

Be There, Be Together, Be Streamed! AR Scenic Live-Streaming for an Interactive and Collective Experience

Zeyu Huang*
The Hong Kong University of
Science and Technology, China
Chengzhong Liu§
The Hong Kong University of
Science and Technology, China

Zuyu Xu†
The Hong Kong University of
Science and Technology, China
Yanwei Zhao¶
Zhejiang University, China

Yuanhao Zhang‡
The Hong Kong University of
Science and Technology, China
Chuhan Shi||
Southeast University, China
Jason Chen Zhao**
FieldFusion Ltd., China

Xiaojuan Ma††
The Hong Kong University of
Science and Technology, China

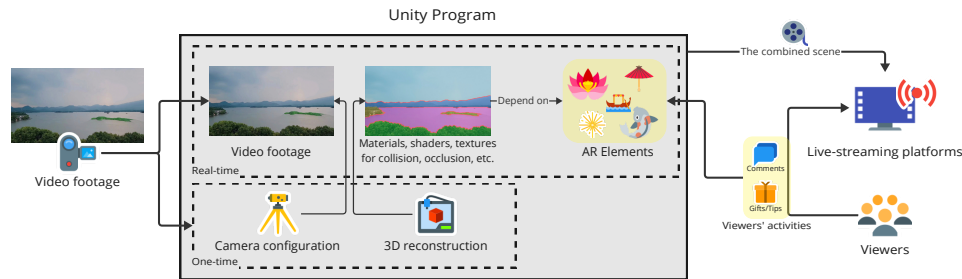


Figure 1: The ARSLS system diagram.

ABSTRACT

Scenic Live-Streaming (SLS), capturing real-world scenic sites from fixed cameras without streamers, combines scene immersion and the social and real-time characteristics of live-streaming into a unique experience. However, existing SLS affords limited audience interactions to engage them in a collective experience compared to many other live-streaming genres. It is also difficult for SLS to recreate important but intangible constituents of in-person trip experiences, such as cultural activities. To offer a more interactive, engaging, and meaningful experience, we propose ARSLS (Augmented Reality Scenic Live-Streaming). Culturally grounded AR objects with awareness of the live-streamed environment can be overlaid over camera views to provide additional interactive features while maintaining consistency with the live-streamed scene. To explore the design space of this new medium, we developed an ARSLS prototype for a famous landscape in China. A preliminary study (N=15) provided initial insights for ARSLS design.

Index Terms: Augmented Reality, AR, live stream, scenic live stream

1 INTRODUCTION

Live-streaming is a popular means of entertainment nowadays where people gather together to watch the same video stream at the same time. Such a time-synced and collective nature often

brings about something beyond pure video-watching experience — a collective experience created by spontaneous interactions among viewers in the same live-streaming room [18, 9]. In recent years, a new form of live-streaming called *scenic live-streaming* (SLS)¹ has emerged: using fixed cameras to capture real-world environments, often scenic views, without human streamers to create a 24/7, scenery-oriented experience. SLS appears on both dedicated websites like SkylineWebcams² and general-purpose video platforms like YouTube³. The situated experience in a remote location provides an alternative to in-person trips [11], including having a taste of new places [11, 13], relieving nostalgia for familiar places [7], and gain educational insights [11].

Apart from the benefits mentioned above, SLS also allows viewers to participate in real-time social interaction. Social connection and community building are common purposes people choose live streams over non-interactive videos [24], and such goals are realized through the exchange of thoughts and feelings in live chats and other channels on the platform [10]. SLS naturally inherits this trait compared to pre-recorded scenic videos. Nevertheless, social interactions and collective experiences in SLS are less engaging because there are no human streamers who constantly “stir up the atmosphere” for more interactions among the audience [9] and timely respond to the audience’s interactions [18, 8, 15, 26, 12]. Contrary to the current limitation, we anticipate an even larger potential of SLS in creating unique collective experiences that transcend sole text communications among viewers: a scenery often bears profound cultural values, and many of them are reflected through collective cultural activities. Recreating such activities will not only help people gain a comprehensive understanding of the live-streamed site, but also leverage the scenery itself and the underlying culture to offer unique real-time experiences.

