The Future of Livestreaming: Strategy and Predictive Analysis for Future Operations of Facebook Live

Nolan Alexander
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
nka5we@virginia.edu

David Brenman
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
db3za@virginia.edu

John Eshirow
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
eshirow@virginia.edu

Joshua Rosenblatt
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
jmr6vb@virginia.ed

Justin Wolter
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
jsw6ch@virginia.edu

William Scherer
Department of Engineering Systems and Environment
University of Virginia
Charlottesville, USA
wts@virginia.edu

James Valeiras
Data Scientist
Facebook, Inc.
New York City, USA
james.valeiras@gmail.com

***Abstract*—Social Network Services (SNS) are systems that allow users to build social relations with one another, with one of the largest SNSs being Facebook, totaling 2.6 billion active monthly users in 2020. Their live streaming service, Facebook Live, is one of the fastest-growing branches of the company, allowing creators to synchronously broadcast original content to the public. However, in the rapidly growing world of technology, Facebook Live faces fierce competition from other live streaming platforms (Twitch, YouTube Live, etc.) and well as other video-on-demand providers (Netflix, Hulu, TikTok, etc.). To better understand current issues and future directions, our team focused on the Facebook Live platform, to develop a three to five-year strategic plan for the platform. We focus on Facebook Live's growth opportunities from a multitude of perspectives, including the competitive landscape, interface modification, future projections based on historical trends, and competitive analysis. Our approach utilizes the systems analysis process, focusing top-down on objectives and metrics. To produce a comprehensive strategy for future operations, we employ analytical methods ranging from quantitative data analysis to qualitative exploration of industry trends. These quantitative methods include statistical analysis, time-series forecasting, and natural language processing. Qualitative methods include domain research into the history, current state, and possible future for live-streaming. Forthcoming, the results for the complete analysis will be synthesized into a multi-recommendation strategic report to provide Facebook with flexible guidance for continuing operations. We also present a comments summarization and visualization feature for viewers and creators, three to five-year market forecasts after COVID-19 lockdowns, and attractive emerging markets including education and morning shows.**

The team would like to thank Facebook, Inc. for sponsoring this project.

Keywords— Livestreaming, Social Media, Market Analysis, Competitive Analysis, Natural Language Processing, Latent Dirichlet Analysis, Sentiment Analysis

I. Introduction

Live-streaming is a novel technology that allows users to watch and interact with video in real-time over the internet. Facebook has capitalized on this market through its successful launch in 2016 of the Facebook Live service. Multiple other competitors have also seized the opportunity, and in the future, not all will be able to withstand the competition. To perform our research, our team at the University of Virginia was provided access to Facebook personnel who work daily with the Facebook Live platform. Our research efforts engaged with Facebook Live personnel to better understand the system and present recommendations for future growth so that Facebook can remain a top competitor in the market.

Developing these requirements required a strong understanding of the Facebook Live ecosystem including its objectives and metrics. Using this understanding, we investigated three primary areas:

Comments summarization and visualization feature – Uses NLP to provide users a view into what topics others are discussing, their emotions on these topics, as well as providing content creators additional information to gauge their viewers.

Forecasting the market – The market forecasts focus on predicting trends after COVID-19 lockdown restrictions.

Competitive analysis and emerging markets –investigates markets that Facebook should also pursue along with their competitors including morning shows. The emerging markets analysis investigates untapped markets to pursue including education and fitness.

The paper will next explore each of these areas.

II. Understanding the System

To develop an understanding of the system, we first identified the objectives and metrics, primarily through interfacing with the client and performing domain research. To improve understanding, we developed system diagrams to visualize interrelationships among system components. Given the recency of the development of livestreaming platforms, we found relatively limited research available on such markets. Furthermore, given the proprietary and internal nature of data relating to the Facebook Live platform, there exists even less research on this specific platform. Our novel research and analysis provide a much-needed foundation to explore trends in the live streaming market, competitive behavior, and improvements to the Live platform, which can be used by Facebook, and companies with similar products, to gain viewers and grow market share.

Regarding objectives and metrics, we identified the main objective, which we determined was split into two primary categories.

- The financial impact of Facebook Live on the bottom line of Facebook as a whole.
- How Facebook Live can help Facebook live up to its mission statement.

