

Beyond Swiping through Short-Form Videos

Burak Kara[†⋆], Gwendal Simon[†]
[†]Synamedia, [⋆]Inria
Rennes, France

Ayse B. Demir, Ali C. Begen
Ozyegin University
Istanbul, Turkiye

Florence Agboma
Sky
Isleworth, U.K.

ABSTRACT

Short-form videos have emerged as a dominant force in online media, and they have been characterized since the early 2010s by their bite-sized format, fast consumption and high engagement. Initially considered niche social media content, they now account for a substantial share of global media consumption. This paper begins by analyzing a dataset of short-form videos from a popular news application, which includes metrics related to videos consumed within the app in a format akin to YouTube Shorts or Instagram Reels. Building on these insights, the paper poses open questions and advocates for community-driven efforts to establish open protocols for short-form video content.

CCS CONCEPTS

• **Information systems** → **Multimedia streaming**.

KEYWORDS

Content delivery network, CDN.

ACM Reference Format:

Burak Kara[†⋆], Gwendal Simon[†], Ayse B. Demir, Ali C. Begen, and Florence Agboma. 2025. Beyond Swiping through Short-Form Videos. In *Mile-High Video Conference (MHV '25), February 18–20, 2025, Denver, CO, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3715675.3715792

1 INTRODUCTION

Short-form videos have rapidly become dominant in online media consumption, transforming the way users interact with digital content. Since their emergence in the early 2010s, these videos—characterized by their brevity, ease of consumption and engaging nature—have evolved from a niche social media feature to a significant segment of global media consumption. Platforms such as YouTube, Instagram and TikTok have been instrumental in popularizing this format and have transformed how audiences interact with digital content, catering to increasingly shorter attention spans.

While short-form video has revolutionized content delivery, it also introduced unique challenges for researchers and practitioners seeking to optimize the streaming experience. The challenges include optimizing data usage and storage overhead, maintaining user engagement, and ensuring high Quality of Experience (QoE) under varying network conditions. Recent works have addressed these challenges using diverse strategies, including dynamic preloading, adaptive bitrate streaming and learning-based models to predict user behaviors and network conditions.

Data Wastage. Frequent video switching is a defining characteristic of short-form video platforms, where users quickly scroll past uninteresting content. This behavior leads to substantial data wastage, nearly half of the mobile data in short-video streaming is wasted due to early video termination, costing streaming vendors millions annually, per [7–9].

Zhang *et al.* in [7, 8] propose DUASVS, a data-saving strategy that dynamically adjusts prefetch thresholds and bitrate levels to minimize data loss without compromising QoE. In [9], the authors tackle the problem with bandwidth-efficient multi-video prefetching. Nguyen *et al.* [4] further improve data efficiency by introducing a network-aware prefetching method that adjusts prefetched video data based on real-time network throughput. This approach reduces data wastage up to 52 %, outperforming traditional sequential video downloading methods by adapting to both network conditions and user behavior. Moreover, the parallel prefetching in their method not only minimizes data wastage but also reduces startup delays for subsequent videos, thereby enhancing the overall QoE.

Fixed Bitrate Streaming. Many short-form video platforms rely on fixed bitrate streaming, which simplifies encoding but increases the chances of QoE degradation under fluctuating network conditions. Fixed bitrates often mismatch available throughput, causing overloading and data loss. To overcome this, several studies [5, 9] propose Adaptive Bitrate (ABR) streaming strategies tailored for short-form content. For instance, Ran *et al.* [6] introduce short-form video streaming and recommendation (SSR), a system that allocates higher bitrates to videos with longer predicted watch times, enhancing visual quality while conserving bandwidth. However, their reliance on user behavior predictions limits robustness, as switching behaviors are influenced by multiple dynamic factors.

Guo and Zhang [1] propose QASVS that employs deep reinforcement learning to improve the consistency of video quality between sessions. QASVS dynamically selects the optimal bitrate for each session, achieving improvements up to 28 % in video quality while reducing rebuffering events.

Li *et al.* [3] proposed Incendio, a multi-agent reinforcement learning framework that optimizes buffer management and bitrate adaptation to improve both QoE and bandwidth efficiency. Their approach improves system performance by 53 % over traditional ABR methods. This learning-based method dynamically adjusts the bitrate based on past throughput measurements, video chunk sizes and buffer status, mitigating the limitations of fixed bitrate strategies, such as bandwidth wastage [3].