¹Other names of this medium include but are not limited to *world cameras* and *live webcams*.

²<https://www.skylinewebcams.com/>

³An example of live-streamed Venice: <https://www.youtube.com/watch?v=HpZAez2oYsA>

*e-mail: zhuangbi@connect.ust.hk

†e-mail: zxubo@connect.ust.hk

‡e-mail: yzhangiy@connect.ust.hk

§e-mail: chengzhong.liu@connect.ust.hk

¶e-mail: 458815330@qq.com

||e-mail: chuhanshi@seu.edu.cn

**e-mail: zhaoch21@gmail.com

††e-mail: mxj@cse.ust.hk

To this end, we propose a medium called *Augmented Reality Scenic Live-Streaming* (ARSLS), where additional virtual content is superimposed and integrated into the real-world scene to actualize proper audience interactions in scenic-oriented live streams. AR has already been a valid solution to enhance in-person, single-user experiences with real-world scenery [22, 21], and previous work has also shown the technical feasibility of applying AR to fixed cameras [19, 16]. For ARSLS, the AR content shall be aware of the scene's structures and boundaries, be able to interact with the real-world scene [2], and respond to viewers' interactions in the live-streaming room, such as comments and tipping, in real-time.

It is worth noting that the additional AR content and interactivity may largely differ from current SLS — where the streamed view is mainly for passive appreciation and separate from viewers' interactions. Hence, we propose ARSLS as a new medium derived from SLS, not a replacement for existing SLS. Focusing on the limited engagement and the cultural activities as two underexploited elements, the aim of ARSLS is to offer a real-time immersive experience combining scenic, social, and cultural emphases.

To explore the design space of ARSLS, we developed a prototype in a real-world setting and obtained preliminary insights from a qualitative study (N=15). The ARSLS prototype is set up at West Lake, a famous scenic spot in Hangzhou, China with intriguing aesthetic and profound cultural values [25]. We designed several concrete interactive features, where viewers' spontaneous and collective interactions are driven by the scenic and cultural elements of the site. The implementation is in partnership with a local corporation, FieldFusion Ltd.

2 RELATED WORK

2.1 AR on Static Cameras

Apart from the mainstream AR application on mobile devices and Head-Mounted Displays, AR on static camera images or video footage is another approach to delivering virtual content in a given scene. This is made possible by remote rendering techniques [14, 17]. Szentandrás et al. [19] applied real-time computed, context-aware AR to video footage captured by surveillance cameras, and Liu et al. [16] integrated interactive AR into physical machine tools and machining processes. However, current works mainly focus on technological feasibility and less on designing an engaging and immersive experience. Our design specifically highlights AR for an engaging SLS experience, including strengthening the connection between the audience and the place, as well as viewers' collective interactions.

2.2 Interactive Live-Streaming

Live-streaming is a new form of multimedia service that has gained increasing popularity in recent years. On modern live-streaming platforms, users can not only watch the stream passively, but also interact with the streamer, other viewers, or the streamed content through live chat (i.e. commenting) and gifting (i.e. tipping or rewarding) [18, 8, 15, 26, 12]. Such interactions offer emotional support, make viewers noticeable to the public, and help them develop social connections [18, 15].

However, most research examines the interactions of more popular streaming categories, e.g., gaming and life sharing [18]. SLS, on the other hand, mainly displays real-time views with limited or no motivation for interactions [11, 13]. Thus, the ultimate goal of ARSLS is to provide incentives and opportunities for a comparably interactive viewer experience with other live-streaming categories.

