Subscriber Number is Not Everything: YouTube Community Engagement Measurement

by

Linxin Xie

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Arts in Sociology

Murfreesboro, TN

May 2017

Thesis Committee:

Dr. Foster Amey, Chair

Dr. Meredith Dye

Dr. Angela Mertig

To

Dr. William Craig Carter

I still skate with you

ACKNOWLEDGEMENTS

This work would not have been possible without direction from my academic advisor Dr. Foster Amey. I am especially indebted to Dr. Meredith Dye and Dr. Angela Mertig, who gave me timely feedback and helped me achieve my goal.

I am grateful to these computer scientists, Dr. Ralph Butler, Dr. Hyrum Carroll, Lisa Lowry and Robert Myers who taught me adapting the modern technology in my research. I want especially to thank Dr. Brian Hinote who worked actively to support my choice.

And of course, I would like to thank my lovely husband Lee and my daughter Zoe without whom this project would not have been completed one year earlier.

ABSTRACT

This study developed a method for measuring engagement in YouTube user-generated content (UGC) communities. To date, subscription number has been the most important index to measure YouTube Channels' value; however, it cannot tell the quality of the channel in terms of community engagement. The main goal of this paper was to develop a measurement method to calculate an engagement index of UGC channels. Using this method, I analyzed 31 YouTube UGC channels and their 2.7 million comments. The result is that engagement index does not have positive correlation with the subscriber number. Some channels have community engagement levels not reflected by their subscriber numbers. This analysis method can serve the interests of businesses and agencies looking for marketing opportunities on the UGC YouTube channels.

TABLE OF CONTENTS

LIST OF TABLES vi
LIST OF FIGURES vii
I. INTRODUCTION AND PURPOSE OF THE STUDY 1
II. LITERATURE REVIEW 3
1. Traditional Studies about Community 3
2. The Virtual Communities Built on Traditional Cultural Products 5
3. Communities in Cyberspace 5
III. BACKGROUND ABOUT YOUTUBE 8
1. YouTube is Gradually Replacing Traditional Television Programs 8
2. The Financial Incentive for Creating YouTube Videos 9
3. How Do the Videos Reach the Audience? 10
4. Subscriber Number is the “Currency” on YouTube 12
IV. METHODOLOGY 12
1. Stratified Sampling Method from YouTube Categories 12
2. Data Collection Procedures 15
1) Coding Standard Based on Exchange Theory. 15
2) Coding Procedure. 15
3) Data Analysis Method. 16
V. FINDINGS 20
VI. DISCUSSION 21
1. About the Channel Content 21
2. About Deep Engagement 22
VII. CONCLUSION 23
REFERENCES 26
APPENDICES 29
APPENDIX A: SAMPLED CHANNELS AND VIDEOS 30
APPENDIX B: QUANTITATIVE DATA OF EACH CHANNEL 32
APPENDIX C: LIST OF THE EMOTIONAL VOCABULARY AND CONVERSATIONAL VOCABULARY 34
APPENDIX D: ENGAGEMENT INDICES 35

LIST OF TABLES

Table 1. Sampling Method……………………………………………………………….14
Table 2. The Correlation Between Engagement Level and Subscriber Number ………..20
Table 3. Summary of Channels Received Fan's Contributed Subtitle…………………..23
Table 4. Sampled Channels and Videos…………………………………………………30
Table 5. Quantitative Data of Each Channel…………………………………………….32
Table 6. Engagement Indices …………………………………………………………...35

LIST OF FIGURES

Figure 1. Illustration of Youtube Hierarchy ……………………………………………13
Figure 2. Shallow Engagement Score by Subscriber Number………………………. 21
Figure 3. Medium Engagement Score by Subscriber Number……………………… 21
Figure 4. Deep Engagement Score by Subscriber Number…………………………...21
Figure 5 Engagement Index by Subscriber Number…………………………………..21

I. INTRODUCTION AND PURPOSE OF THE STUDY

Community is one of the oldest topics of study in social science research. Sociologists studied communities initially as a territorial concept with human beings in interaction with one another in a physical environment (Bell and Newby 1974). This early formulation of the concept of community as linked to place or geographic entity is not held today. The term now includes people in interaction with one another even when they are widely dispersed. Since the development of the Internet, communities in cyberspace have also attracted sociologists' attention. These virtual communities have been observed and studied in the past twenty years. As early as 1995, there were studies about topic-oriented collective discussion groups (Wellman and Gulia 1999). Email lists, online discussion forums, Bulletin Board Systems (BBSs), and World Wide Web sites all have become avenues for human interaction with participants spread across the globe.

Sociologists have also analyzed companionship, supportiveness, information, and sense of identity in cyber communities in previous research (Wellman and Gulia 1999). Cyber community members pass on free advice and provide digital goods which become public goods to other members. In return, the contributors receive help from others for their contributions, build reputations, and become influential persons in the community (Kollock 1999:220:243).

YouTube.com (YouTube) is the world's largest video sharing website. This type of media has highly targeted audiences. Most YouTube channels have a focused topic, such as gaming, fashion, technology, etc. The audiences are actively searching for information about a certain topic. Content creators provide information that can provide knowledge, stimulation, fulfillment, entertainment and relaxation to the audiences. Besides passively receiving information, the audiences also have an opportunity to socialize online. The comment sections allow audiences to have conversation with content creators and with one other. I observed the human interaction on this platform while following hundreds of YouTube channels; content creators have conversation with the audiences and audiences also talk to each other.

Despite the popularity YouTube has gained in the past decade, studies on YouTube communities are very limited. When people evaluate a YouTube channel's efficacy, the most important index is the subscription numbers of the channel. Subscription numbers only reflect the number of followers of the content creator and may not be associated with how engaged the audiences are. Audiences can express emotions, write long arguments, and have conversations with each other. Audiences also give material gifts, purchase merchandise associated with the channel, and even make monetary contributions to the content creators. In the past two years, I observed these activities in YouTube communities by watching thousands of videos and subscribing to hundreds of channels on different topic categories. It became clear to me that these activities could all be indicators of how engaged audiences are in these channels. However, they are currently not taken into account when we evaluate a channel's impact.

In this study, I propose a new measure to evaluate the level of engagement in YouTube's virtual communities, i.e., YouTube User-Generated-Content (UGC) channels. The objective of this study is to contribute a new tool to assess levels of engagement among participants through the viewer comments generated on the channels. This effort could contribute to the sociological literature on this newly emerged virtual community and shed light on its dynamics.

