Twitube : A comprehensive influence analysis tool for TwitchTV and YouTube

ChangHyun Shin* Chan Park
Department of Electrical and Computer Engineering
Seoul National University
Seoul, Korea
schyun9212@snu.ac.kr, cpark@snu.ac.kr

ABSTRACT

Recently, Gaming Video Contents (GVC) market is growing rapidly.. According to [1], by 2017, the GVC market is worth $ 32 trillion, of which the advertising market accounts for 51%. This means that the influence of content providers in the GVC market has a great impact on society. In terms of advertisers, marketing, and agency, it is important to analyze the influence of content providers for effective investment..

The most representative platforms in the GVC market are TwitchTV and YouTube. Each platform basically provides analysis tools that can analyze the impact of users. However, recent GVC content providers generally do not use a single platform, but use multiple platforms at the same time, so traditional analysis tools alone can not capture the overall influence.

To solve these problems, we propose a visualization tool, Twitube, which helps visualize the data of similar characteristics between TwitchTV and Youtube to understand the comprehensive influence of content providers.

Keywords: TwitchTV, YouTube, influence analysis, data visualization, rank algorithm

1 INTRODUCTION

The global market for GVCs is about 660 million consumers worldwide by 2017, which is larger than the sum of consumers of streaming services for other content such as movies and sports. TwitchTV [2] and YouTube [3] are the platforms that dominate the GVC market, and their number of viewers is higher than the sum of all viewers on other legacy platforms [1].

Recently, the GVC market is growing rapidly. According to [1], as of 2017, it has a size of $ 32 trillion, of which the advertising market accounts for 51%. This means that the impact of GVC providers on society is enormous. In fact, it can be seen that much of the society is heavily influenced by the GVC provider when the words of the GVC content provider become buzzwords, ads, and various events and accidents..

As the influence of GVC providers has increased, advertisers and marketing planners need to grasp their influence in order to make efficient investments.

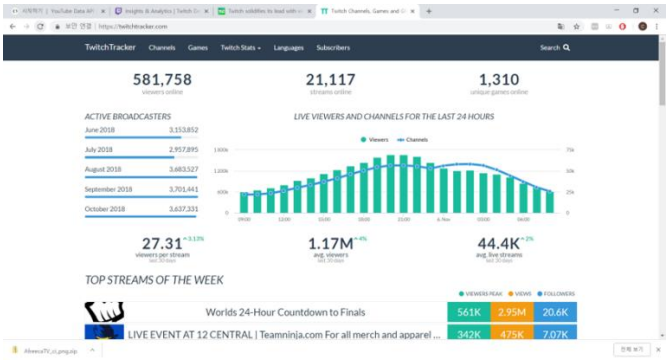

Figure 1: TwitchTV analysis tool.

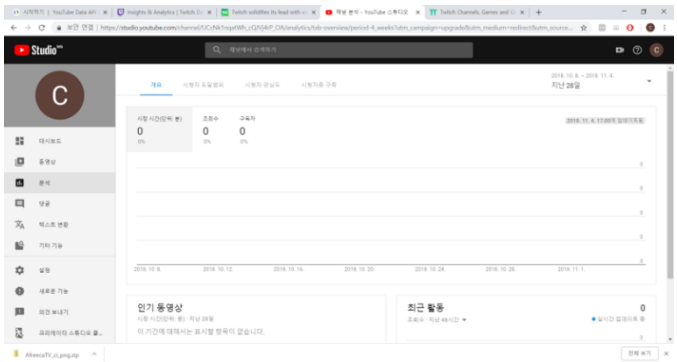

Figure 2: YouTube analysis tool

If the visual influence of the GVC provider is expressed, the advertiser or sponsor will be able to effectively invest in their situation.

