

Content characteristics of College Student Vlog Celebrities on Bilibili

Tingting Chen

*Business School, Sichuan University, Chengdu, 610065, China
Email: arya-t@foxmail.com*

ABSTRACT. As an upgraded version of blog, vlog has the characteristics of personalization and documentary distribution, emphasizing single-person perspective and self-timer mode. It has switched the way people socialize on the web. College students are the main force of vlog creation due to their ability to quickly absorb novelty and having a strong sense of innovation. This paper analyzed vlog content of college student vlog celebrities on Bilibili. In order to identify vlog celebrities, personal attractiveness and audience participation were regarded as identification indicators through PSI theoretical analysis. Accordingly vlog data was crawled for a two-step clustering. After that content analysis method is applied to vlog content analysis. The study shows that the themes of vlogs shooting by college student vlog celebrities on Bilibili are diverse but unevenly distributed. Vlog celebrities more often than not focus on 1 or 2 vlog themes, among which those closely related to college students' identity attributes gain higher popularity. The content tends to be homogenous in the same type of vlogs. The paper suggests that vlogs about campus experience can serve as a new way of university publicity.

KEYWORDS: Vlog celebrity; Parasocial-interaction; Content analysis; Two-step clustering

1. Introduction

With the rapid development of We-media, video blog, vlog, has ushered in its blow-out growth due to the current consumption habits of fragmented content. In China, social media giants such as Weibo and Tencent are actively building the layout of vlog business. However, Bilibili, a website which focuses on building niche circles and cultural endogenous dynamic mechanism, has become a leader in the vlog market. On the one hand, Bilibili is committed to creating a video website with social characteristics, which exactly matches the cultural core of vlog. Similar to the YouTube that is popular all over the world, Bilibili has rich UGC and PUGC video content catering for the needs of people's social interaction. On the other hand, as an emerging market, vlog has various themes and content forms, which is in line with the needs of the millennial generation. According to the prospectus of Bilibili,

81.7% of its users are born in 1990-2009^[1], most of them are still in their studentship. They have a strong curiosity for novelty, and are also eager to accelerate the process of self and social integration via plugging into virtual social space. Thus creating vlogs, watching and commenting them have become an outlet for this desire.

In Bilibili, numerous college student vloggers have accumulated tens of thousands of fans by updating vlogs through normalcy and maintaining interaction with followers. Their vlogs are influential in certain fields, making them gradually grow into vlog celebrities. By showing the charisma or positive behaviors in vlogs and actively engaging in the interaction with followers at a high frequency, vlog celebrities have an impact in some way on their vlog audience. Till now, there is a lack of study conducted on researching vloggers and the vlog content characteristics using quantitative and qualitative research methods. This paper provides a theoretical framework, the combination of PSI theory and two-step clustering method, to identify college student vlog celebrities by observing a specific group of vloggers on Bilibili. Through content analysis method, the author reveals the vlog content characteristics of these vlog celebrities.

2. Literature

Parasocial-interaction(PSI) was originally proposed by Horton and Wohl in 1956^[2]. The theory describes that when users watch or get to know their favorite characters through media, they unconsciously evolve a kind of affective attachment towards these characters. This attachment will promote diverse reactions from users making them develop an imagined interpersonal relationship. After this definition, other scholars' definitions of parasocial-interaction are mostly similar to Horton's, all of which clearly indicate that parasocial-interaction is a one-way behavior, but individuals can form strong relationships with media characters through this^[3-4]. With the advent of emerging media technologies, PSI theory is widely used to explain audience behavior and mass communication effects. Despite the theory is first based on traditional TV or radio media, it is also applicable to the emerging media such as the Internet. Lots of scholars latched onto the effects or motivations of parasocial-interaction. Cole^[5] invited students as participants and investigated their motives for parasocial interactions. Research by Bao Dunan^[6] and others showed that there is a parasocial-interaction relationship in the virtual community environment. Now that Bilibili serves as a video sharing platform, in which the interactive community between the video producers and the audience should be classified as a category of parasocial-interaction.

