



# Assessing the Content Topics of the Educational Videos on Tik Tok for Science Communication

Pengqi Wang<sup>1</sup>, Mengli Yu<sup>2</sup>, and Yadong Liu<sup>2</sup>(✉)

<sup>1</sup> Faculty of Business and Management, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

<sup>2</sup> School of Journalism and Communication, Nankai University, Tianjin, China  
yadongli@sina.com

**Abstract.** Short video platforms like Tik Tok are becoming an increasingly crucial educational platform that fosters learning on demand with educational science videos. Educational videos are perceived as a productive strategy to enhance the users' scientific knowledge and even are applied in some academic scenarios. This study provides an overview of science communication videos on Tik Tok information management and assesses the content topics based on user engagement, time, creator, and textual features. Seven main content topics are identified: Healthcare, Trivia, Nature, Astronomy, Social Science, Engineering, and Agriculture. Further findings demonstrate that topics vary significantly in terms of user engagement features: Topics on Healthcare and Trivia receive more likes, comments, and shares, whereas Social Science, Engineering, and Agriculture attract fewer engagements. Moreover, videos on the Healthcare topic have shorter durations, more followers, more concise video titles, and a smaller number of hashtags. We further elaborated on each topic, combined with the existing literature on related fields of information management.

**Keywords:** Clustering algorithm · Empirical research · Information management · Science communication

## 1 Introduction

Educational videos and films have a long heritage in conveying complicated information in kindergartens, schools, and higher education. However, concerning educational videos of science communication, it is traditionally conquered by professional communicators through mainstream media channels like television and movies [1]. With the emergence of Web2.0, short video platforms (e.g., Tik Tok) have reshaped the educational science communication environment. By eliminating the obstacles for content creators to access an audience, short video platforms offer an alternative to traditional content distribution, creating new channels for popularizing educational scientific knowledge [2]. As the current representative platform, Tik Tok encourages users to create and view educational scientific knowledge short videos on mobile devices. As per a statistical report by ByteDance [3], science communication videos account for 20% of total views on Tik

Tok. The emerging large number of science videos deliver a challenge for information management and topics' recommendation on short video platforms.

Previous research focused on social media platforms (e.g., Facebook and Twitter) and attested to some critical features of the specific topics of science communications [4–6]. Könneker & Lugger [7] alleged that the advancement of the Internet allows scientists and science communicators to discuss scientific topics with multiple audiences with different perspectives by using social media platforms. However, significant gaps exist in research on the content types and communication features of scientific knowledge dissemination on short video platforms [8]. Compared with traditional online video platforms, the number of videos on the short video platform is more significant, the length of content is more disparate, and the distribution mechanism is decentralized [9]. Due to the low threshold of the participation of the common public, the content topics delivered by the short videos are vastly diversified [10]. These features make existing research relatively inapplicable for this novel communication platform.

In this study, we attempt to address these questions:

- (1) What content topics can be identified from science communication videos on Tik Tok information management?
- (2) How do videos of different content topics differ in user engagement features?
- (3) How do different content topics interact with other video features, including time, creator, and textual features?

In answering the preceding research questions, we employed the clustering method and VOSviewer tool to determine seven main content topics. To investigate the topic-level differences, we conducted ANOVA and post-hoc tests to test topics' differences in user engagement and video features.

## 2 Research Method

### 2.1 Data Procurement and Processing

This study collected data from Douyin (the Chinese version of Tik Tok). To narrow the field of study to science communication and popularization videos, we use the search keywords “Ke Pu” (meaning science communication in Chinese) to filter the videos. Data were collected for three months from October 2021 to December 2022. By removing duplicated entries and these empty entries, we finally collected 19,000 entries. The collection of video data mainly includes basic information (e.g., title and hashtags), creator information (e.g., nickname), and video popularity metrics (e.g., likes and comments).

### 2.2 Video Clustering

This study employed VOSviewer to identify categories of the video data from the co-occurrence data. Specifically, by artificially parsing this visual network, we determined the main clusters of the video. Then we categorized the videos based on the Challenge and Hashtags fields in the video information coding. Here, Challenge is the topic of the video