TwitchTV and YouTube provide basic tools for analyzing user statistics. TwitchTV, shown in Figure 1, can perform various analyzes in terms of game, number of viewers, and real-time broadcasting. In Figure 2, YouTube offers a variety of analyzes in terms of views and subscribers. Using these analysis tools, you can fully understand the user's influence on each platform. However, recent GVC providers often use multiple platforms at the same time, not just on one platform. For example, after playing a game on TwitchTV, you edit your own broadcast and upload it to YouTube as a video. GVC providers are not likely to have equal influence across platforms because of the cultural differences

TwitchTV	YouTube
Average Viewer	Average View Count
Followers	Subscribers
SRA	CRA

Table 1 Similar features between TwitchTV and YouTube

between platforms. Therefore, in order to quantify the influence of content providers on various platforms, it is necessary to analyze the statistics of each platform in a complex way.

In order to analyze the statistics of the platform in a complex way, we classified data with similar characteristics on each platform. Table 1 shows the features used in this project. The average number of viewers of TwitchTV, a real-time game streaming service, and the average number of views of YouTube, a video streaming service, are similar in that they represent the number of users who support content provided by a GVC provider. TwitchTV's followers and YouTube's subscribers, who record their favorite GVC providers, have similar characteristics in that they represent the number of users who support GVC providers. Because of the nature of the game, many GVC providers are influenced by each other or playing games together, so I thought it would be meaningful to analyze the scores considering GVC providers' mutual relations. As the Google search engine categorizes a large number of pages on the web through the Page Rank Algorithm (PRA) [4], the influence of GVC content providers is evaluated using algorithms such as SRA and CRA, which transform PRA using tags and follow information.

Twitube analyzes the data collected by TwitchTV and YouTube on the basis of similar features, and visualizes them by digitizing them. This allows us to take complex data from multiple platforms into account so that we can see the exact influence that one platform analysis tool alone could not have..

2 OVERVIEW

Figure 3 shows the overview of Twitube. First, we periodically crawl the data on each platform. Next, analyze the data with similar characteristics of both platforms before delivering it to the client. Finally, the client receives this data and performs the visualization.

2.1 Data Crawling

Data collection was done using APIs provided by TwitchTV and YouTube [5] [6]. TwitchTV's data, which is the main content of real-time streaming, was crawled every 30 minutes, while YouTube's data handling static content was crawled every 6 hours

In TwitchTV, there are real-time user information for grasping the number of users who prefer the content of the GVC provider, real-time streaming information for

knowing the favorite game of the streamer, information of the number of followers for knowing the number of users who prefer the streamer, And follow list

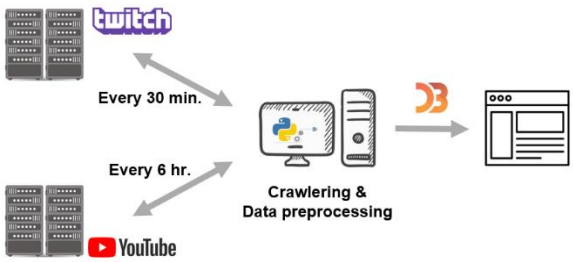

Figure 3: Twitube Overview

information for searching for the link among the streamers.

In YouTube, crawler collects views information to determine the user's preference for the video of the YouTuber, the number of subscribers to know the number of users who prefer the YouTuber, the title and tag information for the relationship between the YouTubers .

2.2 Data processing

To optimize the program, all data is processed as much as possible and input to the client. Since the processed data will be finally output in the form of a network, it is stored in the form of a node and a link, and the daily statistic value for the designated period is additionally stored.

TwitchTV nodes store the average number of viewers in the past week, broadcast time, played games and play time, followers, follow list the streamer follows, and SRA scores. Figure 4 shows the example of node of TwitchTV.

YouTube's nodes store the number of image views, subscribers, and CRA (pra is the name before naming) scores recently. Figure 5 shows the example of node of YouTube.

Each of the data was normalized to compute the overall score of the two platforms.

$$n(d) = \sqrt{\frac{d}{d_{max}}} \; (d \in D) \qquad (1)$$

Equation *(1)* is the normalization expression used for processing D denotes the entire dataset, $n(d)$ denotes normalized data, d denotes data before normalization, and d_{max} denotes the largest value among D.

