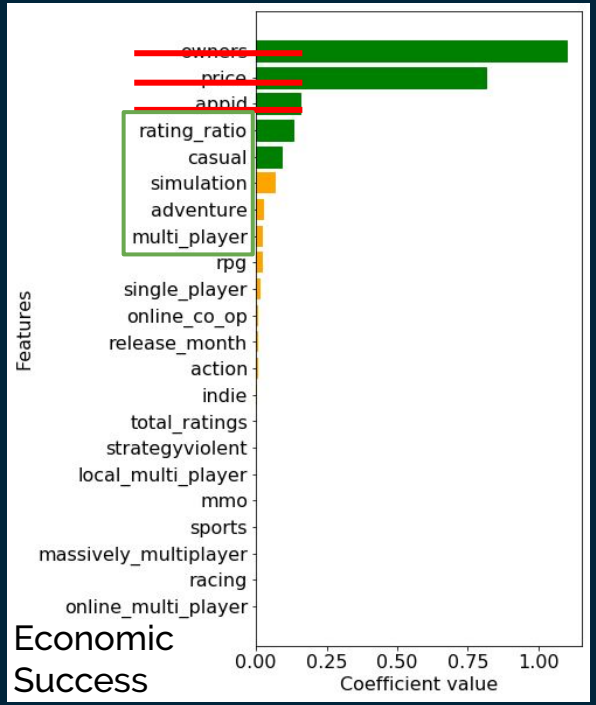




Exploration

Which features are the most important for success ?