Learning-Based Adaptation Models. Recent advancements explore machine learning-based solutions for short-form video



This work is licensed under Creative Commons Attribution International 4.0.
MHV '25, February 18–20, 2025, Denver, CO, USA
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1488-7/25/02
https://doi.org/10.1145/3715675.3715792

streaming. Deep reinforcement learning has been employed to predict user viewing behaviors and optimize video delivery. For instance, DUASVS [7, 8] utilizes asynchronous actor-critic agent (A3C) reinforcement learning to train specialized adaptation models for diverse network environments. Unlike conventional approaches, the solution categorizes network conditions and matches users to pre-trained models at runtime, achieving robust performance across various scenarios. Similarly, in [2], LiveClip leverages deep reinforcement learning to anticipate video switching, optimizing chunk downloads, and minimizing both startup delay and rebuffering. While learning-based approaches outperform traditional heuristics, they face challenges related to computational overhead and deployment feasibility on resource-constrained devices.

Research Gap. Despite significant progress, several gaps persist in the current state of the art. Existing methods often assume predictable user behaviors, as the models are often trained with social media datasets. However, predicted user behaviors may not hold in real-world scenarios where user preferences and network conditions vary dynamically, especially in news applications. Furthermore, while the learning-based and ABR-based solutions improve data efficiency, their reliance on prefetching and prediction models introduces latency and computational overhead. The generated but not consumed segments create a storage waste at the origin. There is a need for scalable, low-latency solutions that integrate adaptive bitrate streaming, real-time user feedback and emerging transport protocols like QUIC to further optimize and standardize short-form video delivery.

In this paper, we examine a dataset from a widely used news application to better understand the impact of the short-form video format outside of social media platforms. Unlike AI-created or user-generated content, the videos in this dataset are professionally produced, providing a distinctive perspective on editor-driven short-form media. The dataset encompasses key consumption metrics for videos presented in formats similar to YouTube Shorts or Instagram Reels. Through this analysis, we aim to identify patterns and trends in user interactions with short-form videos, offering insights that are both actionable and scalable.

Beyond data-driven insights, this paper also addresses broader issues related to short-form video content, emphasizing the need for open standards and protocols. As the widespread of platforms continue to fragment the media landscape, the lack of standardized approaches to video formatting, metadata and analytics creates barriers for researchers and developers. We advocate for a collaborative approach within the research community to tackle these challenges and advance the development of this dynamic media form.

The remainder of this paper is organized as follows. Section 2 provides an overview of the dataset, describing its structure and key characteristics. Section 3 presents an in-depth analysis of the dataset, uncovering patterns and trends in short-form video consumption. Section 4 outlines open research directions, emphasizing the need for community-driven efforts to develop open standards and protocols for short-form videos. Finally, Section 5 concludes the paper, summarizing the findings and their implications.

2 DATASET OVERVIEW

The dataset spans videos published between January 1 and June 30, 2024, encompassing a total of 61 985 videos, with an average daily addition of 339 new videos. These videos collectively accumulated 205 909 559 views, corresponding to a total watch time of 47 465 726 154 s (or approximately 13 184 924 h or 549 372 d). The dataset includes the following fields: *video_id*, *video_duration*, *video_view*, *video_seconds_viewed* and *aspect*. Each entry corresponds to a unique video and provides metrics on its consumption and characteristics.

This dataset, though smaller in scale compared to those derived from platforms like YouTube, Instagram or TikTok, is substantial enough to enable meaningful analysis of short-form video consumption patterns. The unique context of editor-produced content, as opposed to user-generated or AI-driven videos, adds further value by offering insights into professionally curated video strategies.

To better understand the variety within the dataset, we categorize the videos into two distinct types. *Mobile-friendly videos* are optimized for viewing on mobile devices and have aspect ratios of 1:1 or 9:16. These formats are designed to fit seamlessly within mobile app interfaces, offering a portrait-oriented viewing experience. In contrast, *TV-like videos* have an aspect ratio of 16:9, reflecting their origins in television production. Although these videos were not initially created for mobile consumption, they are adapted to be suitable for viewing on mobile platforms.

3 ANALYSIS

This section presents a comprehensive analysis of the dataset to uncover patterns and trends in short-form video consumption. The analysis is divided into three key aspects: *duration*, *viewing rate* and *popularity*. First, we examine the characteristics of video duration and its impact on user engagement. Next, we explore viewing rates, highlighting differences in consumption behavior between mobile-friendly and TV-like videos. Finally, we analyze video popularity, identifying factors contributing to higher user interest and engagement.