Figure 6 shows the link data for TwitchTV and YouTube. Stores a starting and ending point and a score indicating the strength of the connection. Calculating the score will be explained in Section 3.

```
{
    "id": "김블루",
    "alias": "김블루",
    "average_viewer": {
        "viewer": 1043.0666666666666,
        "normalized_viewer": 0.3202326713676539,
        "duration": 15.0
    },
    "games": [
        {
            "game": "Just Chatting",
            "duration": 4.0
        },
        {
            "game": "PLAYERUNKNOWN'S BATTLEGROUNDS",
            "duration": 11.0
        }
    ],
    "followers": 201788,
    "normalized_followers": 0.7309514786965309,
    "sra_score": 18.633640914222408,
    "normalized_sra_score": 0.1748700845104892
},
```


Figure 4: Processed TwitchTV node data


```
{
    "id": "김블루",
    "alias": "김블루",
    "recent_average_view": 449016.0,
    "subscriber_count": 1277072,
    "pra_score": 50204.50807110926,
    "pra_out_score": 72620.5,
    "normalized_pra_score": 0.3317158698672232,
    "normalized_subscriber_count": 1.0,
    "normalized_average_view": 1.0
},
```


Figure 5: Processed YouTube node data


```
"links": [
    {
        "source": "햇살살",
        "target": "중괄호",
        "score": 2.876152347450885,
        "normalized_score": 0.12334825539748855
    },
```


Figure 6: Processed TwitchTV and YouTube link data

2.3 Visualization

Figure 7 shows the visualization results using the processed data.

In the first section, the influence of the GVC provider is expressed through the network type. The size of the circle is determined by the values that follow the checked criteria in the two control panels and represents the magnitude of the influence. The color of the circle represents the game that the GVC provider most frequently plays. The arrows around the circles represent associations between different GVC providers. Through the color of the arrow, we can know the association by yellow (in), green (out), red (mutual), etc., and the thickness of the arrow indicates the intensity. The strength of the link is determined by the score of CRA and SRA.

The second section is the control panel where you can manipulate the results of section 1. In the control panel, you can interactively control the parameters included in the ranking algorithm and filter out specific games to be excluded from the analysis.

The third section is a section that prints a summary of the node when clicking on a node in the network section 1. Ranking information and scores for each aspect, and game play statistics for the GVC provider for a week.

The fourth section is a summary of the results of applying the filter in section 2. It prints more detailed information in text than Section 3. Rankings and scores for each aspect, comparative analysis of impacts among similar GVC providers, recommendation of the most influential GVC provider among similar GVC providers, and recommendation the type of advertising by analyzing potential.

3 Scores

In order to analyze the overall impact of TwitchTV and YouTube, we calculated scores from several aspects. Since there is a large difference in the data range between the two platforms, the influence was calculated by normalizing the processing steps and adding the respective values.

$$radius(sc) = \alpha\big(YouTubeValue(sc)\big) + (1-\alpha)\big(TwitchValue(sc)\big) \quad (2)$$

Equation 2 shows the equation for calculating the influence of TwitchTV and YouTube. sc indicates selected criteria (average view / average viewer, subscriber / follower, CRA / SRA). α is a parameter that can control the ratio of influence between TwitchTV and YouTube. Adjustable in section 2 of Figure 7. The closer to 1, the greater the weight of YouTube. The closer to 0, the greater the weight of TwitchTV. All metrics except for the Merged Rank Algorithm (MRA), which combines CRA and SRA, are all analyzed using this formula.

The connection between GVC providers is a very important factor that can generate synergy between them. However, for smaller links below certain values, it may be meaningless data. To remove this data, we added the ability to adjust the dropout value, which is the criterion for discarding links with low connection strength between nodes. This number serves as a threshold. Adjustable in section 2 of Figure 7.