Considering the financial aspects, we determined that this objective is to drive overall profitability for Facebook. While Facebook's profitability is dependent on both its revenues and costs, we focused on the performance through a set of viewership and content metrics: a balance of content, demand, and viewership growth. Facebook Live can generate revenue for Facebook in two primary ways: generating revenue from ads within Facebook Live; and driving Live users to pursue other features of Facebook, thus generating ad revenue there.

Additional objectives for Facebook Live focus on living up to Facebook's mission statement: to "*give people the power to build community and bring the world closer together*" [Facebook, 2020]. These objectives focus on how Facebook Live can connect friends and family, help people discover what is happening in the world, and share information they feel is important.

To accurately gauge how our systems analysis can work toward achieving these objectives, we identified quantitative metrics to measure the achievement of the identified objectives. Top-line view time is the first potential metric, which would measure the total view time across all viewers across all streams. This metric is the most direct measurement of how much content is being consumed on Facebook Live. A potential concern of using this metric is that it does not capture how well Facebook Live is adding value to other parts of Facebook, as it only measures the users using Facebook Live. Similarly, the total number of viewers across all streams is another top-line metric that gives a broad overview of Facebook Live's performance but does not describe Live's impact on the rest of the platform. The average number of users or peak daily users better helps to capture the proportion of Facebook users that are using Live. This is a helpful metric to see if individual live events are drawing users to Live as well as to compare Facebook Live to other live streaming services. A potential concern with using this metric is that it does not account for how engaged the users are since it does not measure the length of time that the user is watching content for. Another metric is the change in users over time to measure user growth, which can be used to compare Facebook Live to other services and see which competitors are growing to capture market share. These metrics each have their flaws, so no single metric can be used to measure system success. Therefore, all of these metrics, combined with qualitative analysis, must be taken into consideration when evaluating our recommendations.

We focused our research effort on three primary analyses: Comment summarization and visualization, forecasting the market by extrapolating insights to post-COVID, and competitive analysis and emerging markets.

III. Comments Summarization and Visualization

Comments provide a medium for users to express ideas, and for other users to be exposed to these ideas. However, in large livestreams, the volume of comments can become overwhelming as it moves too quickly for users to process. A significant number of Facebook Live users choose to hide the comments section for this reason. Facebook Live attempts to rectify this concern by limiting the number of comments to show only those the specific user is likely to interact with [1]. Other platforms use random sampling or limit the frequency that users can comment. While this reduces the number of comments shown, it discards a significant amount, which can lead to users only receiving exposure to a select set of opinions and ideas.

We propose an alternative approach: providing a visualization of comments using natural language processing (NLP). Specifically, this approach will find the relevant topics that users are commenting on, find the users' sentiment on these topics, and find the emotions that users feel on these topics. To handle the rapidly changing topics, sentiments, and emotions from livestreaming comments, we will employ a rolling window. The model uses two parameters: the lookback period of the rolling window, and the frequency of updating the visualizations. The parameters could be tuned to find optimal values by performing AB tests on various livestreams with this model implemented. This feature can also be used to provide sentiment information to content creators, which they can use to improve their future streams.

Using this feature, users will be able to both view a summary of comment topics with their corresponding emotions, as well as filter their comment feed by these topics. By being able to gauge the sentiment of their fellow viewers at a glance, users will receive an improved live stream experience as well as a greater sense of community within the stream. By filtering the comment stream into relevant topics, commenters

can form smaller, more tightly knit sub-communities within the comment section, which will also lead to a greater sense of community within the Facebook ecosystem. In the long run, we expect that this improved comment feature, via bolstering the users' live streaming experience and fostering a greater sense of community with their peers, will lead to increases in Facebook Live's top-line metrics such as the total number of viewers and overall watch time.

A. Extracting Comments Information with NLP

Comments data must be preprocessed before it can be used in modeling. In NLP, each comment is referred to as a document, and the collection of comments is known as a corpus [2]. Preprocessing initially removes irrelevant words such as "the", which are known as stop-words. It then performs lemmatization, which transforms words to their stem; for example, "watching" would be transformed to "watch". Then, frequently occurring sets of words are grouped into bigram and trigrams, where a bigram is a set of two words joined together and a trigram is three. Finally, any words with no inherent meaning on their own, such as "because" are filtered out.

