# Network Analysis of the Steam Community

Zac Reid

## Introduction –

The Steam community is a vast network of millions of players, it is a dynamic and rapidly evolving community much like the rest of the 'gaming' space. The Steam community is a part of the Steam platform which is a video gaming platform, providing services to 132 million monthly active users, along with providing a store for players to purchase and download there games, it also provides the Steam Community which is a hub for users to exchange there opinions, in-game items, and original content in the form of game mods. In this report I will focus on four separate networks of friends, giving a combined total of approximately 76k users. We will look to find the dynamics that drive how friendships are created in our networks, compare how the networks found in this report look to other networks, along with trying to explain how cheaters are distributed about our networks. (Steamworks - Valve, 2022)

## Data and Methods –

To gather the data, I used the Steam Web API, this is a standardised way of communicating with Steam servers to retrieve player data. To build the networks of players, you simply get a player from the Steam database (you can do this at random by generating random steam ids), then start by getting all the friends of this player, we can then repeat this process for all the friends of this player, and we can begin to build a network of players. To do this we get all the players with public friends lists and begin to connect them on our graph.

Below are some summary statistics of the networks gathered. These networks where created and analysed in NetworkX and igraph.

(See full network summaries Fig 4.x)

Network One – (Fig 1.1)

```
16308 nodes and 110685 edges
VAC Bans: 332, Comm. Bans: 28
Avg. Degree: 13.57
Avg. Clustering: 0.369
Friend Distribution (Fig 2.1)
Cheater network (Fig 3.1)
```

Network Two – (Fig 1.2)

```
9430 nodes and 19409 edges
VAC Bans: 1371, Comm. Bans: 109
Avg. Degree: 4.12
Avg. Clustering: 0.35
Friend Distribution (Fig 2.2)
Cheater network (Fig 3.2)
```

Network Three – (Fig 1.3)

```
13976 nodes and 31503 edges
VAC Bans: 342, Comm. Bans: 213
Avg. Degree: 4.51
Avg. Clustering: 0.295
Friend Distribution (Fig 2.3)
Cheater network (Fig 3.3)
```

Network Four – (Fig 1.4)

```
3433 nodes and 6089 edges
VAC Bans: 485, Comm. Bans: 34
Avg. Degree: 3.547
Avg. Clustering: 0.315
Friend Distribution (Fig 2.4)
Cheater network (Fig 3.4)
```

**Results –**

From our networks, we get a lot of long tailed distributions, for every attribute where users can either earn, collect, or perform, we get trends that follow the power law. This is similar to a lot of real-world social networks, where a lot have few and a few have all. Examples of these distributions can be seen in Fig 2.x and Fig 5.x.

We also get networks with that form sub-clusters around large main clusters, this can be seen in Fig 1.x, this can be thought of as networks with multiple friend groups combined by members linking with multiple groups, this is similar to how real-world social networks work.

Another interesting feature of these networks becomes apparent when separating cheaters (users with one or more VAC bans on there account) and their friends from the network and creating there own, then performing a comparison against each other. The results and figures of this can be seen in Fig 3.x and 4.x. The result of which are networks with higher clustering coefficients versus the default networks. Along with networks with higher or similar degrees versus the default network.

This gets even more interesting when you look at networks of just cheaters, (this is achieved by creating networks with only users with one or more VAC bans on there account). We find these networks are badly connected with low or no clustering coefficients and low average degrees.

When looking to see if specific attributes can explain what causes some users to have more friends then others, you can see the main correlation that can be found is one between a users level (this is a number that represents a users XP, this is earned via completing badges which are earned by competing in community challenges, collecting cards from games, or buying these cards from the steam community market place) and a users friend count. See Fig 6.x.

**Discussion –**

If we look at the at the attributes in our data, we see exceptionally large, long tailed distributions, this is quite similar to what we see in almost every other social environment, for example in the economy, followers on social media, and every competition domain. This trend is seen in user levels and achievements to games owned and hours played. Though some of these could be explained through external factors from steam, for example levels are achieved through the creation of badges which are either earned or bought, if we see income distributed unevenly in the real world it would follow attributes such as level would also follow this trend. The factor of a players income could also affect a users playtime and level, this can work in both ways, if a user has a low income, this could be an indication of a player being unemployed meaning they could spend more time in game whereas a user with a high income would have less time to spend time in game thus spending more to make up for this deficit.