This study may also have some marketing value. Subscriber number is the "currency" on YouTube. When YouTube content creators try collaborating with the other creators to expand to broader audiences, or reach out to business agencies for sponsorship, subscriber number is the only bargaining chip they have to represent the channel's value. The method developed in this research can be generalized in the future to help businesses and agencies looking for sponsorship opportunities in YouTube's UGC channels to invest their capital efficiently.

II. LITERATURE REVIEW

1. Traditional Studies about Community

Community is a word with many meanings. One of the first American sociologists to define community was Robert Park. Park and his colleagues in the Sociology Department at the University of Chicago (The Chicago School) were instrumental in establishing community as a central concept in American sociology. Park (1936) described the community in the following terms:

> The essential characteristics of a community, so conceived are those of: (1) a population territorially organized, (2) more or less completely rooted in the soil it occupies, (3) its individual units living in a relationship of mutual interdependence. (Park 1936:3)

Park's definition was the first attempt to give a clear meaning to the sociological concept of community. It emphasizes the physical environment or territory in which people live and interact as a defining characteristic of the community. Hence, human ecology came to describe this approach to studying the community (Bell and Newby 1974). The human ecology model while studying people living in and identifying with a particular place gives special attention to the type and quality of their interaction (Lyon 1999).

Based on the type of human interaction, Ferdinand Tönnies's (2001[1876]) *Community and Civil Society*, described two types of community. The first, *Gemeinschaft*, is based on small rural village settings, where people connect by tradition, sentiment and common bond. *Gesellschaft* on the other hand is based on industrialized urban settings, in which people connect by exchange. According to Tönnies, both *Gesellschaft* and *Gemeinschaft* are ideal types. No place is purely *Gesellschaft* or *Gemeinschaft*. Emotional, sentimental, value oriented, and commodity exchange based relationships could all be the elements of bonding in a community.

Less than twenty years after Park's attempt, Hillery (1955) found more than ninety-four definitions of community, in which, the community is a group, a process, a social system, a geographic place, a consciousness of kind, a totality of attitudes, a common lifestyle, the possession of common ends, local self-sufficiency, and on and on. The concept of community is thus loosely defined in sociology. Bell and Newby's (1974) examination of Hillery's efforts pointed out that a vast majority of the definitions "agreed on the joint inclusion of social interaction and common ties" (p. 29) as more important than local area. It is clear then that the meaning of community now goes beyond the geographic limitations proposed by the human ecology model.

One such community that has emerged since the technological developments of the late 20th Century is the virtual community. As Erickson (1997) described it, the virtual community is a "computer mediated social interaction among large groups of people, particularly long term, textually-mediated interaction" (p. 1). In the virtual community, communication is by way of typing text via the medium of a computer or similar device. The text can be read by members of an interest group or the general public who may respond by similar text. Participants are not limited by geography and time. This text communication is usually available beyond the time they are posted and thus serves as collective memory of participants' thoughts and behaviors.

2. The Virtual Communities Built on Traditional Cultural Products

In traditional cultural production industries, audiences were considered vulnerable to the cultural production and totally autonomous. Mutual exclusion was a trait of traditional mass media. Producers and audiences did not have the means to connect instantly. Producers could not receive feedback from the market until they received mail from the fans. Also, only enthusiastic audiences would write to the producing company to share their thoughts or critiques of the content. The large majority of audiences did not have the incentive to provide feedback to the program creators.

Jenkins (1993) studied the fan culture on the fandom in film and television, which was identified as "media fandom." Fandom is a fan community surrounding TV shows, movies, books and other cultural products. Fans participation in cultural production had been practiced in traditional media. Fans often write critiques; fans make music videos; fans even write scripts for their favorite shows (Star Trek had taken fan contributed script into their product). However, the overall interaction between the fans and creators is on a small scale.

3. Communities in Cyberspace

Sociologists have been wondering for over a century about how technological changes (along with industrialization, urbanization, bureaucratization, and capitalism) have affected community (Wellman and Leighton 1979). Social interaction is no longer limited to the geographic boundary (Kollock and Smith 2002). Cars, airplanes, and telephones can help people maintain relationships over long distances (Wellman 1988). Sociologists realized that communities do not have to be solitary groups of densely-knit neighbors but could also exist as social networks of kin, friends, and workmates who do not necessarily live in the same neighborhoods (Wellman and Gulia 1999). One's "village" could span the globe. This conceptual revolution moved from defining community in terms of space-neighborhoods to defining it in terms of social networks (Wellman 1988). Social network analysts have had to educate traditional, place-oriented, community sociologists that community can be stretched well beyond the neighborhood (Wellman and Gulia 1999).

Virtual communities in cyberspace have attracted scholar's attention in the past decade. For example, Jenkins (2013) diverted his interests from TV fandom to Internet fandom, online gamers, and bloggers. UGC creators had been studied from the communication science perspective in the past. He observed that the "new knowledge communities will be voluntary, temporary, and tactical affiliations, defined through common intellectual enterprises and emotional investments" (p. 137).

Regardless of what form of media it is, audiences are gaining greater power and autonomy as they enter into the new knowledge culture. The communication obstacles between audiences and cultural products creators could not stop the keen enthusiasts to participate, and the development of technology, to be specific, the Internet, only enabled fans to participate more in their favorite cultural products.

The most actively studied virtual communities have been discussion forums or bulletin boards dedicated to a certain topic, such as health or a specific hobby (e.g.

beachbody.com), communities or practice intended for learning or professionals (e.g. khanacademy.com), enterprise communities or communities of transaction (e.g. amazon.com products review and Q & A), social networks sites (e.g. linkdin.com), wikis (e.g. Wikipedia.com), creative communities including open-source software development (e.g. github.com), and question-answering sites (e.g. quora.com). Individuals who share similar interests or seek specific knowledge could engage in discussion forums online. Members in these communities are connecting by similar interest.

The participatory activities of members of these virtual communities are intended to benefit other members who, in return, offer various rewards or sometimes expressions of dissatisfaction. This process of collaboration, participation, contribution, exchange, etc. lies at the heart of what is termed "community engagement."

The concept of community engagement in human ecology terms "is a process that brings together groups of stakeholders from a neighborhood, city, or region (including individuals, organizations, businesses, and institutions) to build relationships and practical collaboration with a goal of improving the collective well-being of the area and its stakeholders" (Maurrasse 2010:223). This same process is at work in the virtual community where members, by various actions, exhibit a sense of engagement.