3 ARSLS DESIGN

3.1 System Overview

As illustrated in Fig. 1, the ARSLS is implemented as a Unity program. The video footage is streamed to the program, and the per-

spective of the Unity scene is adjusted according to the camera device that captures the video. Since the camera is fixed (hence so is the view), we can manually reconstruct the 3D structure of the scene in Unity. Proper shaders and textures are attached to the nearer geometries of the 3D reconstruction so that they appear transparent but can occlude virtual objects behind them. Hitboxes are also configured, as well as the special ripple effects of the water surface whenever virtual objects are in touch with it. The Unity program also subscribes to the commenting and gifting events in a specific live-streaming room via public API. As such, additional AR objects not only conform to the scene's geometric structure, but also programmatically respond to viewers' actions in real-time. Finally, the combined scene of the Unity program is remote-rendered on a server and streamed to the live-streaming room.

Our implementation of the ARSLS prototype was in collaboration with a local company, FieldFusion Ltd. The company implemented the fundamental Unity program that captured user action events, created necessary shaders and art assets, and provided the deployment server. On top, we built the 3D reconstruction of the West Lake scene, and designed and implemented the following features that leveraged the real-time commenting and gifting data.

3.2 The Live-Streamed View

The West Lake in Hangzhou, China, is one of the best-known scenic sites nationwide. It is also listed as UNESCO's World Heritage [1]. It bears not only the aesthetic value of the landscape with bridges, causeways, pavilions, temples, etc. but also profound historical and cultural values as the theme of countless traditional artworks and poetry [25], making it a suitable ARSLS site. Existing city SLS services, e.g., SkylineWebcams² and Live China⁴, often adopted a top-down camera view that overlooks the landscape. Therefore, we set up a camera that could capture West Lake and the hills far away and on the left from a lookout point (see Fig. 2a).

3.3 AR Features

3.3.1 Chinese Verse Game "Fei Hua Ling"

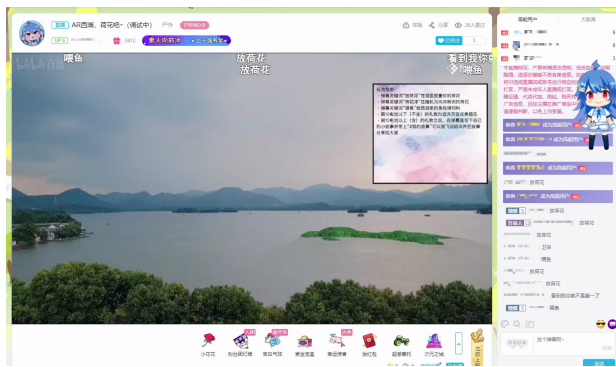
Given the rich collection of classical Chinese poems about West Lake, we adopted a traditional competitive literary game about verses in poems called "Fei Hua Ling" (飞花令 in Chinese, lit. "flying flower command") with slight modifications to suit the live-streaming context. At the beginning of the game, a keyword (e.g., "flower", "snow", and "wine") or a theme (e.g., nostalgia) is announced. The audience collaboratively quotes verses in live comments that contain the keyword. They win the game together if the total number of unique, valid verses exceeds a threshold within a given time (i.e., 5 minutes in our implementation). A virtual game board on the left side shows the verse keyword, the most recent nine valid verses mentioned by viewers, the countdown, the combo, and the total progress. With Fei Hua Ling, we aimed to (1) engage the audience in collective efforts [8, 23] and (2) encourage them to recall and build connections with the local culture depicted in classical literature. We designed two rounds of Fei Hua Ling, 10 minutes apart. The first round uses the most conventional keyword, "flower" (the origin of the game's name). If the audience wins, they will unlock a lasting effect of flower petals flying all over the sky. The second round prompts a theme of "Hangzhou (the city where West Lake is) and Jiangnan (lit. 'South of the Yangtze River', a geographic area where Hangzhou resides)". If the audience wins, a bunch of colorful fireworks will set off, creating a beautiful view.