Another reason for the friend counts following a power law may be due to the fact it is easier to make new friends if you already have more friends, thus leading to a feedback loop causing some players to have disproportionally more friends then others. Players could also have more friends due to notoriety through forms and groups, along with having a high steam level, as players with a high level appear first on other friends lists, thus again causing a feedback affect.

Cheating is an interesting feature of our network, not only because of how cheaters and their friends seem to cluster closer together, but of the questions it arises, both about how these connections between cheaters and normal players form, but if these connections can have an influence on a person’s propensity to cheat. A cheater in this report is defined as a player with one or more VAC bans on record, for a player to get this steam must have detected reported software running on there computers to have received this ban. This itself brings some issues as the cheaters that have been caught in our network our 'obvious' cheaters, it is not uncommon for players to create individualised cheating software or to purchase tailor-made software from other vendors to cheat which would obstruct steam anti-cheat from detecting these cheaters.

As to why we see such a contrast between the networks with cheaters and friends versus just cheaters there could be a few reasons but we cannot say for certain, one reasons could be as mentioned above some of these users labelled as non cheaters may in fact actually be cheaters, thus affecting our analysis, although this is only theory as there is no way to test this. Another reason could be due to a widespread practice known as 'smurfing;' this is when a user creates an alternative account with the sole purpose of cheating in a game. The reason we can observe the results we see in our networks could be due to players adding there smurf's as friends and their friends also adding these accounts causing a higher cohesion between the nodes. Finally, this could also be explained by like-minded people adding each other, if a player is more likely to add a cheater, this could explain why we observe what we observe.

Finally if we see a correlation between a users level and a users friends count this should not be much of a surprise from what we have discussed above, if we think of the inputs of ones level being playtime and money, and the result of which being prestige and a higher level, it would be a natural conclusion that this would result in a user getting more friends, not only from the increase playtime, resulting in increased opportunity to create new friends, but with the increased prestige and the increasing in-gaming spending resulting in more games and more in game items, these players can be though of as having a higher social status versus the average player in the steam community.

**Potential Extensions –**

There are plenty of ways these observations could be taken further with this report. One of which is the inclusion of a player's game collection, this would require both the analysis and to get access to this data from steam as they seem to block these kinds of requests when an excess amount is sent making this data collection impossible. Another possibility would be to investigate how a player's membership in separate groups affects their attributes and friend network.

Another possible avenue for improvement would be to look into the factors of cheating more deeply. This could involve attempting to classify cheaters from non cheaters with machine learning to attempt to derive the driving attributes, along with collecting data on the games these cheaters play and see if connections or conclusions could be drawn from these. This could also be expanded to attempt and identify accounts owned by that are owned by the same user and are used for cheating, which is the practice known as 'smurfing,' as this could be used to make cleaner predictions about how cheaters propagate throughout the network. This could be done by inferring a match through similar details on a profile, or by getting access to steams records and looking at attached mobile numbers or emails.

Another more obvious way to improve this report would be to increase the sample size, this would both be time consuming and require more compute power, since the steam API will only support up to one hundred thousand requests in a calendar day, collecting full player profiles (which takes multiple requests) can take many days, and when attempting to create ever expanding groups of friends the amount of profiles you would need to get a exponentially growing amount of profiles the further you go out.

**Conclusion –**

In conclusion the networks analysed in this report are ones that closely follow the kinds of network seen in social networks outside the steam community such as social media, the placement of these networks on the web allows for both the network they represent and the individual members scale to reach a mass of people. The distribution of degrees in the steam network and how the nodes connect with each other are not only for the experiences of players but can give a clue how trends and opinions can propagate through a network, an understanding how trends spread could help both platform developers and game developers understand how to best cater for player experience. The way players and cheaters are distributed throughout the network could help build better anti-cheat software. This could help build a community free from cheaters leading to a better experience for all.

The way in which the network is constructed also raises social questions, such as do networks with distributions like this led to both issues on an individual level such as mental health issues and or on a societal level, as we are essentially observing what is the entire population of the United Kingdom play daily. You may also ask is it an ideal scenario for some players to be on the extreme end of achievement as this could indicate possible addiction, if we look at how achievements and thus 'status' is distributed throughout the network and how this corresponds with an increase in friend count, it raises that with all this positive feedback, are we essentially creating these social structures to feed addiction.

