

Empirical Examination of Pet-Themed Short Video Popularity and Power-Law Patterns on Douyin

Wancheng Liu

*Yiwu International Academy, Yiwu, China
19548803420@163.com*

Abstract. Although short-video platforms disseminate content constantly, the resulting spread is far from uniform, since a limited set of videos rapidly attract disproportionate attention while the overwhelming majority receive only marginal engagement, a pattern that researchers commonly describe as a video 'going viral.' Because prior studies have largely relied on models or simulations to explain this phenomenon, empirical verification using real-world data remains comparatively rare. To address this gap, the present study examines fifty short video segments about animals that circulated on Douyin between July 26 and August 8, 2026, using likes and comments as indicators of dissemination. A log-log analysis was conducted to determine whether the extent of this spread conforms to a power-law distribution, in which a small number of elements account for most of the total while the majority contribute only marginally, and a Gini index was calculated to capture how concentrated that distribution is, alongside a comparison between video content category and dissemination intensity. The results reveal a pronounced long-tailed distribution: the power-law model fit the data closely, yielding an R^2 of 0.927, while the top 20% of videos accounted for 80% of all likes and the Gini index reached 0.745. Because footage from verified or prominent accounts attracted substantially more likes than that produced by ordinary content creators, these findings lend support to the Matthew effect as it operates within social media diffusion.

Keywords: short video, power-law distribution, Gini coefficient, pet-themed content, Matthew effect

1. Introduction

Because short-video platforms allow information to disseminate swiftly while offering individuals a convenient means of participation, they have become a vital channel for content spread across social networks. Previous studies have identified a distinct pattern in this dissemination [1], whereby a small number of videos rapidly attract a large audience while most content reaches only a limited number of viewers, a dynamic that carries real significance for daily life since it helps explain why certain videos gain sudden popularity while the vast majority remain unnoticed. Experts describe this pattern as a power-law distribution, in which an exponential relationship links a given value to its rank, so that a small number of elements account for a disproportionate share of the whole [2, 3].

Although numerous recent studies have examined this trend through models or simulations [1, 4], comparatively few have investigated it using real-world data.

This study examines short video clips featuring pets on Douyin, which ranks among the premier platforms for short-form video content, sorting the samples by content genre and by the type of creator account while using likes and comments as indicators of dissemination. Beyond testing whether the extent of spread conforms to a power-law distribution, the analysis assesses how much the most-watched videos contribute to overall dissemination through a log-log regression combined with the calculation of a Gini coefficient [5], and it further explores how account classification and content category relate to the degree of spread.

By offering real-life data that reinforces ideas which had previously relied heavily on simulations, this research provides tangible evidence of how content spreads atypically across social media networks, evidence that may help content producers understand how their material disseminates while also benefiting platform analysts.

2. Methods

2.1. Data source and sampling

The sample comprises short videos with pet-related topics sourced from Douyin and posted between July 26 and August 8, 2026, a window chosen deliberately so that every video would have at least five days to accumulate likes, thereby minimizing the bias that disparities in posting time might otherwise introduce. Because the sample was generated through systematic sampling, in which clips were selected at regular intervals from the full list, it encompasses a wide range of videos—both widely recognized and relatively obscure—rather than being limited to the most favored options, yielding a final sample of 50 video segments.

Although the number of views is frequently used as an indicator of spread in other research, it was not applied here because Douyin generally does not display view counts for most of its video snippets; the primary indicator instead was likes and comments, which every user can view and which can be documented in the same way for each video, making them a trustworthy choice for this research.

Each clip was coded according to several basic features, including the number of pets displayed, the use of well-known music, and the style of captions shown, and was assigned to one of four content categories—ordinary footage, skill or display, unexpected or contrast-based appeal, and other material—following a distinct coding scheme. Creator accounts, meanwhile, were split into two groups, ordinary creators and authenticated or institutional accounts, a distinction drawn from follower counts and verification status.

2.2. Statistical methods

The power-law model was tested through a log-based regression [5], in which the 50 video clips were first ranked by their number of likes, from largest to smallest, before the logarithms of both likes and ranking were computed and subjected to a linear regression; the resulting R^2 value indicated how closely the data adhered to a linear trend, with a high R^2 providing strong support for the power-law pattern.

To assess how concentrated the distribution of likes was, a Gini coefficient was calculated, a value ranging from 0, where likes are distributed uniformly, to 1, where a single video accounts for every like, so that a higher value indicates a less evenly balanced distribution.

A Pearson correlation coefficient was also computed to examine the association between the number of days elapsed since a video was posted and the number of likes it received, since videos posted earlier could plausibly accumulate more likes over time; this assessment aimed to determine whether such a time-related bias had actually influenced the results.