The process of finding topics that users are discussing is known as topic modeling, which we performed using Latent Dirichlet Allocation (LDA) [3]. LDA is an unsupervised approach that uses stochastic Bayesian modeling to find a set number of topics and the distribution of these topics in the corpus. LDA is a bag-of-words model, meaning it only uses the frequencies of words in the corpus, rather than ordering or punctuation. LDA assumes that topic distributions follow a Dirichlet distribution because it is a conjugate prior, so its posterior distribution is the same as its prior. LDA initially assigns each word in the corpus to a topic and then iteratively moves these words to another topic according to the conditional probabilities of being in each document, such that it will converge to optimal posterior distributions.

Sentiment analysis is a method used to find users' views on a negative to positive scale from a document [4]. Sentiment analysis generates a polarity score from -1 to 1 that measures the document's sentiment. Sentiment analysis uses pre-trained neural networks to determine the polarity score.
Emotional analysis is similar to sentiment analysis, but instead measures users' views on a multidimensional scale for each emotion rather than only two [5]. Emotional analysis provides the proportions of emotions in a document. The package we used measured emotion proportions for the following emotions: fear, amusement, anger, annoyance, indifference, joy, awe, and sadness. Emotional analysis similarly uses pre-trained neural networks to determine proportions.

We implemented this model using a publicly available comments dataset for the NASA mars rover landing and using Python packages including SpaCy and Genism [6]. Because LDA is an unsupervised approach and can only extract weights of words in a topic, we chose to label the topic according to the highest weighted words. Table 1 shows the topics, sentiment, and standard deviation of the sentiment from the first 10 minutes of the stream, and table 2 shows it for the last 10 minutes of the stream.

Table 1. The topics sentiment and standard deviation of sentiment

Topic Sentiment for the First 10 mins		
Topic	**Sentiment**	**SD**
history	0.079	0.241
love	0.119	0.244
i was here	0.068	0.203
mars	0.087	0.255

Table 2. The topics sentiment and standard deviation of sentiment.

Topic Sentiment for the Last 10 mins		
Topic	**Sentiment**	**SD**
congratulations	0.068	0.267
nasa	0.078	0.263
i was here	0.101	0.252
rick astley	0.117	0.282

The LDA found different topics over time, as table 1 shows users commenting about "history" and "love" before the rover landed, while table 2 shows users commenting about "congratulations" and "Rick Astley" (a well-known internet meme) after it landed. The sentiment on the topics were similar, so it did not provide useful information from this stream, but could be useful for streams where users are expressing more extreme feelings. Figure 1 provides two visualizations of the overall sentiment over time.

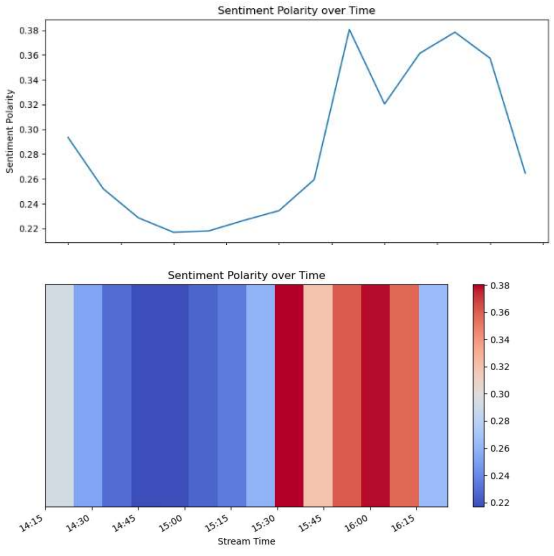

Figure 1. The sentiment over time represented both as a time-series and heatmap.

We propose that these visualizations could be shown to a content creator (e.g., a user generating livestream data) as part of their dashboard. The content creator would have the option to view the sentiment changes with respect to their stream as well as in comparison to the overall sentiment from -1 to 1. Based on similar streams, the y-axis would also include baseline sentiment and standard deviations, so that the content creator can determine how far from the standard the sentiment has deviated.

While the sentiment only provides a negative to positive view, emotions provide a richer view. Table 3 provides the emotion proportions for the first 10 minutes of the stream.

Table 3. The emotional proportions of the first 10 minutes of the stream.