**References –**

Steamworks - Valve. (2022). Steam - 2021 Year in Review. *Steamworks Development Events*: Steam. Retrieved from https://store.steampowered.com/news/group/4145017/view/3133946090937137590

**Figures –**

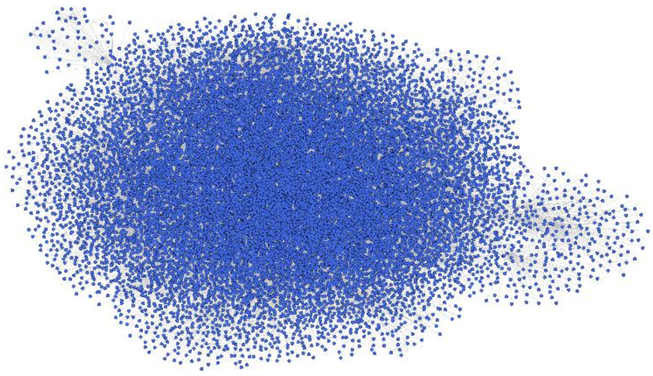

Fig 1.1

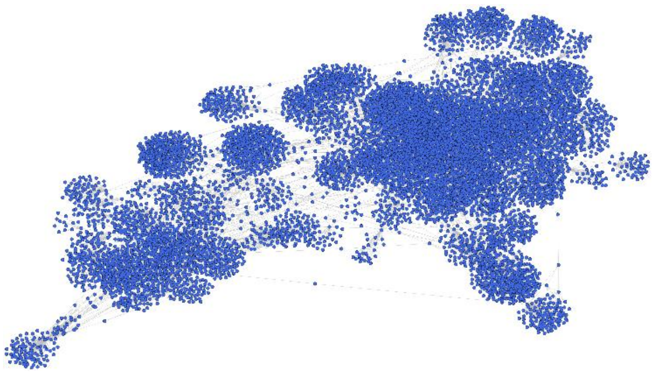