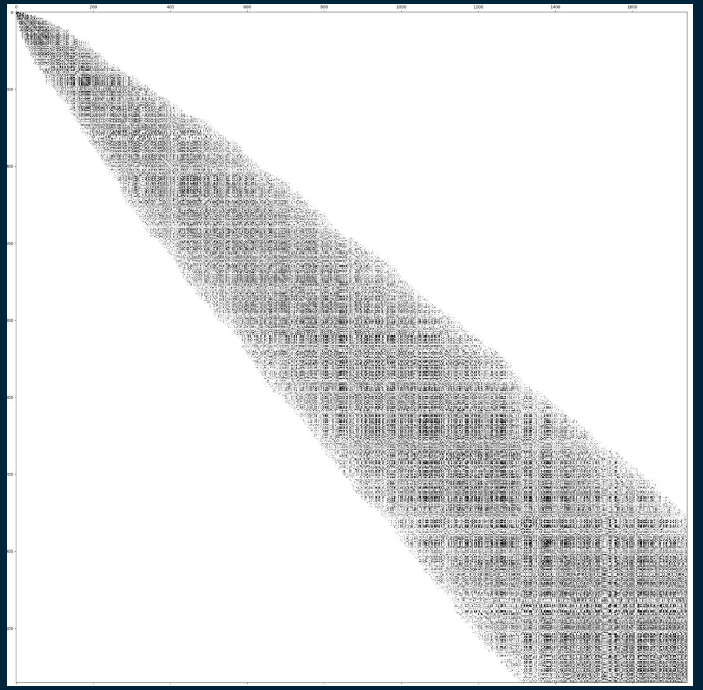




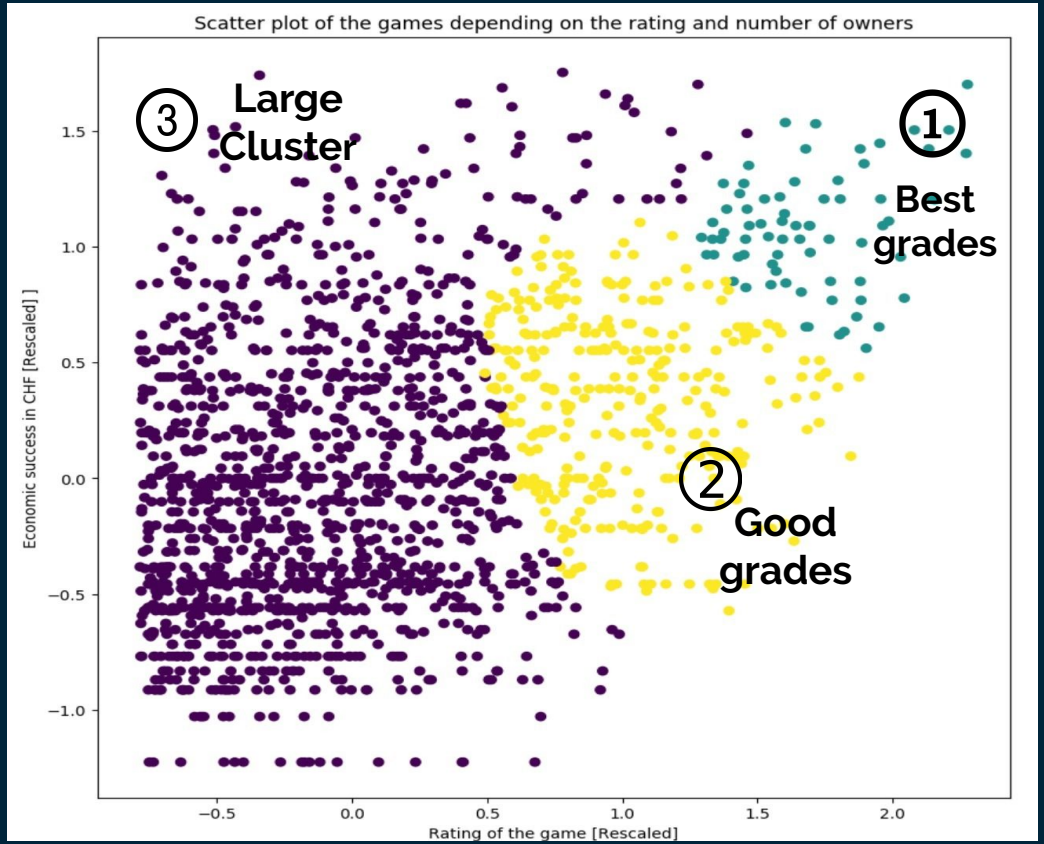
Graph construction

Subset of our data composed of the 2000 best rated games

$$w_{ij} = exp(\frac{- ||x_i - x_j|| _2^2}{2\sigma^2})$$



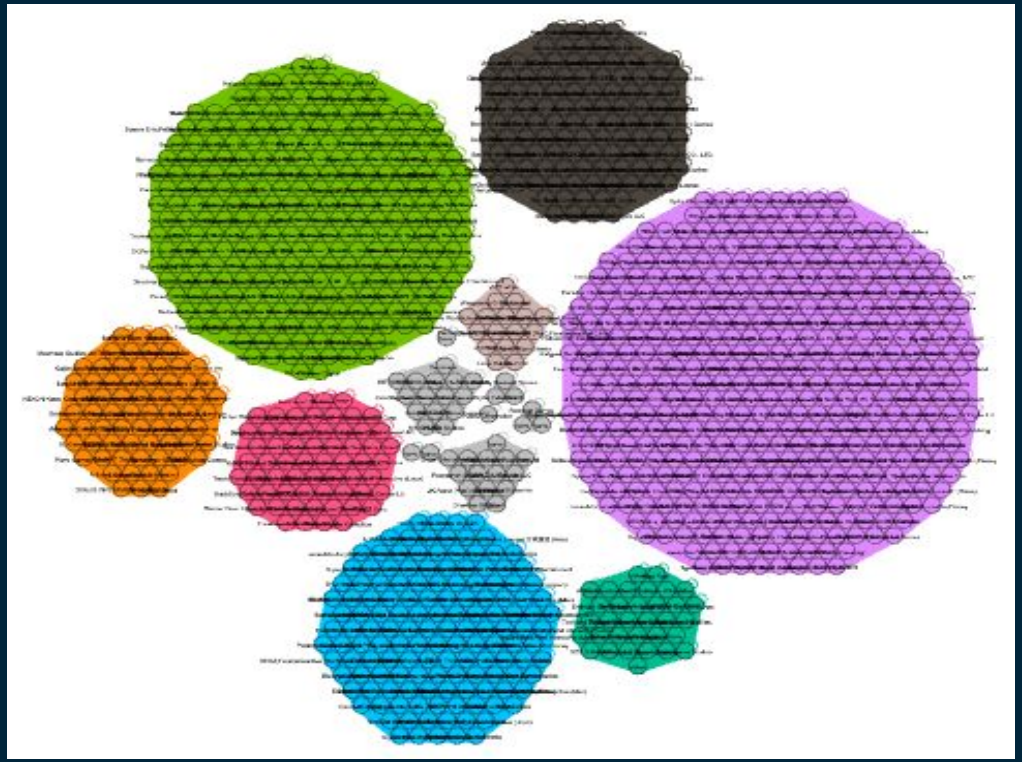
Exploitation



Spectral clustering with Laplacian and K-Means clustering

Our games are divided in three clusters

Gephi implementation



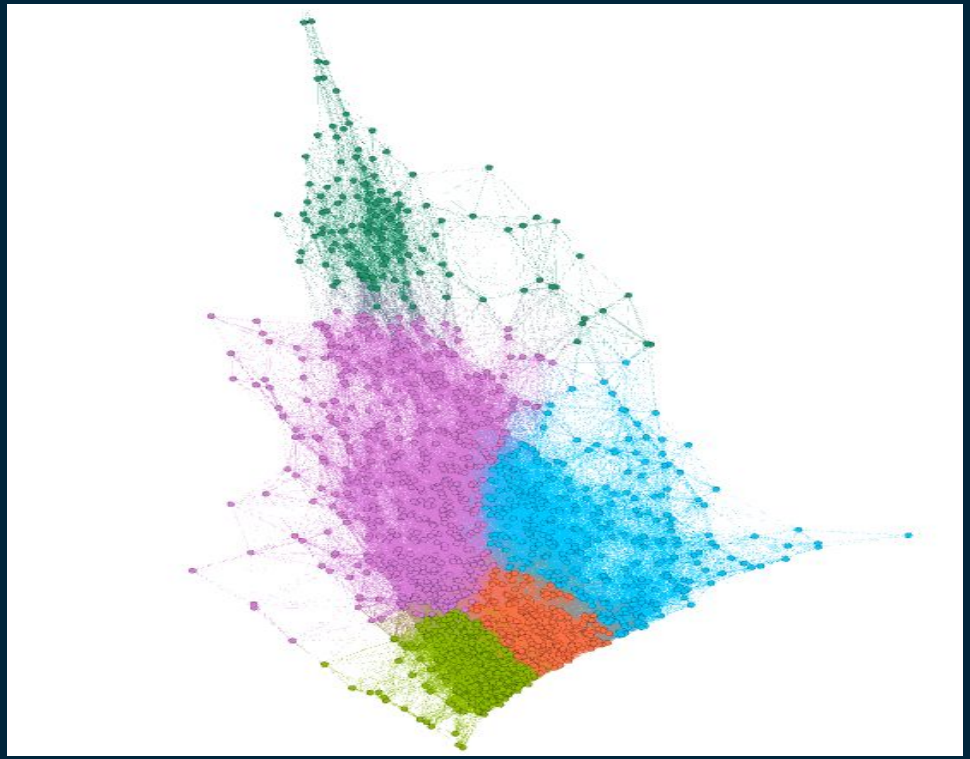
Modularity clusters with Circular Pack layout

Circular Pack Layout

Small World networks:

- Clusters
  → dense ties + short average path length)
- Few degrees of separation

Gephi implementation



Modularity clusters with Yifang Hu layout

Yifang Hu Layout