Topic Emotions for the First 10 Mins				
Emotion	**T1**	**T2**	**T3**	**T4**
Fear	0.09	0.08	0.08	0.09
Amusement	0.18	0.17	0.17	0.18
Anger	0.1	0.09	0.1	0.09
Annoyance	0.12	0.12	0.13	0.12
Indifference	0.14	0.15	0.14	0.14
Joy	0.11	0.11	0.11	0.11
Awe	0.18	0.17	0.18	0.18
Sadness	0.09	0.1	0.09	0.09

For all the topics, the comments primarily showed that the users were in awe and amusement. In addition to being able to extract the topics, sentiments, and emotions, we determined the need for an interface to display this information.

B. Interface Design

As discussed earlier, for streams with a large number of viewers and commenters, comments enter the feed in real-time, and due to the volume of comments entering the feed concurrently, it is nearly impossible to gauge the sentiment of the audience. This is a problem for both viewers and content creators. Viewers, when unable to interact with the comment section, may become discouraged and disconnected from the rest of the audience, which therefore makes them less likely to view streams. This issue impacts both of Facebook Live's top-line metrics of view time and viewer count, as well as going against their value proposition of building communities. Publishers can read the comment stream to gauge the viewer's feelings and adjust their content accordingly, but they will have difficulty if there are a large number of comments. Our solution looks to provide both viewers and publishers an option to view a real-time summary of the comment section that shows which topics are being mentioned the most, and the sentiments surrounding each.

A visual prototype of this comment summary on Facebook Live's mobile interface is shown in Figure 2.

Figure 2. Comment Summary Graphic Prototype.

The live stream depicted in the figure is from a news source discussion on President Biden's COVID-19 recent stimulus plan. As seen in the figure, four primary topics discussed in the comment section are shown: "Stimulus", "Biden", "COVID-19", and "Vaccine". Each topic also has a corresponding emoticon, which depicts the general audience sentiment surrounding each topic. The sentiment is pulled from the emotional analysis described previously and is mapped to a corresponding emoticon. Additionally, the area of each topic "bubble" corresponds to the total number of comments found within each topic. This graphic allows viewers to quickly and easily gauge the popular topics being discussed, the frequency of these topics, and how users feel about these topics. Users can click or tap one of the topic bubbles to see a new comment section filtered to only contain comments on that topic. This allows users to tailor their comment feed and connect with other users interested in the same topic. With this comment summary option, we believe that viewers will be more engaged with the comment section, making them more likely to leave their comments and revisit streams. Additionally, we believe that streamers will be able to better estimate the sentiment of their viewers, and in turn, provide viewers with more beneficial content. All of these outcomes together will increase Live's top-line metrics and a greater commitment to their missions of building communities and bringing the world closer together.

IV. Forecasting The Market: Extrapolating to Post-COVID-19

A key challenge in analyzing the metric data is to account for the impact that the COVID-19 pandemic has on the habits of platform users. Internet usage levels spiked due to countrywide lockdowns, driving significant increases in Facebook Live's top-line metrics, watch time, and view count. To understand the effects of the pandemic and countries'

responses on Facebook Live watch time, we performed an analysis of qualitative country features to explain the statistically significant trends in the platform's driving metrics.

The country-specific analysis focused on underlying characteristics, and their administration's response to the coronavirus. Each country is given a designator in 2 categories of preparation: pandemic and digital economy. Countries with a positive designator in the pandemic demonstrate a higher capability and readiness to handle the virus, while the other countries lack robust channels to deal with such a crisis [7]. For the preparation of a digital economy, countries were graded based on their development of digital infrastructure and ability to shift commerce online during the pandemic. The qualitative country factors assist in understanding the underlying differences between countries before the onset of the pandemic.

To measure the statistical significance of country responses to the coronavirus, a set of countries was chosen to explore the effects of lockdowns on the usage metrics. Lockdown data is assigned to an ordinal scale, with values representing the severity of the lockdown restrictions and guidelines. The severity of lockdowns is mapped to dates for each country, correlating to the countries' level of response to the pandemic and Facebook Live's watch time metrics. To measure the significance of lockdown status and watch time trends, an ANOVA GLM model is fitted to explore the effects of lockdown status on watch time and the interactions between terms.