Fig 1.2

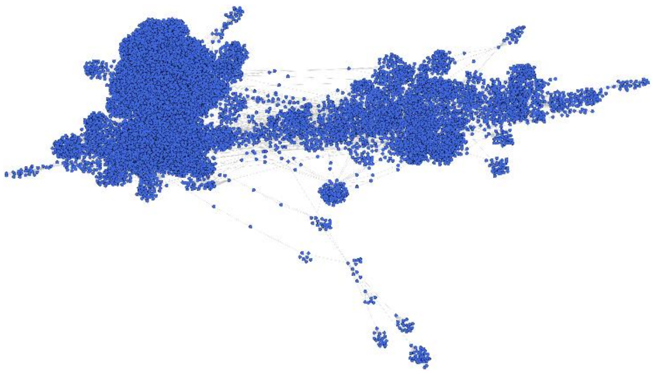

Fig 1.3

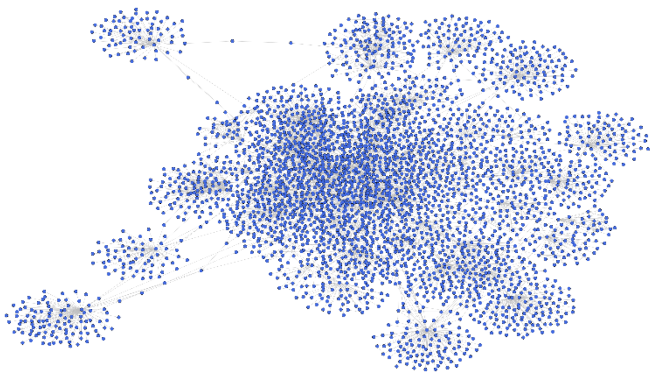

Fig 1.4

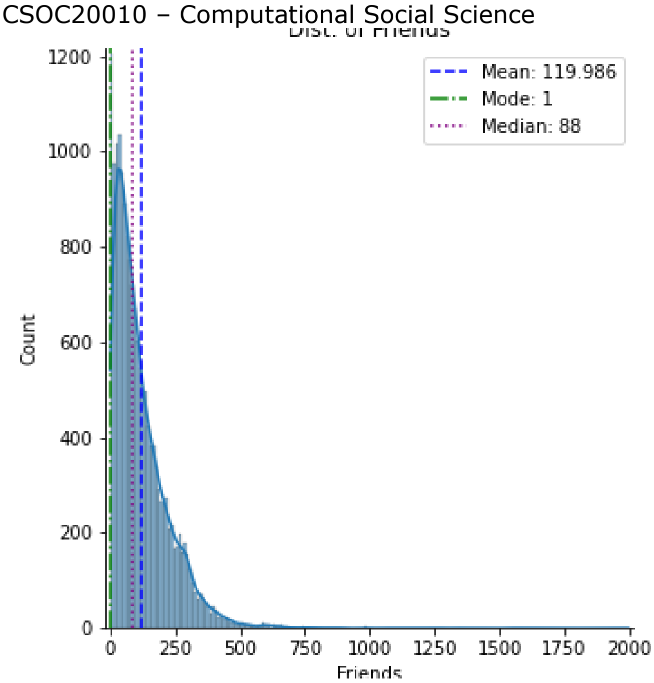


Fig 2.1

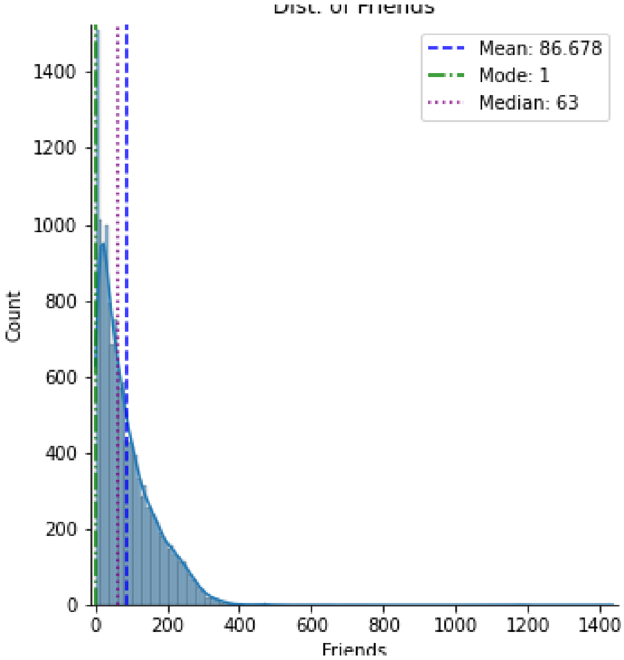


Fig 2.3

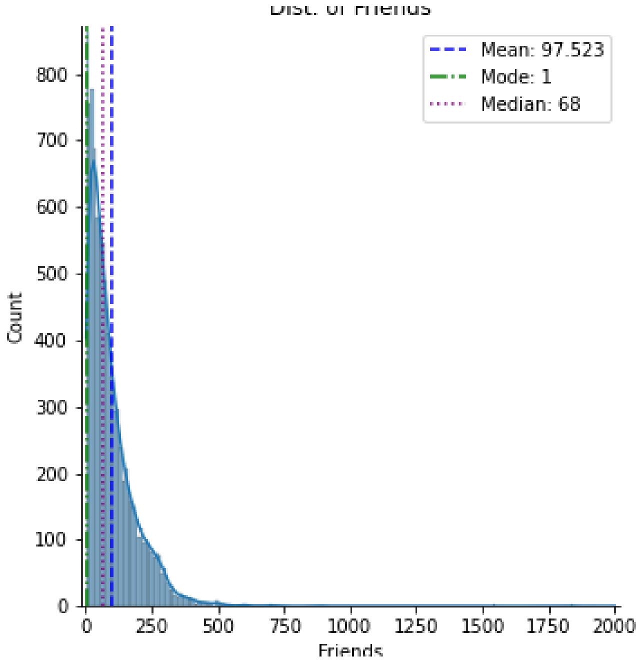


Fig 2.2

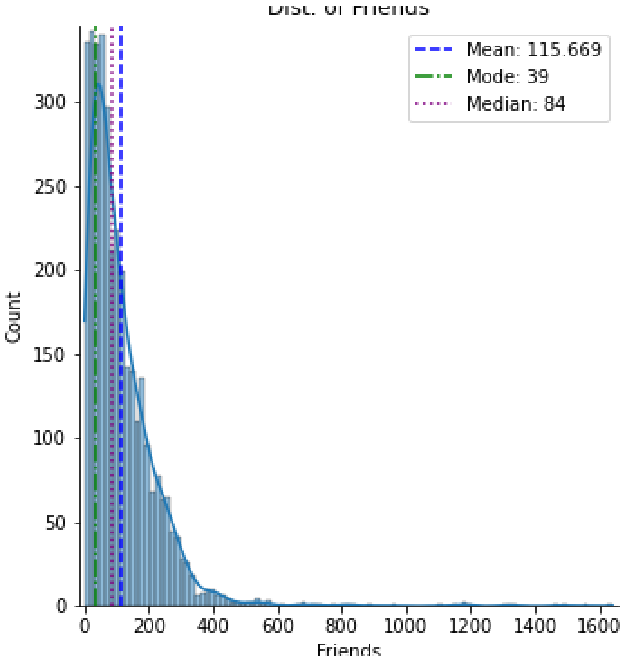