**Table 1.** Content topics in science communication videos on Douyin (N = 5848)

Category	<i>n</i>	Highlighted hashtag words
Healthcare	1663	medicine; Chinese medicine; cardiovascular; skincare; medical beauty; tumor; hypertension; healthy diet; pregnancy; women;
Trivia	1536	unsolved mystery; life tips; interesting knowledge; cold knowledge; curiosity; human body; plane; UFO; incredible; culture
Nature	792	wild animals; marine life; lions; antelopes; plants; insects; dinosaurs; nature offerings; fishing; mystery
Astronomy	737	universe; space; moon; solar system; planet; aerospace; meteorite; alien; space station; milky way
Social Science	535	insurance; law; nation; social security; Chinese history; geography; infrastructure; pandemic; China; finance
Engineering	435	manufacturing; hydraulic press; craftsmanship; manufacturing process; original technology; laborer; the beauty of the utensils; mechanical principle; magnet; creative
Agriculture	150	agricultural planting; farmer; rural; grape; farming period; agricultural technology; weeding; rice cultivation; seed; grain

interpreted by the user or the platform. Hashtags are some keywords related to video content and the topic. The unsupervised learning K-means algorithm was employed to automate the clustering of clusters, which makes the classification more accurate and helps us to find outliers [11]. As some of the video hashtags were vaguely defined or too scattered for us to define the topics accurately, we discarded the data for these videos and kept only the precisely defined ones. Finally, we defined 5848 clustered videos and developed a systematic description based on the seven clusters.

### 3 Results and Discussion

#### 3.1 Overview of Content Topic Clustering

The keyword co-occurrence network generated by VOSviewer is shown in Fig. 1a and its translated version is shown in Fig. 1b. Different clusters are labelled with different colors, and the size of the nodes indicates the number of occurrences of this term, while the thickness of the connecting lines between the nodes represents the strength of the association between two nodes. The co-occurrence network and clustering output define seven categories: Healthcare, Trivia, Nature, Astronomy, Social Science, Engineering, and Agriculture. As shown in Table 1, we highlighted ten hashtag words with the highest link strength for each category. These words represent the content with the highest topical associations for each category.

#### 3.2 Topical Differences in User Engagement Features

User engagements represent how the public recognizes and interacts with the video under specific content topics. The number of likes could reflect how the general public



**Table 2.** Means and standard deviations of user engagement metrics (N = 5848)

Content topic	n	Likes		Comments		Shares	
		Mean	SD	Mean	SD	Mean	SD
Healthcare	1663	19745.55	133782.37	944.03	5639.27	3068.63	14028.91
Trivia	1536	31650.14	148247.70	1439.13	7723.31	2278.66	15386.93
Nature	792	4513.66	31424.82	253.30	1690.98	466.77	3439.88
Astronomy	737	17872.99	120660.20	611.75	3648.80	618.13	4221.38
Social Science	535	2662.17	11928.34	237.06	2149.91	728.40	4752.15
Engineering	435	4910.19	34019.08	294.17	3558.87	219.79	1538.96
Agriculture	150	3677.33	12292.15	200.75	1069.89	1802.71	11589.29
Total	5848	17494.95	114191.80	806.57	5326.22	1741.47	11350.13

many non-professional health content consumers prefer to get their knowledge through online support groups like bulletin boards and discussion groups. People need to share the video in their chat groups to involve the information from the videos in the discussion. Moreover, since the knowledge delivered by Trivia is compelling and interesting enough, the user engagement index for this video category also stands out. We can see that this type of content's likes, comments, and shares are well above average. Trivia has been shown in business case studies to improve engagement rates [17]. It is believed that trivia information can improve the user experience when it comes to entity searches; even a minor improvement in this type of search can significantly improve user engagement [18].

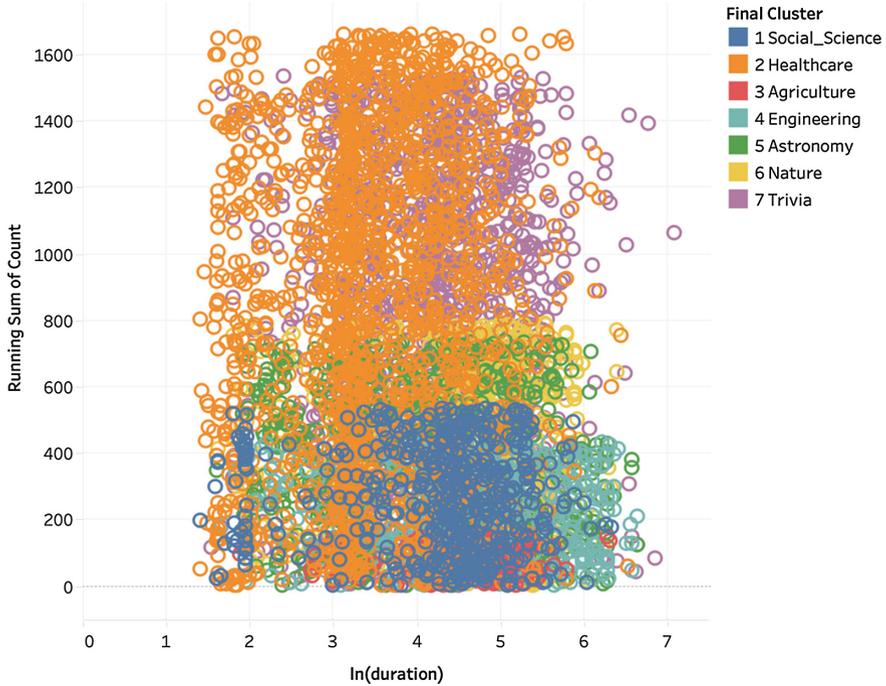
### 3.3 Topical Differences in Time, Creator, and Textual Features