Vlog celebrities with large followings are more likely to have and maintain strong parasocial interactions with their followers. This strong parasocial-interaction helps vloggers build their own commercial brands as well as increase their popularity. In the field of PSI research, many scholars focus on the influencing factors of parasocial-interaction in social media. Studies by Turner^[7] and Rubin^[3] found that role attractiveness is an important factor and a reason that affect the audience's perception of parasocial interactions. Hall^[8] and others discovered that the characters often appearing in the media have stronger parasocial interactions

with their fans. Sengupta^[9] and Horton^[2] including some scholars claimed that the amount of time users spending on watching media characters is related to the amount of effective role information they obtain. This will significantly affect the degree of parasocial-interaction. Based on the research of the above scholars,taking Bilibili into account, it can be summarized that vlogger's personal attractiveness such as appearance and personality, and vlog content quality, like the number of posts, playback volume and comments etc.will affect the strength of the parasocial-interaction between vlogger and its audience,thus helping identify the vlog celebrities.

3. Method

3.1 Vlog celebrity identity design

According to PSI theory, on Bilibili, behaviors of users like following vloggers, sending comments or clicking like are mainly one-way. Similar to stars interacting with fans, fan's stronger willingness to participate in one-way behaviors will strengthen the bond between them, resulting in the fame of stars grows. So the strength of one-way interactions between the audience and vloggers can effectively reflect the popularity of vloggers. Therefore, to identify college student vlog celebrities, we use two-step clustering method, in which two indicators: personal attractiveness and audience participation, are considered. Personal attractiveness is expressed in terms of activity, coverage and platform popularity. In Bilibili, Danmaku or live subtitles reflect the user's timely feelings when watching the video, while the comments in the comment area reflect the user's delayed perspective after viewing. They are different and should be set apart. Thus audience participation is represented by three aspects: real-time participation, delayed participation, and degree of recognition. See table 1 for specific attributes.

Table 1. Indicators for identifying college student vlog celebrities

Indicators	Features	Attributes
Personal attractiveness	Activity	Number of posts
	Coverage	Number of followers
	Platform	Number of plays
Audience participation	Real-time participation	Number of Danmaku ^[10]
	Delayed participation	Number of comments
	Degree of recognition	Number of favorites, Number of likes

3.2 Content analysis method

The data of vlogs from the identified college student vlog celebrities will be saved as a txt file. Then with the aid of the word frequency analysis software ROSTCM6, we handle the textual data through word segmentation and word frequency statistics. Then extract high-frequency words for summarizing the vlog hotspots and specific categories of vlogs based on high-frequency word feature analysis. To investigate further, we design an encoding scheme to encode the vlog data to explore the characteristics of vlog content.

4. Experiments and results

4.1 Data collecting and processing

Taking 6 hours as a time period, we crawled vlog data with "College Student vlog" and "University" as the video keywords on Bilibili. Weekly data from February 20, 2020 to February 27, 2020 were collected, a total of 1567 vlogs, containing these following information about the video: Vlog title, Number of comments, Number of plays, Number of favorites, Number of likes, Name of vloggers, Number of posts and Number of followers. After data cleaning and preprocessing through Excel, deleting duplicate data, non-original videos, incomplete data and no data records, 1044 pieces of data were obtained. Various types of data were unified into units, among which number of followers and plays were counted in ten thousand.

4.2 Two-step clustering for vlog celebrity identity

The two-step clustering method was used to identify vlog celebrities from collected vlog data. Two-step clustering is a hierarchical clustering algorithm applied in the field of data mining and multivariate statistics. It is an improved algorithm proposed by scholar Chiu In 2001 based on the BIRCH algorithm^[11]. This algorithm is suitable for clustering of great sample data (> 1000 items) and can process mixed data with continuous and categorical variables. Using SPSS17.0 software to set the 7 attributes in table 1 as the parameters of two-step clustering, and experimenting with the number of clusters from 1 to 15, the results revealed that the clustering effect was the best when the number of clusters was 5. The comparative analysis of these 5 types of user data shows that the characteristics of the 4th type of users, which account for 14.6%, are consistent with the main characteristics of the vlog celebrities mentioned above. The average value of clustering index of various vlog categories is shown in table 2.