Fig 2.4

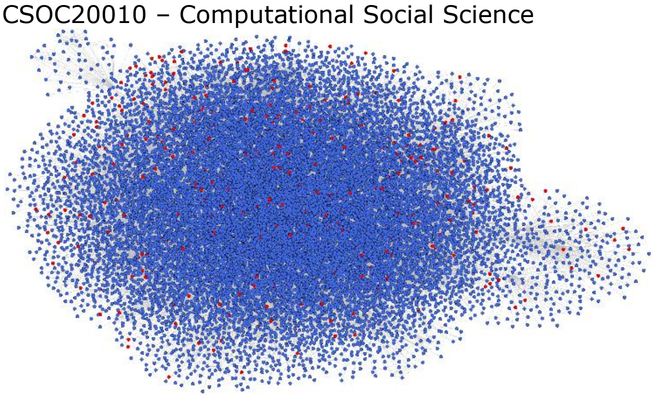

Fig 3.1

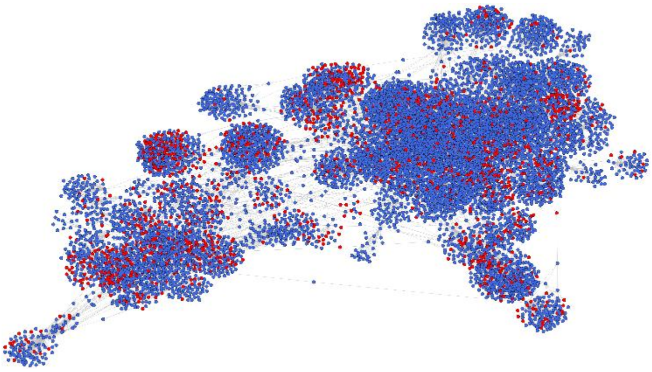

Fig 3.2

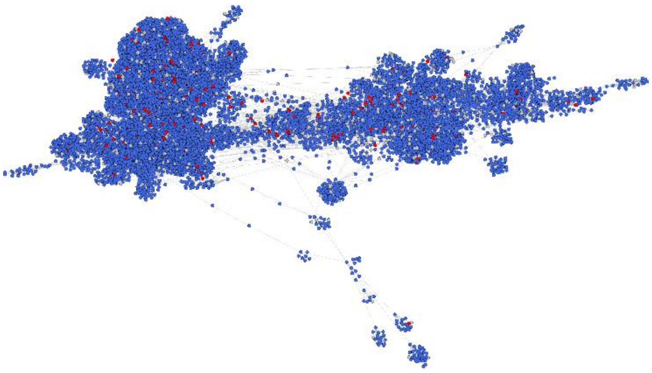

Fig 3.3

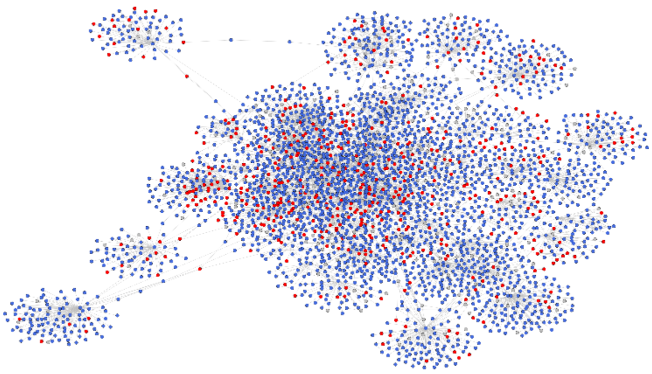

Fig 3.4

Fig 4.1 –

Default Graph -- --

Created a graph with 16308 nodes and 110685 edges!