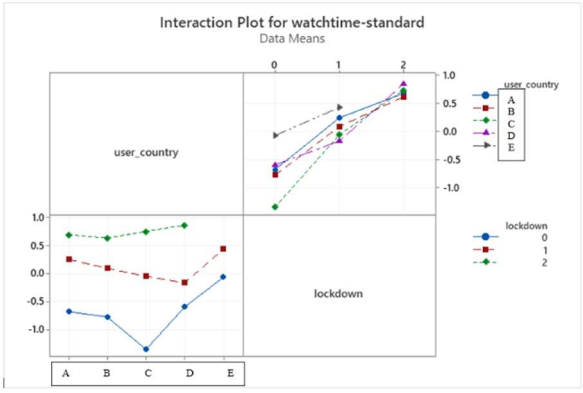

Figure 3: Interaction Model of Watch time and Lockdown Type

Results from the model in Figure 3 show significant increases in watch time in countries with strict lockdowns. Positive coefficients for severe lockdown levels (2), indicate that watch time rises when people are confined to their homes. On the other side, negative coefficients for times without lockdowns (0) demonstrate a downward trend in user activity and watch time. This helps confirm the hypothesis that country-implemented lockdowns force a substitution of physical/in-person activity for digital alternatives, leading to elevated Facebook Live watch time.

Understanding world events similar to the pandemic offers opportunities for Facebook Live to understand and accommodate the substitution for digital activity. A crisis similar to COVID-19 may occur in the future that confines people to their homes. As the prevalence of digital economies grows, the demand for digital communication and interaction increases [8]. Facebook maintains the tools and ability to provide productive channels for online user interactions. The metric shifts observed over 2020 and 2021 demonstrate a clear market for future developments in interactive media and digital services.

V. Competitive Analysis and Emerging Markets

To provide Facebook with recommendations for investments, we performed an analysis on competitors and emerging markets. The competitive analysis provided insight into which markets Facebook should pursue along with its competitors. The emerging markets analysis provides insight into which markets will grow in the future.

A. Competitive Analysis

As the popularity of online, video-on-demand platforms such as Netflix, Hulu, and Disney+ have grown immensely in recent years, producers that create content that traditionally appeared on cable television have looked to move to faster-growing online platforms [9]. We predict that this type of content provides Facebook Live with an opportunity to grow its viewer base. We performed research into content that drives viewership on other platforms or in similar formats to Facebook live to better understand the competitive ecosystem. This research helped to explore and discover markets Facebook Live could expand into. The main driver for viewers to prefer certain platforms over others is content [10]. From this, we determined that there are two primary avenues to deliver the right content to viewers: (1) acquire popular content from other platforms, or (2) create and invest in new, emerging content. In particular, we determined that morning shows, live sports, and live news provide major opportunities for Facebook Live to capture audiences with television-quality content. Investment into general audience content will attempt to draw in users to drive watch time and promote usage of other platform services. Sports content is nearly an assured method to drive watch time but is a particularly expensive avenue in which to expand. Most major sports leagues charge high prices for rights to their content, so the program has to pull a significant number of viewers to achieve profitability.

There also is a steadfast demand for "variety" show content. On television, this takes the form of a morning show or late-night show. These shows offer a consistent source of a variety of news and pop culture and are some of the most-watched programs on television. On competing streaming platforms, there is a significant amount of various content, but it generally consists of influencers playing different games and engaging with their chat in different ways, such as Ask Me Anything (AMA) streams. We believe that there exists a demand for a live-stream variety show that is similar to its television counterparts. We believe that Facebook Live could find potential growth in this market if they invested in a popular host, as the success of this type of content is generally driven by personalities.

B. Emerging Markets

For Facebook to outgrow their primary competitors, they must additionally invest in new, emerging types of content

beyond existing content. The two primary markets that appear to be major opportunities for Facebook Live are education and fitness. To identify these markets, the team performed extensive research on the most popular use cases on both Facebook's and their competitor's platforms. Streamers in the education and fitness space gained significant traction during the global COVID-19 pandemic as individuals were forced to stay home. Both of these categories fit well in a live streaming setting due to the one-to-many information dissemination model. A single educator or fitness instructor can teach a curriculum to hundreds of people at a time. Within education, Facebook Live could produce both formal and informal educational content, such as lectures from professors, professional training, and "Do-It-Yourself" tutorials. A significant use case for education could be public digital learning in developing countries that have a large number of Facebook users. Formal instruction can be modeled after curriculums like *Telesecundarias in Central and South America* [11]. Within fitness, Facebook Live could add fitness classes, yoga instruction, and spin classes. Additionally, fitness influencers with a strong social media presence can broadcast their workouts and host Q&A sessions where viewers ask questions about fitness and health. By investing in these types of new, emerging content to their platform, Facebook could gain an advantage over their competitors, increasing overall watch time and view count.