Homans (1958) brought up the view that human interaction is "an exchange of goods, material and non-material." Blau (1964) also argued that social interaction has value and that people exchange these values. Blau's theory rests on the anticipated rewards of association, with rewards being both intrinsic (pleasure of being with someone) and extrinsic (a good or service that someone can provide). Participants in a virtual community contribute labor and knowledge and express emotions. In return, they receive emotional or other kinds of reward for their participation.

Three aspects of community engagement can be derived from the above in the context of virtual communities. The first is quantitative and can express the amount of participation by means of the number of people participating and the quantity of exchange in a given period. The higher the number, the larger the engagement. The second is qualitative and reflects the emotions expressed during the exchange, some positive and some negative. This emotional expression is quantifiable by counting the frequency of it. The third is labor and monetary contribution to the community. These three aspects are the theoretical foundations of the measuring strategy in this paper.

III. BACKGROUND ABOUT YOUTUBE

1. YouTube is Gradually Replacing Traditional Television Programs

YouTube has a slogan: "broadcast yourself," which means everyone can upload videos to YouTube; everyone can be a content creator on YouTube. As a free broadcast platform, YouTube does not charge any fee for uploading video content or to show up in search results. YouTube does not censor the contents unless users report the video had violated third party's copyrights or privacy.

Currently, YouTube ranks as the second most viewed website after Google. More than one billion unique users visit YouTube each month and over 6 billion hours of video are watched each month on YouTube (Neilsen 2015). That is almost an hour for every person on Earth. At least 100 hours of video are uploaded to YouTube every minute. According to Nielsen, the TV rating company, YouTube reached more 18-49 year-olds in the US than any cable TV network in December 2014 (Neilsen 2015).

2. The Financial Incentive for Creating YouTube Videos

When YouTube started in 2005, it was just a video sharing web site. The majority of the videos were User-Generated Content (UGC). The UGC videos on YouTube were typical Internet videos, generally short, mostly humorous, and easy to access.

After Google Inc. purchased YouTube in 2006, YouTube UGC could be monetized by enabling display of advertisements with the videos. This raised tension between YouTube and traditional media: because a large amount of YouTube videos were not generated by the uploaders. They were rather Professional-Generated Content (PGC) that could infringe copyrighted material.

Copyright infringement issues did not stop YouTube's growth. Instead, YouTube and traditional media utilized each other through a revenue sharing program (Kim 2012). For example, Vevo is a video hosting service YouTube offered to Sony Music Entertainment (SME) and Universal Music Group (UMG). YouTube hosts the music videos for SME and UMG and shares the advertising revenue. YouTube also offers training to UGC creators about copyright knowledge through YouTube's blog (YouTube 2015) and YouTube Creator Academy channel (YouTube Creator Academy 2015). YouTube gives users the restrictions on what videos they can upload, and the opportunity to monetize their videos.

YouTube Partner Program is an advertising revenue sharing program and copyright protection policy. It helps YouTube UGC channels to grow sustainably, since UGC channels also can be potential victims of copyright infringement. The top YouTube creators can make their living by creating UGC videos. Making YouTube videos becomes their profession. Therefore, the boundary between PGC and UGC is blurred on YouTube. In this study, I use the term UGC loosely to refer to the independent production channels, in contrast to the products generated by traditional media companies.

The advertising revenue-based business model allows channels focused on entertainment to grow rapidly due to the nature of the content. Since entertainment content is more accessible and sharable than educational material, it naturally receives more view count than educational content. Another monetization tool is fan funding, launched in 2015. It allows audiences to directly make monetary donations to a YouTube channel. Before YouTube enabled fan funding, fans were already making contributions to content creators, through websites such as Patreon.com. Receiving fans funding allows YouTube UGC channels to have more income which also enables channels to focus on serious topics, such as education and documentaries to grow.

When a channel grows to a certain scale, it can also receive commercial advertisement deals from third party sponsors. Business agencies are willing to invest marketing funds through this targeted media to reach targeted demographic groups. Sponsored income is another stream of income for UGC creators.

3. How Do the Videos Reach the Audience?

In traditional media, audiences passively receive content from TV channels or movie theaters although, there are some choices for them. YouTube audiences, however, actively search for videos which interest them or fulfill the information they need. The audiences are self-motivated to watch the video instead of just receiving whatever they have been offered.

If the audiences watch the video anonymously, the watch history will not be stored. However, because of the large number of Google users, audiences can watch YouTube videos while logging on to their Google accounts. Therefore, the watching history will be stored and analyzed, which leading to the next level of audience reaching strategy, the suggested videos.

Suggested videos will show on the side of the current video the user is watching. At the end of the video, YouTube will also build up a unique home page for the user. The suggested videos are selected by YouTube in regards to the user's watching history which not only includes related videos to the one just watched, but also includes prediction of what this specific user would be interested in.

At the end of each video, YouTube offers embedding codes which allow users to share the video on their own social network websites, blogs, and other public forums. Sharing will bring the video to more audiences who share similar interests with the initial viewer. Thus, YouTube pushes the videos to broader audiences with high possibility of interest in the content by means of the process mentioned above.

In addition, YouTube offers a subscription option to users to help them engage in the channel. Subscribed users receive timely notification about the channel's activities through Google Plus, email, or push notification on their mobile devices. Therefore, the YouTube community is formed. The community is a key component that differentiates YouTube from traditional media.

4. Subscriber Number is the "Currency" on YouTube

When YouTube creators look for collaboration opportunities with other YouTube creators or negotiate with external sponsors, subscriber number is the most important factor to prove the channel's value. In most situations, third party sponsors only look at the subscriber number to estimate the commercial value of the channel.

However, to measure the value of a channel solely based on the subscriber number is limiting. Subscriber number can only show how many followers a channel has; it cannot show how engaged the audiences are. The level of engagement can significantly affect the efficacy of the advertising. Members in high engagement communities would have higher possibility to be influenced by the community leaders' opinion and behavior. When the stakeholders endorse a product, members who are highly engaged into the content are more likely to be convinced. The measure developed in this research takes audience engagement in the UGC into account for a more elaborate way of assessing the value of a channel for advertising and other commercial purposes.