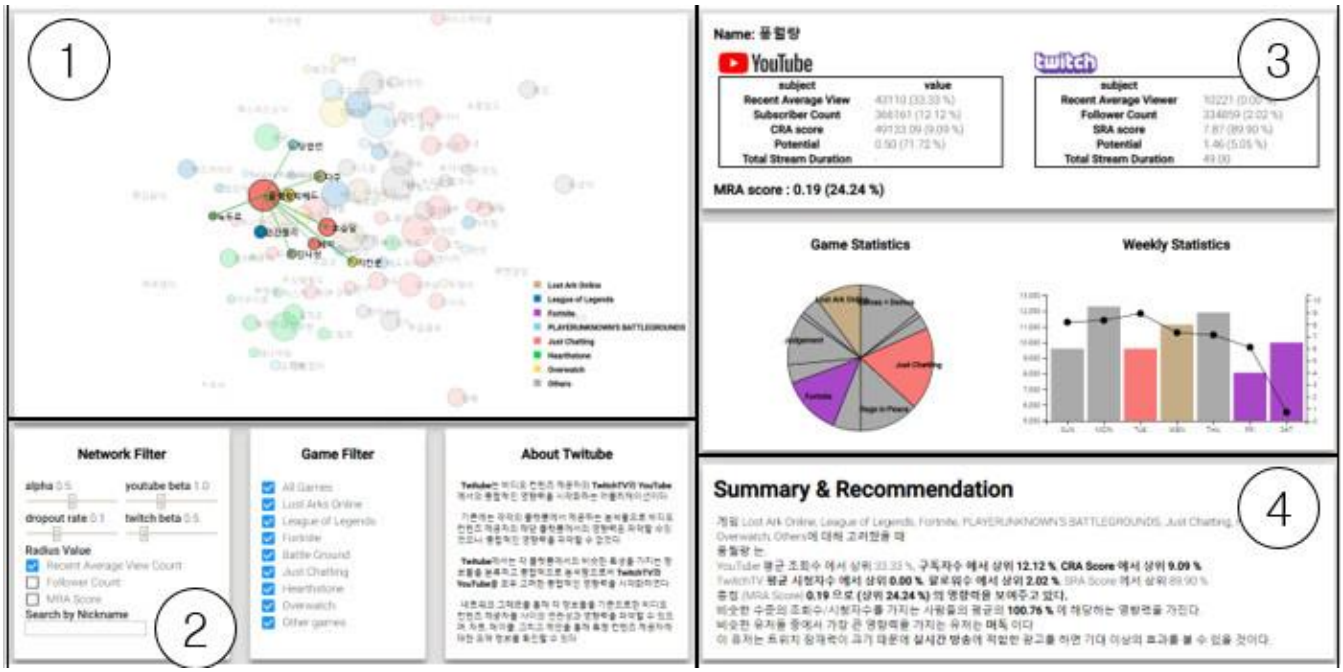

Figure 7: Twitube interface

3.1 Channel Rank Algorithm(CRA)

CRA is a newly defined metric that is inspired by Google's PRA and is used to rank among YouTube channels instead of pages. We used the information tagged in the video to quantify the influence between the channels. The equation for obtaining CRA is as follows.

$$CRA(u) = \sum_{v \in V_u} \frac{n_{u,v}}{N_v} \times \text{view}(v) \qquad (3)$$

u is the channel of some GVC provider. V_u represents videos of channels other than u channel, and v represents one of videos included in V_u. N_v means the total number of tags written in the movie v, and $n_{u,v}$ means the number of tags including u in the video v. view (v) means the number of views of the video v. In summary, this equation can be said to be the sum of the weightings of the number of views of images of other GVC providers by a GVC provider. For example, if a user A uploads a video, creates 10 tags, and includes 5 users in it, 50% of the video views count as the influence of the B user. The sum of CRAs for one YouTuber is used as the total influence, and one CRA is used as the score for the link between two different YouTubers.

3.2 Streaming Rank Algorithm(SRA)

SRA is a metric that uses TwitchTV's follower information to determine the influence of connections between streamers. The formula of SRA is as follows.

$$SRA(u) = \sum_{v \in S_u} \frac{average\_viewer(v)}{\text{average\_viewer}(u)} \qquad (4)$$

u means streamer. S_u means streamer except u and v means one of streamers included in S_u . average_viewer(v) is the average number of viewers of streamer v. Because TwitchTV is a platform mainly for game streaming, there are many cases where streamers play games together or host their viewer to other streamer. The SRA is a numerical value of the influence that may appear in such a relationship. For example, if a user A has an average of 100 viewers and a user B has 1000 viewers, if they follow each other, the impact on each other will be 10 A and B 0.1. This means that already famous streamers are less influenced by unfamiliar streamers, but unfamiliar streamers are heavily influenced.