VI. Conclusion

This research provides a top-down view of the Facebook Live system and three novel proposals for Facebook Live to increase their long-term metrics. Our initial system analysis revealed the primary objectives to be increasing usage and helping build community, with the primary metrics being watch time and view counts. We presented a prototype of a comments summarization and visualization tool that would allow users to see topics that other users are commenting on, as well as the emotion on these topics. Enhanced comment processing and topic recognition seek to encourage a better quality of life for user interactions on the platform. Our analysis of metric shifts due to COVID-19 identified a correlation between active country lockdowns and elevated watch time levels. Recent developments surrounding lockdowns have promoted increased investment in the development of digital economies. Facebook is poised to take a leading role in meeting the demand for online media and features that boost the value of user interactions online. Through competitive analysis, we determined that live-stream morning shows and sports are markets that Facebook and similar competitors should consider further investing in. Our research into emerging markets identified underutilized content genres, investment into which will enable Facebook to add content to its portfolio and increase market share. We conclude that if Facebook focuses attention on education, fitness, and other worthwhile content genres, its platform will grow through diversification and the attained satisfaction of content demand.

While our research has led to some novel results, there exists further research we, or Facebook, can perform to expand upon these areas. Because the LDA model cannot determine a label for the topics, and we chose to use the highest weighted words in each topic as the labels, topic labels may not all be correct. For future research, we propose exploring other topic models, including supervised approaches, and using tools such as Word2Vec. We would also research improving the interface with AB testing. To better understand the implications of future disruptive events similar to COVID, we propose research opportunities to invest in content and advertising when people are faced with the prospect of spending increased time digitally. We also suggest estimating the costs of investing in markets such as morning shows and education and performing tradeoff analysis to determine their viability.

Acknowledgment

The team would like to thank Professor Michael Lenox for his technical expertise and guidance throughout the project.

References

[1] Investor.fb.com. 2021. Facebook - Resources. [online] Available at: <https://investor.fb.com/resources/default.aspx#:~:text=Facebook%20Investor%20Relations%3F-,What%20is%20Facebook's%20mission%20statement%3F,express%20what%20matters%20to%20them> [Accessed 12 April 2021].

[2] B. Srinivasa-Desikan, *Natural language processing, and computational linguistics: a practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Birmingham, UK: Packt Publishing, 2018.

[3] D. M. Blei, *Latent Dirichlet Allocation*, vol. 3. Boston, MA: Journal of Machine Learning Research, 2003.

[4] T. Wilson, *Recognizing contextual polarity in phrase-level sentiment analysis*, vol. 5. Human Language Technology and Empirical Methods in Natural Language Processing.

[5] Y. Zhang, "A Co-Attention Neural Network Model for Emotion Cause Analysis with Emotional Context Awareness," *Empirical Methods in Natural Language Processing*, vol. 2018, pp. 4752–4757, Oct. 2018.

[6] T. Konstantin, "Perseverance Land on Mars YouTube Live Comments," *Kaggle*, 23-Feb-2021. [Online]. Available: https://www.kaggle.com/thomaskonstantin/perseverance-land-on-mars-youtube-live-comments. [Accessed: 12-Apr-2021].

[7] Chakravorti, B., Chaturvedi, R., Filipovic, C., Brewer, G. (2020). *Digital in the Time of Covid*. The Fletcher School at Tufts University. https://sites.tufts.edu/digitalplanet/files/2021/03/digital-intelligence-index.pdf

[8] Spilker, H., Colbjornsen, T. (2020). "The Dimensions of Streaming: Toward a Typology of an Evolving Concept." *Media,Culture & Society.* https://doi.org/10.1177%2F016344372090458

[9] B. Morgan, "What Is The Future Of Television?," *Forbes*, 07-Jul-2019. [Online]. Available: https://www.forbes.com/sites/blakemorgan/2019/07/05/what-is-the-future-of-television/?sh=5e8e1a859de8.

[10] J. Rajala and A. Korhonen, "Streaming wars : competitive dynamics in the online video streaming industry," *JYX Digital Repository*, 2020.

[11] RC Flores and AM Rebollar Albarrán, «La Telesecundaria, before the knowledge society», *RIEOEI*, vol. 44, No. 7, pp. 1-11, Jan 2008.