IV. METHODOLOGY

1. Stratified Sampling Method from YouTube Categories

On September 1, 2016, YouTube.com classified videos into the following 12 categories: Music, Gaming, Film & Entertainment, Comedy, Sports, Beauty & Fashion, Technology, Science & Education, News & Politics, Cooking & Health, Automotive, and Animation. However, the channel amount in each category is not evenly distributed. Some categories contain more channels than other categories.

For statistical analysis purpose, I intended to sample 30 channels. However, after I calculated the ratio of channels in each topic category, several channels had such low proportions that a sample of size 30 would not have a video from each category. Therefore, I adjusted the sample size to allow for at least one video from each category. This resulted in a total sample size of 31. Figure 1 illustrates the hierarchy of YouTube video categories. The sampling process is detailed in table 1.

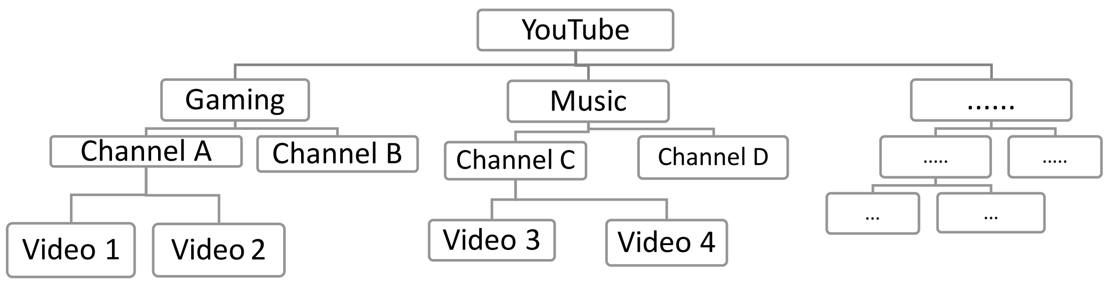

Figure 1. Illustration of Youtube Hierarchy

Table 1. Sampling Method

(Sampling was conducted on September 1st, 2016)

Category	Total Channel	Percentage	Proportional Sample (Number of Sample)	Number of Channels
Gaming	1405	27%	8.0	8
Music	1349	26%	7.7	7
Film & Entertainment	574	11%	3.3	3
Comedy	553	11%	3.2	3
Sports	412	8%	2.4	2
Beauty & Fashion	419	8%	2.4	2
Tech	64	1%	0.4	1
Science & Education	78	1%	0.4	1
News & Politics	58	1%	0.3	1
Cooking & Health	97	2%	0.6	1
Automotive	100	2%	0.6	1
Animation	148	3%	0.8	1
Total	5257	100%	30.0	31

Once the number of channels to include within each category was determined, I selected the specific channels to include based on the following criteria:

The channel had to be:

1. An independent production channel. For example, musicians who signed with traditional label companies or TV show programs uploaded by traditional TV production companies were excluded.

2. Using English as the primary language.

3. If a channel met these criteria, channels were then picked beginning with the highest ranked channel and selecting the next highest and so on, up to the total number of channels to be selected within a topic category.

For each channel selected, I analyzed the video with the highest view count in the selected channel. If comment was disabled for that video, I chose the video whose subscription number ranked next to it. See APPENDIX A for details of sampled YouTube channels and their most popular videos.

2. Data Collection Procedures

1) Coding Standard Based on Exchange Theory.

I was specifically looking for what the content creators offer, and what they received from the audiences, material and non-material. YouTube carries content from calculus tutorials to how to put on mascara; from what is happening in Syria to how to clear a course in Angry Bird. Each content has a highly targeted population, which forms an online community. The different contents may attract different audiences. Different content stimulates different behaviors. YouTube audiences receive information, entertainment, and friendship through following certain creators. They communicate with each other. When we measure community engagement, community members' sentiments and activities can be observed among group members. The length of the comments, the keyword indicators of strong emotion (positive and negative emotion), appreciation, request, suggestion, and the conversation could be used to measure the community engagement.

2) Coding Procedure.

The first level of quantitative data of each video was collected before coding the other content. I recorded like/dislike count, comment count and publishing time of each video into a spreadsheet. Then I looked at the overall channel creator's interaction with the audiences: 1. Does he/she make conversation with audiences? 2. Does he/she receive monetary contribution? These answers were recorded in the spreadsheet.

Next, I downloaded the comments of the sampled videos. Some comments could not be downloaded, however, the comments represented at least 80% of the total comments in the sampled video. See APPENDIX B for detailed quantitative data.

Next, I coded the qualitative video content. I specifically looked for the following traits: 1. Does it provide some information? 2. Is the content entertaining? 3. Does the creator express emotion? 4. Does the creator invite the audience into the future decision making process? I recorded the results in a spreadsheet in binary format.

The next step was coding of the comments. The length of the comments (word count in each comment), the vocabulary indicators of strong emotion (positive and sad emotion), insulting vocabulary (cursing language), appreciation, request/suggestion for future content, and the conversation between audiences were coded into a spreadsheet. If the audiences wrote long comments, it was considered deep engagement. The emotional expression and conversation between the audiences counted as the exchange frequency.

3) Data Analysis Method.

I developed the engagement indices to measure the community engagement level. There are three layers of engagement: shallow engagement, medium engagement and deep engagement. I wrote a customized Python program to analyze the vocabulary and the length of the comments.

Shallow engagement

The first measure of engagement is shallow engagement. Shallow engagement was calculated from four quantitative variables: view count, like count, dislike count, and comment count. See APPENDIX B for detailed counts.

The shallow engagement score is calculated by the following formula:

$$shallow\ engagement\ score = \frac{comment\ count + like\ count + dislike\ count}{view\ count} * 1000$$

I multiplied the number by 1000 for easy interpretation. Also, considering the view count is the divisor, it is not necessarily generated by a unique visitor. A single user could loop the video multiple times which would increase the view count, but they may like or comment on the video only once. Therefore, I inflated the shallow engagement score by multiplying it by 1000.

Medium engagement

Medium engagement was measured by the frequency of the following three types of interactions:

1. The conversation between the community members. If a comment was made in response to an existing comment, it was counted as one interaction.
2. The comment intended to speak to the others who are watching the video was counted as one interaction. These comments usually start with the subject “who” such as, “Who is still watching this in 2016?” “Who also feels … as me? (See APPENDIX C for a full list of the vocabulary that indicates seeking companions.) Since this type

of comment is intended to look for people who share similarity, I counted every single record seeking companions as one interaction.