3.1 Video Duration

The duration of a video is a key factor in user engagement, particularly for short-form content. Across the dataset, the average video duration is 186 s. When categorized by format, *mobile-friendly videos* have an average duration of 69 s, aligning closely with the typical definition of short-form content. In contrast, *TV-like videos* are significantly longer, with an average duration of 336 s, reflecting their origins as segments adapted from full-length television programs.

Figure 1, plotted on a log scale, highlights these distinctions. Mobile-friendly videos overwhelmingly fall under 100 s in duration, with only a small fraction exceeding the three-minute mark. This brevity is consistent with their design for fast consumption on mobile devices. Conversely, TV-like videos exhibit a broader range of durations, with only 20 % being shorter than 100 s. The majority of these videos range between two and eight minutes, capturing the longer viewing patterns typical of television content.

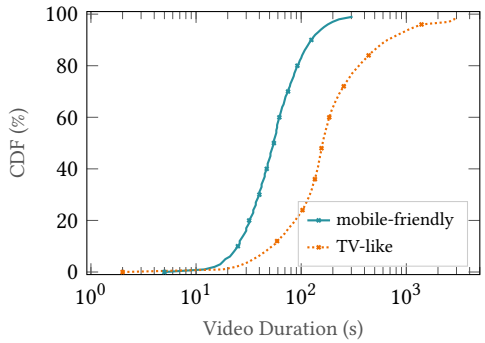

Figure 1: The Cumulative Distribution Function (CDF) of video duration per video type.

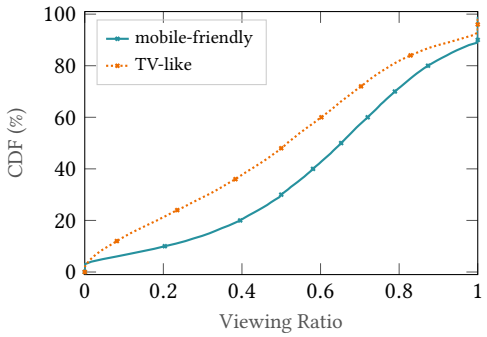

Figure 2: The CDF of video viewing ratio per video type.

The dataset also reveals insights into average view durations. Users watch videos for an average of 69 s, though this metric varies significantly between categories. Mobile-friendly videos have an average view duration of 40 s, reflecting their short format and quick consumption cycle. TV-like videos, on the other hand, demonstrate a higher average view duration of 107 s, indicating greater viewer retention for longer content.

This analysis underscores the contrasting consumption behaviors associated with mobile-friendly and TV-like videos. While short-form content is optimized for brief, engaging interactions, longer videos still capture significant viewer attention, suggesting a nuanced relationship between video length and user engagement.

3.2 Viewing Ratio

A defining characteristic of short-form videos is the swipe interface, allowing viewers to seamlessly switch to a new video before finishing the current one. To measure engagement, we calculate the *viewing ratio*, defined as the average number of seconds watched relative to the total video length. As shown in Figure 2, mobile-friendly videos exhibit a higher viewing ratio than TV-like videos. However, only about 10 % of videos—regardless of type—are viewed in their entirety, corresponding to a viewing ratio of one. Across all videos, the average view rate is 57 %, with mobile-friendly videos achieving a higher average view rate of 62 %, compared to 50 % for TV-like videos.

When we investigate the relation between the video duration and the viewing ratio, shorter videos tend to achieve higher viewing ratios. To analyze this, videos were grouped by length in 10-second increments in Figure 3, considering only those with more than ten views and groups containing at least ten videos. The average viewing ratio decreases with increasing video length, and longer videos are typically watched between one-third and halfway through.

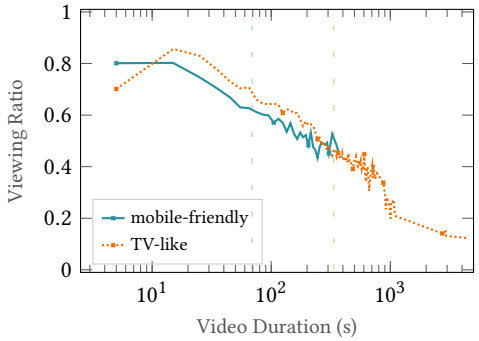

Figure 3: The effect of video duration on the viewing ratio per video type. The horizontal lines indicate the average video duration per type.