3.3.2 Lotus Spreading

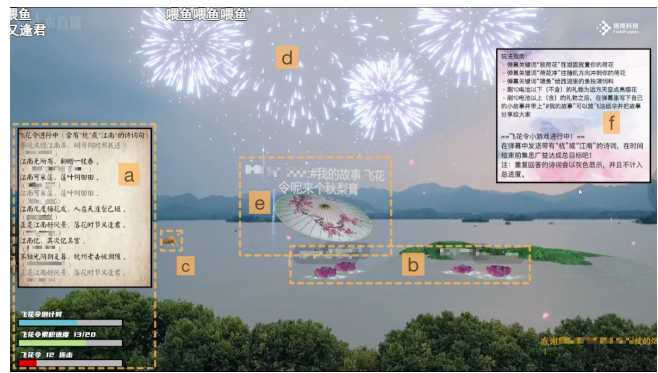
Releasing water lanterns is another traditional collective activity⁵. In Chinese culture, a water lantern is often in the form of a lotus

⁴<https://livechina.cctv.com>

⁵https://en.wikipedia.org/wiki/Water_lantern



(a) The streaming room and the plain scene before any AR object is created.



(b) A screenshot at the middle of the session when all features are triggered. ① Chinese verse game “Fei Hua Ling”; ② Lotus spreading; ③ Fish feeding; ④ Fireworks; ⑤ Oiled-paper umbrella with stories; ⑥ Instructions.

Figure 2: The ARSLs design. The features aim to recreate in-person collective activities related to local cultures. Usernames are occluded.

flower. Thus, we mimic this collective cultural activity in ARSLs. Viewers can place their own virtual lotuses on the lake by issuing a specific comment, `release my lotus`. The lotus displays the user's ID above it, slowly drifting to the right to prevent the scene from being completely occluded. Each user can only have one lotus on the scene; however, if the lotus goes beyond the screen, it will be destroyed, and the user can spawn another one. While the lotus drifts, it triggers virtual ripple effects on the lake. Users can send `dash my lotus` to make their lotus dash in a random direction.

Apart from the ornamental value of lotuses and the cultural meaning of this activity, the feature was also aimed to reinforce viewers' sense of participation through co-creating a shared view [7, 4, 9]. The unique object association might turn lotuses into virtual embodiments of the audience members [6], hence offering a proxy for interpersonal interactions to meet audiences' social needs in live-stream sessions [23].

3.3.3 Fish Feeding

In the live-streaming, we designed fish feeding as a basic interaction with minimal cognitive effort. When users sent the comment `feed fish`, fish food was dropped at a random location over the lake. Then, a virtual fish jumped out of the water to consume the food with a water-splashing sound. The fish picked up a random look from common koi and goldfish in China. This feature was for viewers to familiarize themselves with the ARSLs medium.

3.3.4 Fireworks

Fireworks were ancient Chinese inventions and often launched in batches, with spectators gathering together as a collective activity. In addition, people who launch fireworks first are considered to have more luck [20], which adds nuanced social indication to this activity. Therefore, we associated a fireworks feature with viewers' tipping action. Participants often expect acknowledgment when they tip in a live stream, as it can provide a sense of satisfaction and privilege [26, 18]. Apart from verbal thanks, users may also expect that their ideas are conveyed exclusively, or that other audiences appreciate the outcome of their tipping [18, 12]. As a one-time response to users' cheap gifts (below 10 CNY), virtual fireworks are set off from the far end of the scene and explode in the sky. The tipper's name will be displayed in the meantime. This way, the scene recognizes the monetary input with delicate decorations and a similar sense of privilege.