3.3 Merged Rank Algorithm(MRA)

The MRA is a metric that calculates the combined influence of both platforms, by adding the normalized CRA and SRA. MRA has the following formula.

$$\text{MRA}(u) = \alpha \times \{\beta_{youtube} \times \text{CRA}(u)\} + (1 - \alpha) \times \{\beta_{twitch} \times SRA(u)\} \qquad (5)$$

The MRA is the addition of a β parameter that can set a specific weight for each platform in the equation in section 3.1. In general, the influence of the link on TwitchTV and the link on YouTube may be different, so we added a parameter that can be controlled in detail.

The tags on YouTube are significant because they are created by the GVC provider and are typically added when the GVC providers go through the content.

The follow on TwitchTV is relatively minor because it is interested but cannot guarantee the content to run together.

The MRA calculated by applying each rank algorithm will be able to more accurately determine the influence of the GVC provider when both platforms are considered together.

3.4 Potential

We have newly created a data called potential score.

Potential score refers to the potential of the GVC provider on each platform. The potential on YouTube means how many average views are coming from subscribers, and the potential on Twitch is how many average viewers are coming in relative to followers.

4 INTERFACE

4.1 Network Graph

The network graph in section 1 of Figure 7 shows the magnitude of the influence among the GVC providers, their relationship to each other, and the game they play most often. Each circle represents the GVC provider, the size of the circle means the size of the influence, and the circle color means a game that is often played. Pressing the circle highlights other circles associated with the GVC provider corresponding to that circle, and is linked by an arrow. The direction and color of the arrows indicate the direction of influence and the size of the arrows indicate the link strength between the GVC providers.

The visualized graphs can be viewed from different sides by the control panel, and filtering can be used to check the results among people playing a particular game.

In addition to emphasizing on the network when clicking on the circle, the right side shows the numerical values, ranking, statistics, and analysis results according to each analysis element.

4.2 Control Panel

The control panel in section 2 of Figure 7 serves to determine what the network graph should visualize or to adjust the parameters of the points mentioned in section 3. When hover mouse near the label, it prints a simple tooltip that describes its parameters. The check box can be used to perform filtering based on the game information to check the analysis results for a specific game. It is responsible for making it adjustable. We implemented the function to emphasize the GVC provider by inputting the ID of the GVC provider considering that it is difficult to find by too many nodes.

4.3 Basic Information Panel

The basic information panel in section 3 in Figure 7 visualizes the overall statistics and scores of the GVC providers corresponding to the circles clicked on the network graph. It can be checked the numerical value and rank of each element for each platform, the game play time and the ratio of the game for a week, game statistics for each day, and so on.

4.4 Summary and Recommendation Panel

The summary and recommendation panel in section 4 of Figure 7 outputs the analysis result of the result filtered or controlled by the control panel as a sentence.

First, we show the scores and rankings for each element, and if it is determined that the number is high, the number is emphasized. Next, we show how much influence the GVC providers with similar numerical scores have on the aspects checked in the control panel. Given that similar scores would require a similar amount of investment, the most influential GVC provider among similar GVC providers would also be output to recommend the most efficient user. Finally, we analyze the potential of the user, and if it is high, print a sentence that suggests which ad should be good.

5 CONCLUTION

In this project, we used Twitube to visualize the influence of a comprehensive analysis of data between each platform that was not available with existing analysis tools. By analyzing data with similar characteristics or by applying a rank algorithm, we can check the results in various aspects by calculating new scores.

We added tooltips to labels and charts to enhance user convenience. We've also been able to manipulate various parameters or filter games to test more complex situations. By manipulating these parameters, user can get the results that fit for purpose.

REFERENCES

[1] https://www.superdataresearch.com/market-data/gaming-video-content/

[2] https://www.twitch.tv/.

[3] https://www.youtube.com/

[4] Page, Lawrence, et al. The PageRank citation ranking: Bringing order to the web. Stanford InfoLab, 1999.

[5] https://dev.twitch.tv/docs/v5/

[6] https://developers.google.com/youtube/v3/getting-started?hl=ko