3. Results and analysis

3.1. Descriptive statistics

Among the 50 video segments, the average number of likes reached 29,071, whereas the median stood at only 7,329, a gap that offers an initial sign of an uneven distribution, since a more uniformly spread pattern would show a mean and median closer in value—a resemblance that does not hold here and that instead points toward the power-law distribution examined in the following section. Because the highest number of likes recorded was 293,575 while the lowest was only 572, an increase exceeding 500-fold, Figure 1 makes clear that most videos cluster at the lower end of the range while only a small number achieve an extremely high number of likes, a pattern characteristic of a long-tail model.

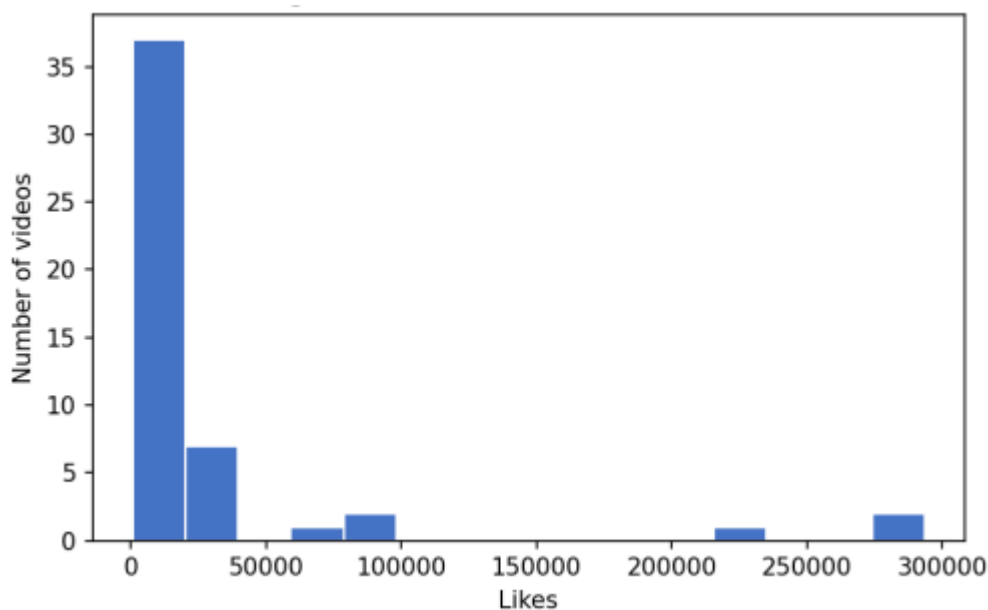


Figure 1. Distribution of likes (n=50)

3.2. Power-law distribution fitting

After the video segments were ranked by their number of likes in decreasing order, the logarithm of ranking and likes was plotted, as shown in Figure 2, where the resulting dots form a pattern that closely resembles a straight line; because the regression line fits the data closely, yielding an R^2 value of 0.927, this outcome validates the power-law tendency and aligns with the early signs already identified in the descriptive analysis.

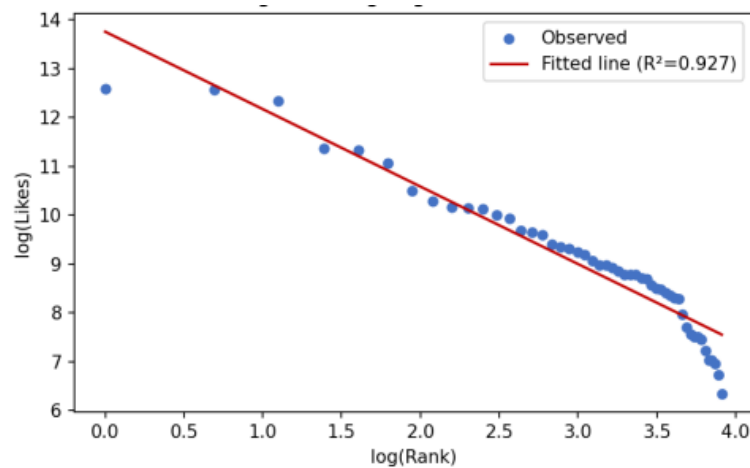


Figure 2. Log-log rank-size plot

3.3. Gini coefficient and concentration of diffusion

The Gini coefficient calculated here reached 0.745, a figure that significantly exceeds the 0.3 to 0.5 range typical of income Gini coefficients in many countries, suggesting that attention on this platform is distributed even less evenly than income in most nations, since it is concentrated within a limited group of videos rather than spread broadly. Because the top 20% of videos, ten in total, account for 80% of all likes, this finding gives a concrete value to the so-called 80/20 rule and aligns with the pattern already presented in Figures 1 and 2.

3.4. Supplementary analysis

When the sample was classified by creator-account type, verified or institutional accounts averaged 72,472.5 likes, roughly four times the 18,220.4 likes averaged by ordinary content producers, a disparity suggesting that an account's history can independently shape how far a video spreads, quite apart from the content itself.