We further analyze the topical differences in time (i.e., duration), creator (i.e., followers), and textual features (i.e., title length and hashtags). In that, duration represents the length of the video when videos can fully express details. Followers reflect the video creator's influence on this platform. Textual features play a key role in searching and recommending for users on the platform. Table 3 summarizes the mean values of duration (mean = 78.93), followers (mean = 845916.10), title length (mean = 21.50), and number of hashtags (mean = 3.50). As the standard deviations of duration (SD = 100.88) and the number of followers (SD = 2445908.49) is excessively high, we took their natural logarithms in the analysis and used their processed values in the following analysis. We performed one-way ANOVA and Tamhane post-hoc tests to examine how each metric's seven content topics interact. Based on the results of ANOVA, all the metrics ( $p = 0.000$ ) are significant at a 95% level, revealing differences among the seven content topic groups.

Overall, the average duration of science communication videos (mean = 97.03) is higher than the intermediate level of 40 s for Tik Tok videos [19]. Figure 2 shows the distribution of the duration after logarithmic transformation. The post-hoc test reveals that videos under Healthcare and Nature topics have significant differences from other

**Table 3.** Means and standard deviations of time, creator, and textual metrics (N = 5848)

Content topic	n	Duration		LN (Duration)		Followers		LN (Follower)		Title length		# of hashtags	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Healthcare	1663	52.05	59.54	3.51	0.95	1285204.84	3093095.40	12.40	2.64	19.77	14.94	2.66	1.51
Trivia	1536	94.56	94.03	4.19	0.88	1167649.83	3025392.31	11.26	3.48	23.79	10.76	3.36	1.39
Nature	792	133.60	93.37	4.57	0.95	373552.66	1626293.91	10.89	2.49	21.61	16.73	4.52	1.96
Astronomy	737	113.16	125.08	4.17	1.12	489042.46	1391604.69	10.12	3.58	24.59	16.63	4.02	2.10
Social Science	535	103.63	76.47	4.31	0.96	362717.34	783823.67	10.57	3.04	20.01	12.20	3.07	1.60
Engineering	435	165.86	157.37	4.63	1.07	261030.79	481135.86	11.04	2.40	19.79	11.30	4.65	2.45
Agriculture	150	125.52	78.93	4.62	0.73	348209.26	380727.68	12.15	1.58	11.63	11.17	4.41	1.51
Total	5848	97.03	100.88	4.10	1.04	845916.10	2445908.49	11.34	3.09	21.50	14.11	3.50	1.87



**Fig. 2.** Distribution of natural logarithm of video length ( $\ln\_duration$ ), colored by content topics

groups in terms of length of time ( $p = 0.000$ ). It can be concluded that the videos' duration of Healthcare (mean = 52.05) is the shortest topic, and the duration of Nature (mean = 133.60) is the most extended topic. Suppose the topic is related to a complex process, such as topics of Engineering (mean = 165.86) and Agriculture (mean = 125.52), or relatively professional background knowledge, such as the topic of Nature (mean = 133.60). In that case, the duration is extended for more elaboration. Moreover, it is worth noting that videos under the Healthcare topics, which are proven to have the highest user engagement features, are of the least duration among those topics. This conclusion is consistent with the findings of Zhu et al. [20], which reveal that the general public prefers shorter videos.

The video creator's followers for Healthcare (mean = 1285204.84) and Agriculture (mean = 348209.26) are significantly higher than the other topics ( $p = 0.000$ ), while the videos on the Astronomy (mean = 489042.46) topic are of the least mean value. We conclude that the audiences of the videos under the Agriculture topic are not the people who are accustomed to using the Internet. But the number of users who watch Agriculture videos is significant. Therefore, such data feature emerges. Because the content of the Astronomy theme is the furthest away from people's lives, people pay relatively little attention to this topic. This further illustrates that Tik Tok reaches a larger audience in the lower market [21].