Table 2. Results of two-step clustering

Clustering index	Comparison of clustering results				
	Category 1 (27.1%)	Category 2 (23.6%)	Category 3 (17.3%)	Category 4 (14.6%)	Category 5 (17.4%)
Number of posts	324.34	112.67	81.84	113.05	114.53
Number of followers (10,000)	2.86	1.81	1.39	6.00	2.57
Number of plays (10,000)	0.44	0.89	1.78	1.83	1.74
Number of Danmaku	150.38	194.98	139.06	275.09	195.99
Number of comments	218.56	308.69	276.34	427.21	226.83
Number of favorities	256.77	420.39	210.27	410.45	285.55
Number of likes	391.01	645.39	350.56	794.93	693.62

4.3 Content analysis of vlogs

In order to study the characteristics of vlog content from vlog celebrities among college students, we collected 894 vlogs from 153 selected college student vlog celebrities, including "Peng Jiangjiang", "Lam Xiaofang" and so on. Since the explicit content of a vlog can be generally reflected by its video name, we used ROSTCM6 software to segment it. To increase the accuracy of segmentation, it's necessary to edit the custom segmentation table in advance and add noun words which represent the characteristics of vlog content. After counting the frequency of the word segmented files, enabling filtering vocabulary, and deleting quantifiers, prepositions and other meaningless words, final results were outputted into one file. Aiming to reflecting the characteristics of vlog content more reasonably, nouns that belong to the same theme were merged into a new category, which was included in the final word frequency table. Finally, 196 words are obtained. The keywords were sorted according to frequency, and the first 30 keywords were selected as high-frequency keywords. The specific keywords and their frequency statistics are shown in table 3.

Table 3. Words with high frequency (Top 30)

Keywords	Frequency	Keywords	Frequency	Keywords	Frequency
Daily	241	Internship	24	First-time	14

University	163	Advice	23	Semester	14
Study	86	Food	23	Love	14
Experience	54	Fitness	21	Abroad	14
Travel	52	Reading	20	Skincare	13
Work	46	Dormitory	20	Chat	13
Makeup	46	Live	19	Friend	12
Vacation	41	Game	19	Movies	8
Specialist	26	School	17	Homework	8
Exam	26	Sports	16	Plan	6



Figure 1. Vlog hotspots from college student vlog celebrities

Further analysis of vlog content requires coding and classification. There are generally three sources for encoding: data, existing related research and theory. The coding scheme can be established by induction or deduction. In the research without relevant theory as the basis, the coding scheme and data categories need to be summarized from the data itself. Through statistical analysis of all high-frequency words, overall, the content hotspots published by college student vlog celebrities on Bilibili mainly focus on the following themes: Campus/Experience, Learning/Advice, Travel/Entertainment, Work/Internship, and Makeup/Appearance. Setting these themes as 5 first-level indicators of vlog content, 894 vlogs were coded and classified. At the same time, the proportion of vlogs in different themes and the number of celebrity vloggers in it were counted. The results are presented in figure 2. In order to increase the objectivity of the research, after crawling the required vlog data, two enlisted coders pre-encoded before coding to test the coding consistency. 100 vlogs were selected to test the consistency of the two coders' codes and SPSS was applied to perform Cohen's kappa consistency. The resulting kappa value was 0.864, indicating that the consistency was good and the credibility was high.

Table 4 shows the coding results of our research, including the distribution of various vlog duration. The duration of vlog can be regarded as one of its content characteristics. In this paper, the duration of vlogs was divided into different time periods: less than 5minutes, 5-10minutes, 10-15minutes and more than 20 minutes. Meanwhile the average views of vlogs in each category are displayed in figure 3.