When the data were categorized by content type, videos that captivated viewers through unexpected or contrasting appeal attracted the most attention, averaging 37,082.3 likes, followed by ordinary video snippets at 29,148.6 likes and other material at roughly 24,037.9 likes, while talent and performance clips averaged only 18,567.5 likes, the lowest figure recorded. Although each group contained a limited number of videos, this trend is consistent with the straightforward idea that content combining originality with emotional appeal tends to perform well.

Because videos that received more likes generally also attracted more comments, a notable connection emerged between the two measures in this dataset, suggesting that widely viewed videos tend to perform well across multiple dimensions rather than a single one. When the relationship between days elapsed since posting and number of likes was examined via Pearson correlation, the result came to only 0.11, a value close enough to zero to suggest that the timing of a post's release did not meaningfully influence how many likes it ultimately received, an outcome indicating that the time frame used in this research did not introduce significant bias.

4. Discussion

The finding that a limited selection of animal-focused clips attracts the lion's share of attention can be understood by examining how the platform's recommendation mechanism operates: a newly

made clip is initially shown to only a limited group of viewers, and only when it achieves a high level of interaction—such as a significant like rate or a high completion rate—does the platform expand its exposure to a wider audience, whereas a clip that receives little interaction simply remains at that early stage. This pattern resembles the rapid, erratic growth documented in various digital media studies [6].

Because authenticated individuals and organizations already possess an established readership, they tend to pass through the platform's initial review phase more easily, since their existing audience boosts interaction from the outset [7] and this initial advantage then allows them to reach a considerably wider readership, a pattern that accords with trends observed in numerous large-scale diffusion studies [8]. Given that official and verified profiles in this study received roughly four times as many likes as regular users, the finding reinforces the Matthew effect, whereby content that has already sparked interest tends to continue attracting additional attention [9].

These findings carry a practical implication: because a content producer with few followers may struggle to reach a large audience for reasons that extend beyond content quality alone—chiefly, the difficulty of moving past the platform's initial exploratory phase—this raises a question that platform researchers might examine more closely, particularly regarding how the recommendation system shapes visibility across different account categories, a question that lies beyond the scope of this study yet merits further attention.

Several limitations should be acknowledged. Because the dataset contains only 50 videos and is restricted to a single category, clips related to animals, whether these findings would apply to other types of content remains uncertain. This study likewise did not examine how likes accumulated over time following a post, even though prior research has found disparities in the speed at which different kinds of content spread [10]; subsequent studies could draw on time-series approaches to investigate this dimension.

5. Conclusion

This study pursued two main aims: determining whether the scale of dissemination adheres to a power-law distribution, and ascertaining what portion of that spread the top-ranking videos account for. The sample of 50 videos provides a clear answer to both questions, since likes follow a power-law distribution with an excellent correlation coefficient of $R^2 = 0.927$, while the top 20% of videos account for 80% of all likes, demonstrating that attention is concentrated rather than evenly dispersed—a conclusion reinforced by a Gini coefficient of 0.745 that surpasses the income-inequality levels found in many countries.

Additional results show that verified or institutionally linked accounts receive roughly four times as many likes as regular creators, while among content types, videos that captivate viewers with unexpected twists or striking contrasts tend to achieve the highest average number of likes, implying that both account background and content type play a significant role. Because this pattern appears to follow a systematic structure that aligns with the concept of accumulated advantage found in prior diffusion research, the study offers a practical, data-grounded example drawn from a widely recognized content genre on a leading Chinese platform.

This study also faces certain limitations, since its sample of 50 videos remains limited in size and its focus on pet-themed content means the findings may not directly apply to other kinds of videos. Future research could address this by expanding the sample and exploring additional categories, such as food-related or fashion-type content, while incorporating time-associated data to track how likes accumulate after a post, an approach that could yield a deeper understanding of the mechanism behind a video's rise to virality.

References

- [1] Dong, J., Chen, B., Liu, L., Ai, C., Zhang, F., & Qiu, X. (2018). Impact of attractiveness and influence of information on cascade size distribution. *Journal of System Simulation*, 30(10), 3624–3631.
- [2] Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351.
- [3] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.
- [4] Li, F., & Wei, Y. (2020). Research on evolution model of online public opinion based on temporal network. *Journal of System Simulation*, 32(3), 394–403.
- [5] Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51(4), 661–703.
- [6] Yang, J., & Leskovec, J. (2011). Patterns of temporal variation in online media. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 177–186.
- [7] Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: Quantifying influence on Twitter. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 65–74.
- [8] Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1), 180–196.
- [9] Leskovec, J., McGlohon, M., Faloutsos, C., Glance, N., & Hurst, M. (2007). Patterns of cascading behavior in large blog graphs. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 551–556.
- [10] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.