3.3 Video Popularity

The distribution of video popularity in the dataset follows an exponential trend, which is a common feature of large online catalogs. Figure 4, plotted on a log-log scale, depicts video views arranged in descending order of popularity. While the most popular video significantly outpaces the second in view counts, the overall slope remains gradual, indicating a relatively modest exponential factor. This distribution suggests that user attention is dispersed across a wide range of videos rather than concentrated on a small subset.

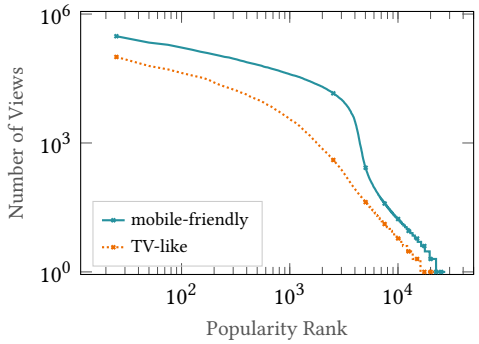

Figure 4: The distribution of video views in decreasing order of video popularity.

The average number of views per video in the dataset is 3841, with mobile-friendly videos receiving 5233 views on average, compared to 2072 views for TV-like videos. This disparity highlights the higher engagement typically associated with short-form mobile content. Moreover, the swipe-friendly interface of the app likely contributes to the observed distribution by encouraging users to explore a larger number of videos within a single session. As a result, the exponential distribution is less steep than those observed in traditional web or video-on-demand (VoD) catalogs.

In addition, viewing ratios were analyzed in relation to popularity, grouping videos into popularity brackets of 100 views (*e.g.*, 0–100, 100–200, *etc.*). Interestingly, the analysis revealed no clear correlation between the popularity of a video and its viewing ratio, suggesting that user engagement is independent of overall view counts.

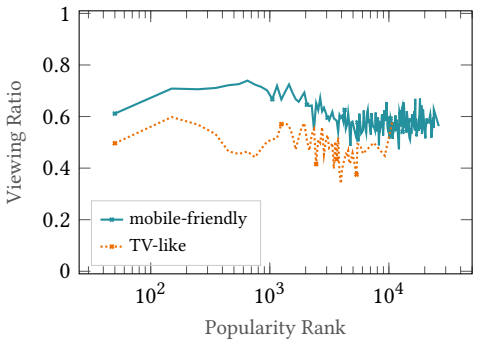

Figure 5: The relation between video duration and viewing ratio per video type.

4 RESEARCH DIRECTIONS

The rapid growth of short-form videos demands advancements in delivery mechanisms to jointly address high user experience and efficient network resource management. This section explores potential research directions, focusing on segmentation and adaptive bitrate streaming, protocol optimization, caching strategies, energy efficiency, open standards and user-centric delivery models.

4.1 Segmentation and ABR Streaming

As revealed in the viewing ratio analysis, short-form videos are rarely consumed in their entirety, resulting in significant data wastage due to preloaded but unwatched content. This inefficiency highlights the need for tailored strategies optimized for the unique consumption patterns of short-form videos.

One promising approach to improving short-form video delivery is the adoption of segmentation and ABR streaming, a method widely used for traditional video content. Segmenting videos into smaller chunks enables dynamic bitrate adjustments based on network conditions, thereby reducing buffering and ensuring a seamless viewing experience. Although recent works in [5, 6, 9] have introduced promising ABR techniques for short-form video streaming, these solutions remain largely confined to research prototypes and have yet to see widespread real-life implementation.

The integration of the ABR streaming with recommendation engines offers a significant opportunity for optimization. Current systems typically treat video recommendation and ABR delivery as separate modules. However, combining these components can enable more intelligent preloading strategies. For instance, recommendation engines can prioritize preloading segments of videos with higher predicted engagement probabilities, reducing unnecessary data downloads for content unlikely to be consumed. By aligning ABR manifests with recommendation algorithms, short-form video delivery systems can further reduce data wastage while maintaining high QoE.

Future research should investigate whether existing ABR algorithms can be adapted or if entirely new methods are required to address the distinct characteristics of short-form videos. These algorithms must consider not only dynamic network conditions but also user behaviors such as frequent video swiping and partial consumption, ensuring both data efficiency and high QoE.