In terms of title length, the videos on Astronomy (mean = 24.59) and Trivia (mean = 23.79) topics are of significantly longer titles than other topic groups ( $p = 0.000$ ),

while the videos on Agriculture (mean = 11.63) topics are of the shortest titles on average. The most apparent purpose of the video title is to attract the users' attention and concisely elaborate on the main content of the video. The topics of Astronomy and Trivia are somewhat mysterious, and therefore longer titles are utilized to be eye-catching. Agriculture, covering the knowledge of some agricultural production techniques and precautions, is relatively straightforward and therefore doesn't need lengthy descriptions. In terms of the number of hashtags, the videos on the Engineering (mean = 4.65) and Nature (mean = 4.52) topics are of significantly more hashtags than other topic groups ( $p = 0.000$ ), while the videos on Healthcare (mean = 2.66) topics are of the least hashtags. The number of hashtags reveals the concentration of the topics covered in the video. Videos under the Engineering and Nature topics may contain multiple contents, such as subcategories and comprehensive knowledge, so they have more hashtags. Videos under the Healthcare topic usually focus on a single point of expertise and therefore have fewer hashtags.

## 4 Conclusion

This paper aimed to assess the content topics and the potentials of educational videos on Tik Tok for science communication. Educational short videos contribute to public scientific information demand in fragment time. This research focused on the Tik Tok platform and conducted empirical reviews based on 5848 educational videos. We categorized the science communication videos into seven classifications: Healthcare, Trivia, Nature, Astronomy, Social Science, Engineering, and Agriculture. ANOVA and post-hoc tests are then conducted based on the metrics procured from the preprocess of the raw data. Findings demonstrate that topics have significant differences in user engagement (i.e., likes, comments, and shares) and other short video features (i.e., duration, followers, hashtags, and title length). In that, Healthcare and Trivia show higher attractiveness than other topics. In addition, we argue that the difference in topics' textual features and duration reflect these topics' profession and content expression. By answering the proposed questions, this study provides a content framework for future research on video-based science communication. By revealing some topics' more significant impact over others, it contributes to the platform's information management through optimizing topics' recommendation and educational videos' distribution.

Furthermore, the results provide an understanding of the perception of educational scientific knowledge among Chinese audiences and empirical evidence of how time, creator, and textual features interact with videos under specific topics. In the future, regression analysis will be conducted to explore the relationship between the popularity of science communication videos and the topic features, textual features, etc. We will keep focusing on the content topics and digging more dissemination characteristics of educational videos to improve the efficiency of information management.

**Acknowledgments.** This work was supported by the Tianjin Philosophy Social Science Project [Grant No. TJXC21-007], Liberal Arts Development Foundation of Nankai University [Grant No. ZB21BZ0333], and Fundamental research fees for central universities [Grant No. 63222203].