Table 4. Encoding results and distribution of video duration

Category	Video duration				
	< 5mins (11.74%)	5mins-10 mins (37.25%)	10mins-15 mins (25.50%)	15mins-20 mins (12.76%)	> 20mins (12.75%)
Campus/Experience (28.52%)	30 (11.76%)	87 (34.12%)	54 (21.18%)	36 (14.12%)	48 (18.82%)
Learning/Advice (25.17%)	12 (5.33%)	84 (37.34%)	84 (37.33%)	27 (12.00%)	18 (8.00%)
Travel/Entertainment (23.49%)	36 (17.14%)	78 (37.14%)	42 (20.00%)	24 (11.43%)	30 (14.29%)
Work/Internship (14.43%)	21 (16.28%)	63 (48.84%)	30 (23.26%)	9 (6.98%)	6 (4.65%)
Makeup/Appearance (8.39%)	6 (3.08%)	21 (9.23%)	18 (32.31%)	18 (27.69%)	12 (27.69%)

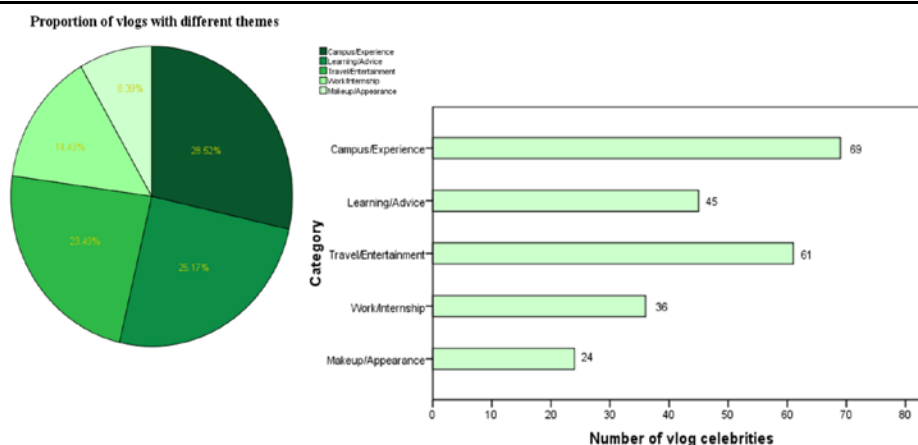


Figure 2. Proportion of vlogs with different themes and number of vlog celebrities in each vlog theme

Here a multi-dimensional line graph helps to visually observe the difference in specific playback volume levels of vlogs of different themes and video duration, as showed in figure 3.

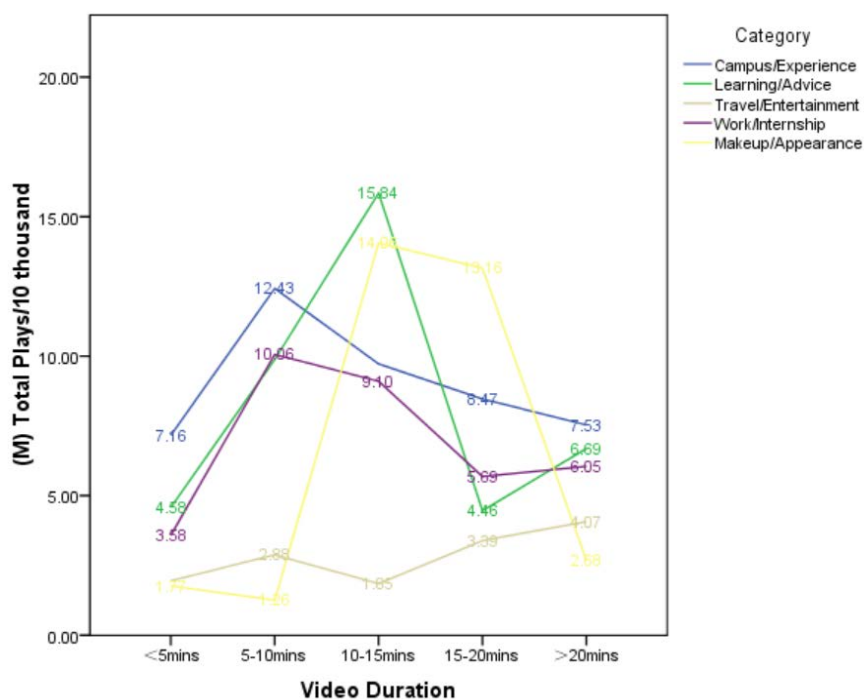


Figure 3. Average playback volumes of vlogs of different themes and video duration(10,000)