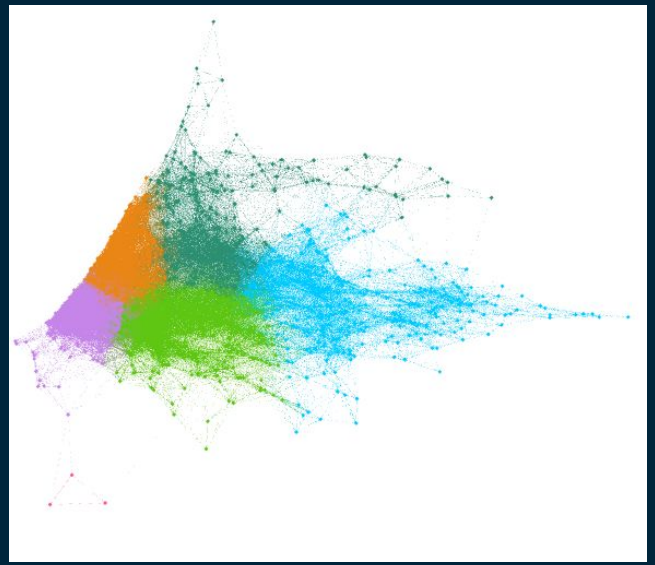
1. A mass of well-rated games
2. More restricted + better rated
3. well rated + lot of owner.

➢ very "**trendy**" games (bulk of traffic)

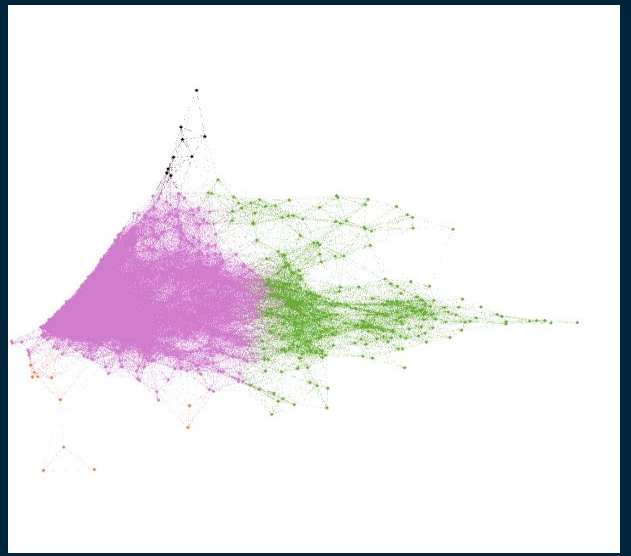
➢ the more "**standard-kind**" games are more on the outskirts.

Gephi implementation

Clusters with Leiden algorithm



Modularity clusters with Yifang Hu layout



Leiden Algorithm clusters with Yifang Hu layout

Conclusion

- Unable to find an absolute given parameter of success
- Economical dependency: number of owners X the overall rating
- Additional parameters for success definition
- Success determined by the gamer's community...
  - ...but handful of players required before