VAC Bans: 332, Comm. Bans: 28, Econ. Bans 0

Avg. Degree: 13.574319352465048

Avg. Clustering: 0.3693663307623806

Cheater Graph -- --

Created a cheater graph with 2132 nodes and 15464 edges!

Avg. Degree: 14.50656660412758

Avg. Clustering: 0.403322602610922

Just Cheater Graph -- --

Created a cheater graph with 332 nodes and 27 edges!

Avg. Degree: 0.16265060240963855

Avg. Clustering: 0.0

Fig 4.2 –

Default Graph -- --

Created a graph with 9430 nodes and 19409 edges!

VAC Bans: 1371, Comm. Bans: 109, Econ. Bans 0

Avg. Degree: 4.11643690349947

Avg. Clustering: 0.3501047265061107

Cheater Graph -- --

Created a cheater graph with 3014 nodes and 7821 edges!

Avg. Degree: 5.189781021897811

Avg. Clustering: 0.4494924427203749

Just Cheater Graph -- --

Created a cheater graph with 1371 nodes and 542 edges!

Avg. Degree: 0.7906637490882568

Avg. Clustering: 0.07513652042834101

Fig 4.3 -

Default Graph -- --

Created a graph with 13976 nodes and 31503 edges!

VAC Bans: 342, Comm. Bans: 213, Econ. Bans 0

Avg. Degree: 4.508156840297653

Avg. Clustering: 0.29474381783032794

Cheater Graph -- --

Created a cheater graph with 1332 nodes and 3001 edges!

Avg. Degree: 4.506006006006006

Avg. Clustering: 0.3586834977729632

Just Cheater Graph -- --

Created a cheater graph with 342 nodes and 31 edges!

Avg. Degree: 0.18128654970760233

Avg. Clustering: 0.006822612085769979

Fig 4.4 –

Default Graph -- --

Created a graph with 3433 nodes and 6089 edges!

VAC Bans: 485, Comm. Bans: 34, Econ. Bans 0

Avg. Degree: 3.5473346926886107

Avg. Clustering: 0.3152502145793823

Cheater Graph -- --

Created a cheater graph with 1222 nodes and 2621 edges!

Avg. Degree: 4.289689034369886

Avg. Clustering: 0.40627127079219805

Just Cheater Graph -- --

Created a cheater graph with 485 nodes and 217 edges!