4.4 Discussions

As the hotpots in figure 1 reveals, in combination of the vlog coding results in table 4, it can be seen on Bilibili, the popular themes of vlogs published by college student vlog celebrities are widely distributed, which are in accordance with the mental state of college students who feel curious about novelty. However, there still exists differences in the content and form of vlogs in various themes. The top two vlog themes with the largest proportion are closely related to the identity attributes of college students. That is, Campus/Experience (28.52%) and Learning/Advice(25.17%). Statistics show that the proportion of vlogs in which the vlogger mentioned the name of his/her university reaches 44.38%, owning an average playback volume of more than 150,000. In comparison, vlogs of the same theme that do not refer to any information about the vlogger's university have only a playback volume of 64,300. Details of the results are shown in figure 4. Travel/Entertainment vlogs account for 23.49% while Work/Internship vlogs and

Makeup/Appearance vlogs account for 14.43% and 8.39% respectively. By separating vloggers with regarding to these 5 themes, the research also finds that the vast majority of vlog celebrities among college students focus on publishing 1 or 2 types of vlogs. Campus/Experience and Learning/Advice vlog celebrities are the most, Travel/Entertainment and Work/Internship vlog celebrities are the second, and fewer vlog celebrities choose to post Makeup/Appearance vlogs.

Name of university is mentioned or not

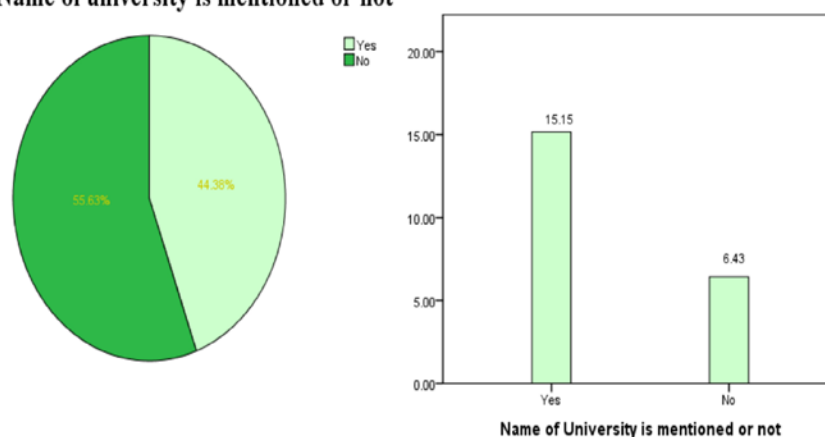


Figure 4. Name of university is mentioned or not

The popularity of various types of vlogs can be reflected by figure 3 showing the average playback volume of each theme in different video duration. Among short vlogs lasting 0-10 minutes, Campus/Experience vlogs are the most popular, while long vlogs lasting 10-20 minutes with themes about Learning/Study and Makeup/Fitness gain relatively high popularity. When video duration exceeds 20 minutes, the average playback volume of various vlogs is at a low level, barely reaching 80,000 plays. It is interesting to find that Travel/Entertainment vlogs, which account for a large proportion (23.49%), have been viewed the least on average, floating within limits between 30,000 and 40,000 plays. This reveals that the influence and popularity of Travel/Entertainment vlogs fail to coincide with the considerable proportion it occupies in Bilibili.

5. Conclusions and suggestions

In our research, PSI theory was applied to distinguish the college student vlog celebrities. The proposed identification indicators are from two aspects: personal attractiveness and audience participation. After that the two-step clustering method was used to identify the required vlog celebrities and accordingly vlog data from them were collected for content analysis. Through word frequency analysis, five themes of vlogs were summarized while distribution and average playback volume

of each theme coded in coding mode were observed. To further investigate the characteristics of vlog content from the selected vlog celebrities, the average playback volume of vlogs in each theme is analyzed with different video duration. The following conclusions can be drawn from the analysis and results in part 4:

(1) College student vlog celebrities tend to shoot vlogs in themes that match their identity attributes. According to the classification of vlog content, the spot vlog content of vlog celebrities among college students revolve around the college students' concerns like daily occurrences and life experience. There are a variety of themes, among which vlogs about campus life and learning experience that are consistent with the identity attributes of college students occupy a proportion of more than 50%, being the most popular with viewers. It's probably because that followers of vlog celebrities mostly are in their studentship, thus watching vlogs with themes related to campus life from their adored vloggers will psychologically shorten the distance between them. In other words, the similarity followers feel out of these vlogs will benefit enhancing the relationship between vlogger and its followers. With this conclusion, it's reasonable that vlog on themes like traveling and entertainment are less attractive, even though they've occupied a large proportion. Because the video content makes followers harder to resonate with the vlogger.