4.2 Protocol Optimization with QUIC

The QUIC protocol (RFC 9000), designed for low-latency and high-throughput communication, offers a significant opportunity to improve short-form video delivery. Its features, such as multiplexing, fast connection setups and built-in encryption, align well with the swipe-based interaction patterns of short-form video platforms. By eliminating head-of-line blocking and enabling rapid video transitions, QUIC can enhance both latency and delivery efficiency compared to traditional TCP-based protocols.

A key area of exploration is the integration of the emerging Media over QUIC (MoQ) standard. MoQ leverages QUIC to enable efficient media delivery, such as real-time adaptive bitrate streaming and low-latency playback. For short-form video platforms, MoQ could facilitate dynamic segment fetching, allowing content to be streamed in smaller, more manageable chunks that align with network conditions and user behavior. This integration could enhance delivery efficiency by reducing preloaded but unwatched data, a significant challenge highlighted in the viewing ratio analysis.

Additionally, MoQ offers potential advantages for edge-based caching strategies. With the ability of QUIC to support connection migration and the fine-grained control of MoQ over media delivery, content providers can optimize edge server utilization and minimize network congestion during peak usage periods. Future research should evaluate the feasibility of deploying QUIC and MoQ at scale, comparing their performance to traditional protocols in terms of latency, reliability and QoE for short-form video users.

While QUIC and MoQ hold significant promise, challenges remain in their adoption. These include the need for widespread client and server support, compatibility with existing infrastructure, and optimization for the unique consumption patterns of short-form videos. Addressing these challenges will require collaborative efforts between researchers, platform developers and standardization bodies to ensure a smooth transition to QUIC-based delivery frameworks.

4.3 Caching Strategies for Short-Form Content

The rapid consumption patterns of short-form videos, driven by swipe-based interfaces, present unique challenges for caching strategies. Users frequently request new videos, often stored across different cache servers, requiring interaction with the Content Delivery Network (CDN) gateway (or router) to locate and fetch the content. This process introduces additional load on the CDN and increases latency, both of which can degrade the user experience.

Prefetching has been suggested as a solution to mitigate these issues by storing segments of anticipated content near the user. While this can reduce startup delays and improve content availability, it may also lead to significant data wastage if the prefetched videos are not consumed. The decision to prefetch must, therefore, balance the trade-off between reducing latency and avoiding unnecessary bandwidth consumption.

Edge caching within CDNs offers another avenue to optimize content delivery. By strategically positioning popular or trending videos closer to end-users, edge servers can minimize the need for repeated fetch requests from origin servers. However, the dynamic nature of short-form video trends makes traditional caching strategies less effective. Incorporating predictive caching algorithms that analyze user behavior and video popularity in real time could enhance cache hit rates and reduce latency. For instance, algorithms could prioritize caching highly engaging videos based on viewing ratios and user interaction metrics.

Future research should also consider the implications of caching on network load and traffic routing. The high churn of requested videos may strain CDN infrastructure, requiring advanced traffic engineering solutions to distribute requests efficiently across servers. Moreover, evaluating the effectiveness of hybrid approaches, such as combining prefetching with dynamic edge caching, could provide insights into balancing latency reduction with resource efficiency. Addressing these challenges is crucial for ensuring scalable and responsive caching systems tailored to the unique demands of short-form video platforms.

4.4 Energy Efficiency in Delivery

The delivery of short-form videos raises concerns about energy efficiency due to the high volume of generated but unconsumed segments, redundant caching and unnecessary transmissions. These inefficiencies contribute to wasted computational resources during encoding and storage, as well as increased energy use during content delivery.

On-demand encoding and just-in-time segment generation offer potential solutions by reducing the overhead of pre-generating and storing unused segments. Optimizing caching strategies to avoid storing low-demand content can also conserve energy. Furthermore, energy-aware ABR algorithms can balance user experience with resource efficiency, dynamically adjusting delivery to reduce energy consumption without compromising quality. Incorporating renewable energy-powered edge servers and energy-efficient network hardware could further support sustainable short-form video delivery.

4.5 Open Protocols for Interoperability

The lack of standardized delivery protocols for short-form videos creates barriers to interoperability across platforms, increasing resource inefficiency and complicating content distribution. Developing open, platform-agnostic protocols could streamline these processes, ensuring consistent delivery quality while reducing redundancies in network infrastructure.