3. The vocabulary in the comment indicates an emotional expression. I counted the frequency of emotional words. Any word in the list increased the emotional point of the corresponding emotion category: positive emotion (funny, awesome, etc.), sad emotion (devastated, pain, etc.), appreciation (thanks, thank, etc.), request/suggestion (please, next, etc.), complimentary (quality, real, etc.), and insult (kill, fuck, etc.). A full list of the vocabulary is in APPENDIX C.

$$Medium\ engagement\ score = \frac{conversation\ frequency + seeking\ frequency + emotion\ frequency}{downloaded\ comment\ count}$$

Deep engagement

Deep engagement was measured by three variables related to contributing labor and money:

1. Audiences contributing subtitles. Audiences contribute subtitles in English and other foreign languages. Each subtitle (including English) was counted as one deep engagement record. If one video has seven subtitle languages, it counted for seven points of deep engagement. Making subtitles and translating the narrative languages into foreign languages is donating free labor to the content creators to help their channel access larger audiences.

2. Long comments. Comment length more than 250 words was considered as a long comment. The reason I choose 250 word-count to define the long comment is that a one-page document with 12-point font format, double spaced is about 250 words. Writing an essay length in the comment section constitutes a long comment.

If for one video two or more comments were longer than 250 words, this video only received one deep engagement point. The reason I only counted one point was that regardless of the possibility of multiple long comments, there were potential spams in the comment section. Audiences could spam the comment section by copying & pasting of some text. During the analysis, I found out quite a few comment lengths were identical, which raised my caution about spam comments. However, because I used a Python program to analyze the length of the comments, I could not see the detailed content of the comments.

3. Audiences donate money to the channel. When the videos are free to watch, audiences are willing to donate money to the creator. This is considered as *deep engagement*. I could only see whether the sampled channels enabled the fan funding function, but I could not see how much funding he/she raised. Therefore, whoever enabled the fan funding function received one deep engagement score.

$$Deep\ engagement\ score = subtitle\ amount + \lambda * long\ comment + \beta * fan\ funding$$

*λ=0 if the sampled video did not receive comment longer than 250 words;
λ=1 if the sampled video received comment longer than 250 words;
β=0 if the sampled channel did not enable fan funding
β=1 if the sampled channel did not enable fan funding

In order to analyze the relationship between subscriber number and engagement, I ranked the sampled channels by their subscriber numbers. Then I calculated the engagement index by the following formula:

$$\text{Engagement index} = 1 * \text{shallow engagment rate} + 3 * medium\ engagement\ rate + 5 * deep\ engagment\ point$$

Every shallow engagement score is worth one point, medium engagement score is worth three points, deep engagement score is worth five points.

V. FINDINGS

I calculated scores for each level of engagement and the engagement index for each channel. Table 2 shows the direction and magnitude of correlation of each level of engagement with subscriber number. Figure 2, Figure 3, Figure 4 and Figure 5 show the channels' different engagement levels by their subscriber number. See APPENDIX D for detailed engagement score of each channel.

Table 2. The Correlation Between Engagement Level and Subscriber Number

Engagement Level	Direction	Magnitude
Shallow	positive	0.68
Medium	positive	0.03
Deep	negative	-0.25
Index	negative	-0.18

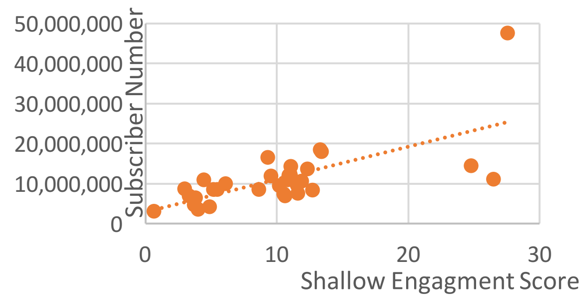

Figure 2.Shallow Engagement Score by Subscriber Number

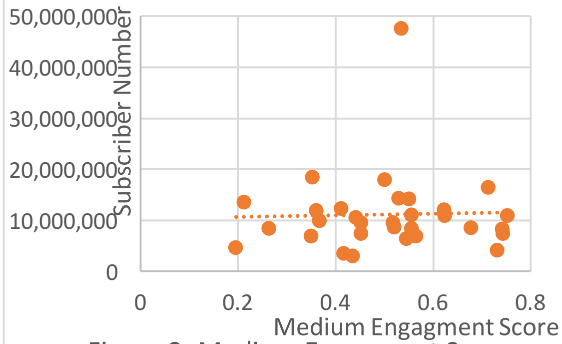

Figure 3. Medium Engagment Score by Subscriber Number

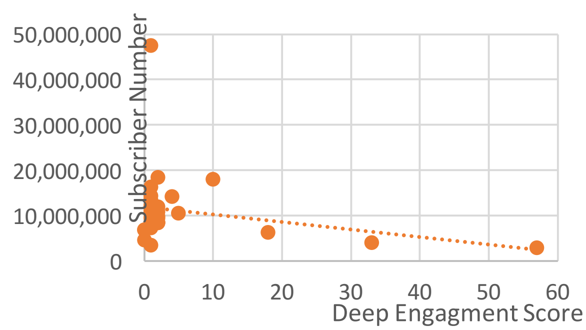

Figure 4.Deep Engagment Score by Subscriber Number

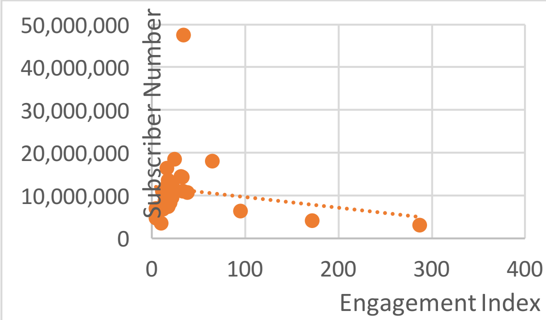

Figure 5. Engagement Index by Subscriber Number

What we can see here is that subscriber number is positively correlated with shallow level of engagement. More subscribers indicate more likes, dislikes and comments. However, subscriber number does not reflect medium and deep engagement in the channel's community. What makes the engagement level different among the channels is the deep engagement score. Some channels attract more fans to contribute their labor, which leads to higher engagement indices. The medium engagement among channels does not differ much, regardless of what content the video provides.