(2) Homogenization problem of vlog content exists in the same vlog theme. It hinders the development of these vlogs and reflects a disappearance of personal characteristics of vloggers on Bilibili. Through word frequency analysis and the average playback volume of vlogs in various themes, it can be found that there is a lack of differentiation in vlogs with same themes in which the vlog titles used are highly similar in terms of words as well as the shooting methods, thus causing a relatively low average playback volume in this theme.

(3) Vlogs on themes about campus life have become a means of publicizing universities. Not only does a vlog show the life details of the vlogger, it also serves as an effective form of propaganda for universities. The research observes that vlogs presenting the name of the university in the title is much more popular than vlogs that simply shoot the campus daily without disclosing any information about that. The reason is that knowing the vlogger's university helps increase the authenticity of the vlogger's identity. Usually the more authentic the vlog is, the more popular it can be. On the flip side, showing the university's name in a vlog will increase the popularity of the university in some degree, for it can demonstrate the unique and beautiful scenery in college through the vlogger's camera. It has become a way of propagating out of the blunt publicity of the official university website, but from a more natural and vivid recording perspective.

In this paper, we conducted content analysis and research on the specific group of college student vlog celebrities on Bilibili. The combination of theory and quantitative tools shows some innovations, but there are still limitations and deficiencies that need to be further improved: First, the coding standard needs continuous improvements, while the subjective factors of coding category setting are relatively comprehensive, and it is still impossible to summarize and refine all vlog themes. Second, the number of collected samples needs to be further expanded, and

the generality of the conclusion needs to be tested. Future research can expand the amount of vlog data and test the rationality of the conclusions, in the hope of providing theoretical guidance for the development of vlog formats and content innovation.

Reference

- [1] Tencent's "Generation Z" and their Bilibili[EB/OL].[2019-04-02]. <https://new.qq.com/omn/20190228/20190228A13SG5.html>.
- [2] HORTON D,WOHL R R. Mass communication and para-social interaction; observations on intimacy at a distance.[J]. *Psychiatry*,1956,19(3):188-211.
- [3] Rubin,A.M.,Perse,E.M.&Powell,R.A.(1985).Loneliness,parasocial interaction,and local television news viewing.*Human Communication Research*,12(2):155-180.
- [4] Song J,George M,Determinants of Perceived WebSite Interactivity[J].*Journal of Marketing*,2008,72:99-113.
- [5] Trim Cole. Attachment Styles and Intimate Television Viewing: Insecurely Forming Relationships in a Parasocial Way[J]. *Journal of Social and Personal Relationships*,1999,16(4).
- [6] BA0 Dunan et al,Parasocial Interaction between Browser and Poster in Virtual Communities: An Empirical Study on dianping.com[J]. *Journal of Management*, 2011, 8 (07): 1010-1020.
- [7] John R. Turner. Interpersonal and psychological predictors of parasocial interaction with different television performers[J]. *Communication Quarterly*, 1993, 41(4).
- [8] Hall J,et al. Improving the Understanding of Parasocial Interaction:A review of its effects, conceptualizations, and antecedents[J]. *NCA 93rd Annual Convention*, 2007.
- [9] Jaideep Sengupta, Gavan J. Fitzsimons. The Effects of Analyzing Reasons for Brand Preferences: Disruption or Reinforcement?. 2000, 37(3):318-330.
- [10] JIANG Mengjiao. Changes in Users' Demands of Video Website—From the Study of Danmaku Video[D]. *Jilin University*, 2017.
- [11] Chiu T, Fang D, Chen J, et al. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment[C]//*Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2001.