Standardization bodies, such as the Internet Engineering Task Force (IETF), Moving Picture Experts Group (MPEG) and Streaming Video Technology Alliance (SVTA) play a crucial role in defining specifications and guidelines that overcome the unique requirements of short-form video delivery. Efforts like the MoQ standard provide a foundation for advancing interoperability by leveraging features like low-latency communication and efficient segment delivery. Collaboration among researchers, platform developers and industry stakeholders will be critical to establishing these standards and addressing practical challenges, such as rapid content consumption and diverse network conditions. Open protocols would not only enhance efficiency but also foster innovation and scalability across the short-form video ecosystem.

4.6 User-Centric Delivery Models

User-centric delivery models that adapt to individual viewing behaviors and preferences represent a promising direction for short-form video streaming. Real-time analytics can inform dynamic adjustments to video quality, preloading strategies and playback settings, ensuring a personalized and seamless user experience. For example, predicted high-engagement videos could be preloaded in higher quality, while less engaging content might use lower bitrates to conserve bandwidth.

Context-aware optimizations further enhance personalization by considering factors such as device type, network conditions and viewing environment. Additionally, user-driven customization options, such as adjustable preloading or quality preferences, empower users to tailor their experience while improving satisfaction and engagement.

5 CONCLUSION

Short-form videos have transformed digital media consumption, offering users highly engaging and easily consumable content. This paper analyzed a dataset from a popular news application to uncover patterns in short-form video consumption, highlighting differences between mobile-friendly and TV-like videos. Our findings reveal the importance of video duration, viewing behavior and popularity trends in shaping user engagement.

Building on these insights, we proposed several research directions to enhance the delivery of short-form videos. These include exploring segmentation and adaptive bitrate streaming, leveraging protocols like QUIC, optimizing caching strategies, improving energy efficiency, and developing open standards and user-centric delivery models. Addressing these challenges requires a collaborative effort from researchers, industry stakeholders and platform developers.

As short-form videos continue to dominate online media, advancing our understanding of their delivery and consumption is critical. With this paper, we aim to direct questions as a foundation for future research and innovation in this rapidly evolving domain.

REFERENCES

[1] Jing Guo and Guanghui Zhang. 2021. A Video-Quality Driven Strategy in Short Video Streaming. In *ACM MSWiM*.

[2] Jianchao He, Miao Hu, Yipeng Zhou, and Di Wu. 2020. LiveClip: towards intelligent mobile short-form video streaming with deep reinforcement learning. In *ACM NOSSDAV*.

[3] Yueheng Li, Qianyuan Zheng, Zicheng Zhang, Hao Chen, and Zhan Ma. 2023. Improving ABR Performance for Short Video Streaming Using Multi-Agent Reinforcement Learning with Expert Guidance. In *ACM NOSSDAV*.

[4] Duc Nguyen, Phong Nguyen, Vu Long, Truong Thu Huong, and Pham Ngoc Nam. 2022. Network-aware Prefetching Method for Short-Form Video Streaming. In *IEEE MMSP*.

[5] Nguyen Tien Phong, Truong Thu Huong, Pham Ngoc Nam, Truong Cong Thang, and Duc Nguyen. 2023. Joint Preloading and Bitrate Adaptation for Short Video Streaming. *IEEE Access* 11 (2023).

[6] Dezhi Ran, Yuanxing Zhang, Wenhan Zhang, and Kaigui Bian. 2020. SSR: Joint Optimization of Recommendation and Adaptive Bitrate Streaming for Short-form Video Feed. In *Int. Conf. Mobility, Sensing and Networking (MSN)*.

[7] Guanghui Zhang, Ke Liu, Haibo Hu, and Jing Guo. 2021. Short Video Streaming With Data Wastage Awareness. In *IEEE ICME*.

[8] Guanghui Zhang, Jie Zhang, Ke Liu, Jing Guo, Jack Y. B. Lee, Haibo Hu, and Vaneet Aggarwal. 2023. DUASVS: A Mobile Data Saving Strategy in Short-Form Video Streaming. *IEEE Trans. Services Computing* 16, 2 (2023). https://doi.org/10.1109/TSC.2022.3150012

[9] Xutong Zuo, Yishu Li, Mohan Xu, Wei Tsang Ooi, Jiangchuan Liu, Junchen Jiang, Xinggong Zhang, Kai Zheng, and Yong Cui. 2022. Bandwidth-Efficient Multi-video Prefetching for Short Video Streaming. In *ACM Multimedia*.