Avg. Degree: 0.8948453608247423

Avg. Clustering: 0.05871667817392499

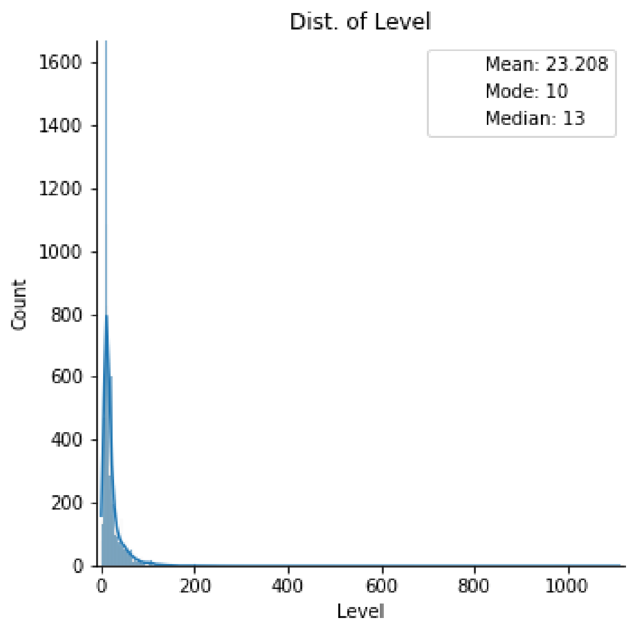


Fig 5.1

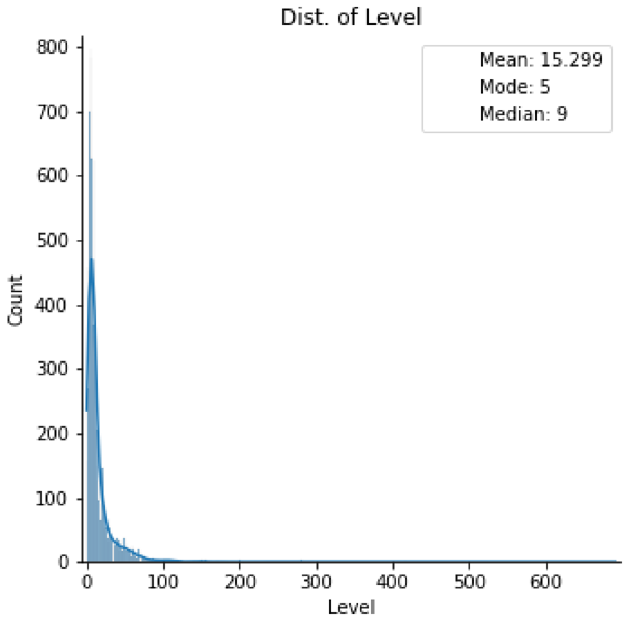


Fig 5.2

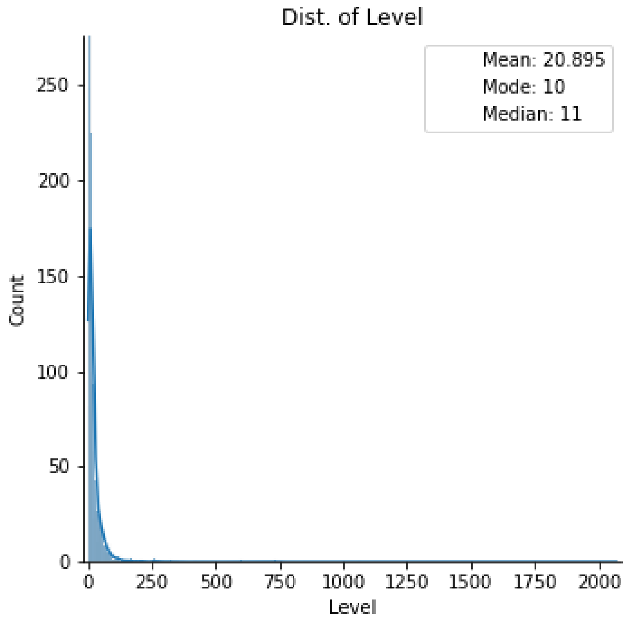


Fig 5.4

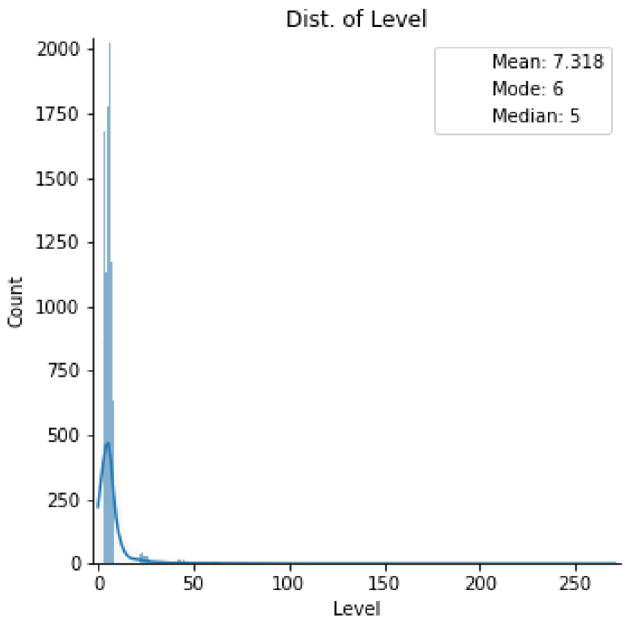


Fig 5.3

Fig 6.1 – COR = 0.48

Fig 6.2 – COR = 0.6

Fig 6.3 – COR = 0.45

Fig 6.4 – COR = 0.55