VI. DISCUSSION

1. About the Channel Content

During the coding process, I discovered that every creator connects with their audiences through means other than posting videos on YouTube. Everyone is active on other social network platforms, such as Facebook, Twitter, Instagram, Snapchat, etc. Twenty out of 31 channels have a secondary vlog channel, which mainly talks about the creators' personal lives and behind the scene stories. I consider these behaviors could bring audiences and creators closer. Almost everyone makes the video content conversational or "talks" to audiences in the video description. Every sampled channel sells something: T-shirts, mugs, posters with channel logo on them, or music. Some (3 out of 31) have P.O. boxes to receive material gifts (e.g. post cards) from fans.

2. About Deep Engagement

Audiences volunteering to translate the narrative into foreign languages makes the degree of deep engagement different. Seven out of 31 sampled videos received fan-contributed subtitles (auto generated English subtitle does not count). Interestingly, they mainly focus on educational and informative content. Among eight sampled gaming channels, only one received fan contributed subtitles. In general, the education and information sharing channels are more likely to receive fan-contributed subtitles. Nigahiga's channel self identifies as a comedy channel. The sampled video from his channel offers sarcasms about culture. Nigahiga's video, *Nice Guy*, talks about the dilemma that nice guys face in the dating market: girls are more attracted to bad boys. That video received 10 subtitles in different languages. Table 3 shows the channels that received fan contributed subtitles.

Table 3. Summary of Channels Received Fan's Contributed Subtitle

Channel Name	Category	Subscribers	Subtitle Contribution
nigahiga	Comedy	18,020,353	10
Markiplier	gaming	14,274,344	2
Vsauce	Tech	10,652,521	4
Michelle Phan	Beauty & Fashion	8,705,936	1
How It Should Have Ended	Animation	6,434,420	17
Smarter Every Day	Science & Education	4,152,649	31
TYT (The Young Turks)	News & Politics	3,059,315	56

What we can see here is that *The Young Turks* which focuses on politics and news received the most subtitles. Their video has 56 subtitle entries in different languages. *Smarter Every Day* focuses on recreational science content. The sampled video in this channel has 31 subtitles. *How It Should Have Ended,* an animation channel, produces animated parody alternative endings to major motion pictures. Movie fans are deeply engaged in this channel.

VII. CONCLUSION

This study developed an index to measure the level of community engagement in 31 YouTube UGC channels by analyzing the comments and other elements of their most popular videos (popularity is defined by view count). The result shows subscriber number does not always positively correlate with community engagement level.

By studying YouTube's virtual communities, I clearly observed community members' interaction, contribution, and exchange (both emotional and material) in the virtual community in the same way as these would occur in traditional communities based on territory. Members in this virtual community communicate with each other through a textual approach. The "village" has expanded to a global level. The largest YouTube channel, *PewDiePie* had more than 44 million subscribers, which is more than four times greater than its home country Sweden's population of 9.9 million (US Bureau of the Census 2016). The Internet connects people worldwide. Virtual community members' shallow and medium interaction is easily quantifiable by the measuring method I developed in this study. Deep engagement behavior is measured more qualitatively than quantitatively. Furthermore, the findings confirm that a smaller subscriber number is associated with stronger engagement in the community just as Durkheim's and Tonnies' works indicate stronger social bonds in smaller communities.

One limitation of this study is that I did not sample any channel with relatively low subscriber number. They may have higher community engagement than larger channels. Sociologically speaking, small groups are expected to have more intimate relationships among group members. Because I only sampled one video from each channel, I could not include some other variables to measure the engagement, such as repeat visitor rate in the channels. Shallow engagement is positively correlated with the subscriber number. It could be because big channels have more fans to like, dislike and comment on the videos. It also could be a result of the algorithm YouTube uses to promote videos. For example, YouTube may be likely to put a video that received high view count, or high like/dislike count on the front page so more viewers can reach it, which leads to more subscribing to this channel. Also, another limitation of this study is that some sampled videos were old. YouTube launched the fan contributing subtitle function in November 2015 (YouTube 2015). Some sampled videos do not have fan-contributed subtitles, but the newer videos in that channel have. The sampled video may not represent the overall engagement level of each channel. Despite this fact, the ones that have fan-contributed subtitles were all uploaded before November 2015. It means the audiences visited the old videos of the channel, and translated the content after YouTube offered this function. Audiences are clearly attracted to the content of the videos. Finally, it is also possible that the correlations reported were influenced by outliers in the data.

Delivering advertisements on web pages enables companies such as google.com and facebook.com to grow into such large scale operations. However, quite recently, a new business model developed. Instead of receiving free information bundled with advertisements, consumers now have a choice to pay a fee to remove advertisements. From the music streaming services provided by Spotify and Apple Music to newly launched YouTube Red, consumers can pay a monthly subscription fee to have an ad-free experience. Whoever pays for a free service could have better consuming power than the users who prefer to tolerate advertisements. Advertisements can therefore not reach the former. In addition, 16% of the US online population uses ad-blocking software to screen off advertisements on web pages (The Pagefair Team 2015). In the battle between consumers and advertisements, how could business agencies deliver their products information to the consumers? Sponsoring the UGC creators could be an effective means.

REFERENCES

Bell, Colin and Howard Newby. 1974. *Community Studies: An Introduction to the Sociology of the Local community*. New York: Praeger Publishers.

Blau, Peter. 1964. *Exchange and Power in Social Life*. New York: John Wiley and Sons.

Burgess, Jean and Joshua Green. 2009. *YouTube: Online Video and Participatory Culture*. Malden, MA: Polity.

Calhoun, Craig, Joseph Gerteis, James Moody, Steven Pfaff, and Indermohan Virk. 2007. *Contemporary Sociological Theory*. 2nd ed. Malden, MA: Blackwell Publishing.

Erickson, Thomas. 1997. "Social Interaction on the Net: Virtual Community as Participatory Genre." *Proceedings of the Thirtieth Hawaii International Conference on System Sciences* 6(10).

Hillery, George A. J. 1955. "Definition of Community: Areas of Agreement." *Rural Sociology* 20:779-791.

Homans, George C. 1958. "Social Behavior as Exchange." *American Journal of Sociology* 63(6):597-606.

Jenkins, Henry. 2006. *Fans, Bloggers and Gamers*. New York, NY: NYU Press.

Jenkins, Henry. 2013. *Textual Poachers, Television Fans and Participatory Culture*. Updated twentieth anniversary edition ed. New York, NY: Routlege.