## References

1. D. Brossard, “New media landscapes and the science information consumer,” *Proceedings of the National Academy of Sciences*, vol. 110, no. Supplement 3, pp. 14096–14101, 2013, doi: <https://doi.org/10.1073/pnas.1212744110>.
2. Juhasz, “Learning the five lessons of YouTube: After trying to teach there, I don’t believe the hype,” *Cinema Journal*, vol. 48, no. 2, pp. 145–150, 2009, doi: <https://doi.org/10.1353/cj.0.0098>.
3. ByteDance, “2021 Tik Tok Knowledge Content Data Report,” Cyberspace Administration of China, Oct. 2021. [Online]. Available: [http://news.xhby.net/tuijian/202110/t20211013\\_7265103.shtml](http://news.xhby.net/tuijian/202110/t20211013_7265103.shtml)
4. K. Chirumalla, P. Oghazi, and V. Parida, “Social media engagement strategy: Investigation of marketing and R&D interfaces in manufacturing industry,” *Industrial Marketing Management*, vol. 74, pp. 138–149, Oct. 2018, doi: <https://doi.org/10.1016/j.indmarman.2017.10.001>.
5. N. Hara, J. Abbazio, and K. Perkins, “An emerging form of public engagement with science: Ask Me Anything (AMA) sessions on Reddit r/science,” *PLoS ONE*, vol. 14, no. 5, p. e0216789, May 2019, doi: <https://doi.org/10.1371/journal.pone.0216789>.
6. D. J. Welbourne and W. J. Grant, “Science communication on YouTube: Factors that affect channel and video popularity,” *Public Underst Sci*, vol. 25, no. 6, pp. 706–718, Aug. 2016, doi: <https://doi.org/10.1177/0963662515572068>.
7. C. Könniker and B. Lugger, “Public Science 2.0—Back to the Future,” *Science*, vol. 342, no. 6154, pp. 49–50, Oct. 2013, doi: <https://doi.org/10.1126/science.1245848>.
8. Z. Chen, Q. He, Z. Mao, H.-M. Chung, and S. Maharjan, “A study on the characteristics of douyin short videos and implications for edge caching,” in *Proceedings of the ACM Turing Celebration Conference - China*, Chengdu China, May 2019, pp. 1–6. doi: <https://doi.org/10.1145/3321408.3323082>.
9. X. Cheng, J. Liu, and C. Dale, “Understanding the Characteristics of Internet Short Video Sharing: A YouTube-Based Measurement Study,” *IEEE Transactions on Multimedia*, vol. 15, no. 5, pp. 1184–1194, 2013, doi: <https://doi.org/10.1109/TMM.2013.2265531>.
10. Y.-H. Wang, T.-J. Gu, and S.-Y. Wang, “Causes and characteristics of short video platform internet community taking the TikTok short video application as an example,” in *2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, 2019, pp. 1–2.
11. S. Pavelle and C. Wilkinson, “Into the Digital Wild: Utilizing Twitter, Instagram, YouTube, and Facebook for Effective Science and Environmental Communication,” *Front. Commun.*, vol. 5, p. 575122, Oct. 2020, doi: <https://doi.org/10.3389/fcomm.2020.575122>.
12. S. Kay, R. Mulcahy, and J. Parkinson, “When less is more: the impact of macro and micro social media influencers’ disclosure,” *Journal of Marketing Management*, vol. 36, no. 3–4, pp. 248–278, Feb. 2020, doi: <https://doi.org/10.1080/0267257X.2020.1718740>.
13. E. Katz, J. G. Blumler, and M. Gurevitch, “Uses and Gratifications Research,” *Public Opinion Quarterly*, vol. 37, no. 4, p. 509, 1973, doi: <https://doi.org/10.1086/268109>.
14. L. de Vries, S. Gensler, and P. S. H. Leeflang, “Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing,” *Journal of Interactive Marketing*, vol. 26, no. 2, pp. 83–91, May 2012, doi: <https://doi.org/10.1016/j.intmar.2012.01.003>.
15. C. Munaro, R. Barcelos, E. C. Francisco Maffezzolli, J. P. Rodrigues, and E. Paraiso, “To engage or not engage? The features of video content on YouTube affecting digital consumer engagement,” *J Consumer Behav*, vol. 20, no. 5, pp. 1336–1352, Sep. 2021, doi: <https://doi.org/10.1002/cb.1939>.

16. M. Wikgren, "Health discussions on the Internet," *Library & Information Science Research*, vol. 23, no. 4, pp. 305–317, Dec. 2001, doi: [https://doi.org/10.1016/S0740-8188\(01\)00091-3](https://doi.org/10.1016/S0740-8188(01)00091-3).
17. D. Tsurel, D. Pelleg, I. Guy, and D. Shahaf, "Fun Facts: Automatic Trivia Fact Extraction from Wikipedia," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, Cambridge United Kingdom, Feb. 2017, pp. 345–354. doi: <https://doi.org/10.1145/3018661.3018709>.
18. H. Bota, K. Zhou, and J. M. Jose, "Playing your cards right: The effect of entity cards on search behaviour and workload," in *Proceedings of the 2016 acm on conference on human information interaction and retrieval*, 2016, pp. 131–140. doi: <https://doi.org/10.1145/2854946.2854967>.
19. Q. Chen, C. Min, W. Zhang, X. Ma, and R. Evans, "Factors Driving Citizen Engagement With Government TikTok Accounts During the COVID-19 Pandemic: Model Development and Analysis," *J Med Internet Res*, vol. 23, no. 2, p. e21463, Feb. 2021, doi: <https://doi.org/10.2196/21463>.
20. C. Zhu, X. Xu, W. Zhang, J. Chen, and R. Evans, "How Health Communication via Tik Tok Makes a Difference: A Content Analysis of Tik Tok Accounts Run by Chinese Provincial Health Committees," *Int. J. Environ. Res. Public Health*, vol. 17, no. 1, p. 192, Jan. 2020.
21. X. Pan, E. Yan, and W. Hua, "Science Communication and Dissemination in Different Cultures: An Analysis of the Audience for TED Videos in China and Abroad," *JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY*, vol. 67, no. 6. WILEY, 111 RIVER ST, HOBOKEN 07030–5774, NJ USA, pp. 1473–1486, Jun. 2016.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