Kim, Jin. 2012. "The Institutionalization of YouTube: From User-Generated Content to Professionally Generated Content." *Media, Culture & Society* 34(1):53-67.

Kollock, Peter. 1999. "The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace." Pp. 220-242 in *Communities in Cyberspace*, edited by Marc A Smith and Peter Kollock. New York, NY: Toutledge.

Kollock, Peter and Marc Smith. 2002. *Communities in Cyberspace*. New York, NY: Routledge.

Lyon, Larry. 1999. *The Community in Urban Society*. Long Grove, IL: Waveland Press.

Malinen, Sanna. 2015. "Understanding User Participation in Online Communities: A Systematic Literature Review of Empirical Studies." *Computers in Human Behavior* 46:228-238.

Maurrasse, David J. 2010. "Standard of Practice in Community Engagement." in *Handbook of Engaged Scholarship*, edited by Cathy Burack, Sarena D. Seifer Hiram E. Fitzgerald. Michigan State University Press.

McInnerney, J.M. and T.S Roberts. 2004. "Online Learning: Social Interaction and the Creation of a Sense of Community." *Education Technology & Society* 7(3):73-81.

Neilsen. 2015. "The Evolution of Digital Video Viewership." Retrieved September 19, 2016 (http://www.nielsen.com//content/dam/corporate/us/en/reports-downloads/2015-reports/nielsen-google-case-study-sept-2015.pdf).

Park, Robert. 1936. "Human Ecology." *American Journal of Sociology* 17(1:1-15):3.

Postman, Neil. 2005. *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*. New York, NY: Penguin Group.

Seth C. Lewis, Avery E. H. M. C. 2013. "Reciprocal Journalism." *Journalism Practice*.

Tönies, Ferdinand. 2001[1876]. *Community and Civil Society*. New York, NY: Cambridge University Press.

The Pagefair Team. 2015. “The Cost of Ad Blocking.” The Pagefair Team. Retrieved December 19, 2016. (https://downloads.pagefair.com/wp-content/uploads/2016/05/2015_report-the_cost_of_ad_blocking.pdf)

U.S. Bureau of the Census. 2016. “International Data Base.” United States Bureau of the Census. Retrieved December 15, 2016. (https://www.census.gov/population/international/data/idb/region.php?N=%20Results%20&T=10&A=separate&RT=0&Y=2016&R=-1&C=SW).

Wellman, Barry. 1988. *The Community Question Re-evaluated*. New Brunswick, NJ: Transaction Books.

Wellman, Barry and Milena Gulia. 1999. *Net Surfers Don't Ride Alone: Virtual Communities as Communities*. New York, NY: Routledge.

Wellman, Barry and Barry Leighton. 1979. “Networks, Neighborhoods and Communities.” *Urban Affairs Quarterly* 14(3):363-390.

YouTube. 2015. “Make Your Videos Even More Interactive with Cards.” YouTube Creator Blog. Retrieved November 19, 2016 (https://youtube-creators.googleblog.com/2015/03/make-your-videos-even-more-interactive.html).

YouTube. 2015. “Manage Fan-Contributed Captions and Subtitles.” YouTube. Retrieved November 21, 2016 (https://www.youtube.com/watch?v=jiEybiElXgk&t=1s).

YouTube. 2016. “Youtube Channels.” YouTube. Retrieved September 1, 2016 (https://www.youtube.com/channels).

YouTube Creator Academy. 2015. “YouTube Creator Academy.” YouTube. Retrieved November 19, 2016 (https://www.youtube.com/user/creatoracademy).

APPENDICES

APPENDIX A: SAMPLED CHANNELS AND VIDEOS

Table 4. Sampled Channels and Videos

(Sampling was conducted on September 1st, 2016*)

Channel Name	Category*	Subscribers number	Sampled VidID
PTXOffical	music	10,941,411	3MteSlpxCpo
boyceavenue	music	8,540,846	fvEZUbzqqyM
Bart Baker	music	8,456,976	U5gT8hf0Z_M
Lindsey Stirling	music	8,387,559	aHjpOzsQ9YI
NOCopytightSounds	music	7,516,653	bM7SZ5SBzyY
Kurt Hugo Schneider	music	7,409,216	a2RA0vsZXf8
Miranda Sings	music	6,898,213	eHpEIDpLg0g
PewDiePie	gaming	47,607,898	lxw3C5HJ2XU
VanossGaming	gaming	18,464,201	POmH7dDMDEc
KSI	gaming	14,409,167	708mjaHTwKc
Markiplier	gaming	14,274,344	iOztnsBPrAA
TheDiamondMinecart // DanTDM	gaming	12,304,504	T_vGAPvjOc8
Sky Does Minecraft	gaming	12,085,656	3V7wWemZ_cs
TheSyndicateProject	gaming	9,960,237	IrJkyFeJ2Fc
CaptainSparklez	gaming	9,540,870	cPJUBQd-PNM
ERB	Film & Entertainment	13,599,024	dX_1B0w7Hzc
Romanatwood	Film & Entertainment	9,616,096	R7AXBOT8KzU
The Slow Mo Guys	Film & Entertainment	8,460,348	j_OyHUqIIOU

Table 4. Sampled Channels and Videos(Continues)

Channel Name	Category*	Subscribers number	Sampled VidID
nigahiga	Comedy	18,020,353	xfeys7Jfnx8
JennaMarbles	Comedy	16,445,463	OYpwAtnywTk
CollegeHumor	Comedy	10,951,837	HIPJrrQlxzY
Dude Perfect	Sports	11,930,251	UtsfUAHkyWQ
freekickerz	Sports	4,664,237	zFRZfUCknUg
Zoella	Beauty & Fashion	11,076,047	b0vPzYhxy9c
Michelle Phan	Beauty & Fashion	8,705,936	J4-GRH2nDvw
Vsauce	Tech	10,652,521	jHbyQ_AQP8c
SmarterEveryDay	Science & Education	4,152,649	kxLoycj4pJY
TYT	News & Politics	3,059,315	IzOKhaiEz2A
Rosanna Pansino	Cooking & Health	6,948,000	q53HUAKB9oU
colinfurze	Automotive	3,524,411	bKHz7wOjb9w
How It Should Have Ended	Animation	6,434,420	JEdZ-yjxHLI

*YouTube lists 12 categories on the browse channels page (YouTube 2016). Because YouTube does not provide a list of the ranking in each category, I retrieved the subscriber number ranking data from the website: vidstatsx.com on September 1st, 2016 in order to complete the sampling process.

APPENDIX B: QUANTITATIVE DATA OF EACH CHANNEL

The quantitative data were collected on Sep 1st, 2016. Most of the comments were downloaded between Sep 1st to Sep 30th, 2016. The comments of *boyceavenue, nigahiga,* and *Sky Does Minecraft*, were downloaded on Nov 17th, 2016, due to initial download has too much missing cases.

Table 5. Quantitative Data of Each Channel

Channel Name	View Count	Like	Dislike	Comments	Downloaded Comment
PTXOffical	206,395,390	2,020,187	61,127	151,524	151,040
boyceavenue	108,383,137	875,243	13,207	46,416	45,061
Bart Baker	117,326,605	523,507	79,338	40,483	38,234
Lindsey Stirling	155,359,483	1,742,580	28,591	202,006	201,727
NOCopytightSounds	160,231,633	1,698,198	24,468	136,494	125,631
Kurt Hugo Schneider	125,389,494	1,165,446	12,096	143,530	143,226
Miranda Sings	52,928,351	373,071	149,426	39,675	37,515
PewDiePie	44,196,305	1,033,165	10,529	174,777	164,422
VanossGaming	32,206,571	405,432	4,745	18,039	17,387
KSI	51,783,111	1,113,887	79,886	89,700	83,881
Markiplier	51,910,688	410,385	21,755	142,186	129,479
TheDiamondMinecart // DanTDM	33,253,025	323,174	7,426	35,867	32,933
Sky Does Minecraft	71,499,061	615,585	20,757	143,465	140,598
TheSyndicateProject	25,292,907	119,448	15,110	19,541	19,268
CaptainSparklez	164,889,726	1,279,638	43,910	357,180	347,202
ERB	123,731,447	896,357	23,241	607,322	494,206
Romanatwood	83,255,568	864,559	32,632	63,369	59,530
The Slow Mo Guys	146,588,564	655,687	25,875	77,727	70,675
nigahiga	67,230,054	698,089	9,738	191,495	156,984
JennaMarbles	64,872,386	483,773	24,557	94,888	94,755
CollegeHumor	72,234,562	277,398	17,934	26,975	25,712
Dude Perfect	64,035,496	562,530	16,784	31,517	30,988
freekickerz	31,151,347	99,184	6,781	9,419	8,950
Zoella	19,024,828	462,531	11,809	29,890	27,574

Table 5. Quantitative Data of Each Channel (Continues)

Channel Name	View Count	Like	Dislike	Comments	Downloaded Comment
Michelle Phan	65,089,833	138,496	18,883	36,360	29,102
Vsauce	22,465,187	218,721	7,001	41,457	38,244
SmarterEveryDay	43,984,623	198,546	7,263	8,951	8,374
TYT	28,046,874	7,377	6,558	4,657	3,739
Rosanna Pansino	132,435,965	358,624	32,111	48,631	50,487
colinfurze	21,195,340	76,064	3,007	5,902	5,727
How It Should Have Ended	26,507,309	83,377	5,373	12,723	10,597

APPENDIX C: LIST OF THE EMOTIONAL VOCABULARY AND CONVERSATIONAL VOCABULARY

There are the lists of the vocabularies express emotion and seeking conversation with the other audiences.

Positive emotion: magical, harmony, wow, :D, emotion, emotional, awesome, love, beautiful, laugh, laughing, joke, jokes, joking, lmao, funny, favorite, favourite, miss, right, better, amazing, cute, bravo, holy, omg, aww, good, best, better, fabulous, entertaining, hilarious, happy, lol, feeling, memories, memory, impressive.

Insulting: kill, fuck, bitch, bad, fucked, annoying, nasty, nasties, hating, ashamed, idiot, terrible, inappropriate, ugly, WTF, stupid

Sad emotion: devastated, never, sad, cry, unfair, R.I.P, tragedy, pain, pains, hate

Appreciate: helpful, thanks, thank

Request/suggestion: want, wonder, hope, hey, request, wish, plz, please, nex

Seeking: we, who, who's, 2016,

Complement: real, quality, great, nice, impressive

APPENDIX D: ENGAGEMENT INDICES

Table 6. Engagement Indices

Channel Name	Subscribers number	Shallow	Medium	Deep	Engagement Index
PewDiePie	47,607,898	27.57	0.53	1	34.17
VanossGaming	18,464,201	13.30	0.35	2	24.35
nigahiga*	18,020,353	13.38	0.50	10	64.88
JennaMarbles	16,445,463	9.30	0.71	1	16.44
KSI	14,409,167	24.79	0.53	1	31.37
Markiplier	14,274,344	11.06	0.55	4	32.72
ERB	13,599,024	12.34	0.21	1	17.98
TheDiamondMinecart // DanTDM	12,304,504	11.02	0.41	1	17.26
Sky Does Minecraft*	12,085,656	10.91	0.62	2	22.78
Dude Perfect	11,930,251	9.54	0.36	1	15.62
Zoella	11,076,047	26.50	0.56	1	33.17
CollegeHumor	10,951,837	4.46	0.75	1	11.72
PTXOffical	10,941,411	10.82	0.62	2	22.69
Vsauce	10,652,521	11.89	0.44	5	38.22
TheSyndicateProject	9,960,237	6.09	0.37	2	17.19
Romanatwood	9,616,096	11.54	0.52	1	18.09
CaptainSparklez	9,540,870	10.19	0.45	2	21.55
Michelle Phan	8,705,936	2.98	0.52	2	14.54
boyceavenue*	8,540,846	8.63	0.68	1	15.66
The Slow Mo Guys	8,460,348	5.18	0.26	2	15.97
Bart Baker	8,456,976	5.48	0.56	2	17.15
Lindsey Stirling	8,387,559	12.70	0.74	1	19.93
NOCopytightSounds	7,516,653	11.60	0.45	1	17.96
Kurt Hugo Schneider	7,409,216	10.54	0.74	1	17.77
Rosanna Pansino	6,948,000	3.32	0.35	0	4.37
Miranda Sings	6,898,213	10.62	0.57	0	12.32
How It Should Have Ended	6,434,420	3.83	0.55	18	95.46
freekickerz	4,664,237	3.70	0.20	0	4.29
SmarterEveryDay	4,152,649	4.88	0.73	33	172.08
colinfurze	3,524,411	4.01	0.42	1	10.26
TYT	3,059,315	0.66	0.44	57	286.97