3.3.5 Oiled-Paper Umbrella with Stories

Previous studies [3, 5] discovered the emotional value of sharing personal stories, especially to relieve people's nostalgia and build strong communities. Considering the symbolization of romance and reunion in Chinese culture of the oiled-paper umbrellas, we combined this imagery with storytelling to help users convey their emotions through AR. Users who send gifts of 10 CNY or above are assigned an oiled-paper umbrella with a random texture. They can then type their stories in a live comment with the hashtag `#MyStory`. The umbrella, with the name and words attached, will be generated at the bottom of the scene and slowly fly up to the sky. We aimed to give tipped users a more eye-catching way to express themselves. The prompt for personal stories also aimed to create emotional appeals to other audiences.

4 PRELIMINARY FINDINGS

A qualitative study (N=15, 9 females and 6 males) was conducted against our ARSLs prototype. Participants were invited to the streaming room simultaneously and remotely. They freely experienced ARSLs together for 20 minutes. All instructions were displayed on the screen; hence, no prior instructions were given to best simulate real-world live-streaming scenarios. The two rounds of verse games (see Sec. 3.3.1) were inserted at the third and the eleventh minutes of the session. We asked participants to interact with the system and use the gifting feature similarly to their regular live-streaming experiences. Only after the study did we offer to reimburse the money they spent on gifts. Apart from the reimbursement, each received a stipend of 80 CNY. Once the streaming ended, participants were invited to semi-structured interviews separately to reflect on their general perception of the ARSLs, the interactive feature design, and the integration of AR in the real scene.

One aspect that emerged in the interview was visual diversity and personalization. The lotuses served as virtual proxies of viewers, participants desired their own proxy to “stand out” and become “identifiable on the screen”. Otherwise, visually repetitive objects would cluster when more viewers joined in. Nevertheless, it suggested the affordance of virtual embodiment in our prototype, which can be an effective way of creating stronger connections between viewers and the environment in ARSLs.

Some participants have complained about the lack of competitive elements to engage them. For example, the verse game did not provide enough mechanisms to encourage viewers' participation, and the lotus movements are too gentle to create an entertaining taste. The rationale behind can be twofold: some may see competitiveness and entertainment as incentives to join an *interactive experi-*

ence, while some — driven by the perceived virtual embodiment — desired teasing and rapports with other viewers as a collective experience.

Also relevant to previous aspects was the emphasis of identifiability of their input, and this demand was intensified by the innate delay of live streams. Especially for the fish feature, when people issued a command and did not know which fish is triggered by them, their interest in using the same feature again would diminish.

The last aspect is vicariousness and immersion. Unlike regular SLS, the strengthened sense of virtual embodiment may have imposed stricter design considerations for this aspect. In our case, viewers were implicitly guided to immerse in first-person views, so the overlooking scenic view fell short. They could not identify representative landmarks in recognizable sizes, nor could they relate themselves to such a view. Consequently, they could not feel immersed in or connected to the scene.

Overall, the relationship between virtual embodiment and immersion presents a major design challenge. Both elements are crucial for a successful ARSLS experience, yet they must be carefully balanced. The virtual embodiment recovers social interactions — an integral part of live streams — in a playful way with no facilitators. The immersion ensured enough significance of the scene, so that viewers do not simply join a game, but can obtain personal, educational, or other values from the scene.

5 CONCLUSION AND FUTURE WORK

Through our prototype and the preliminary results, we have identified participants' intertwined desire for an engaging, embodying collective activity and towards a vicarious immersion. This aligns with, hence generally verifies, our proposal of ARSLS as a combination of scenic and interactive experiences. However, it also indicated a complex design process to balance every aspect. In the future, we would like to respond to such tensions in ARSLS designs via continuous refinement and evaluation of the prototype.

REFERENCES

- [1] W. H. C. . State of conservation of properties inscribed on the World Heritage List, June 2019. 2
- [2] R. T. Azuma. A survey of augmented reality. *PRESENCE: Teleoperators & Virtual Environments*, 6(4):355, Aug. 1997. doi: 10.1162/pres.1997.6.4.355 2
- [3] K. Bennett. Telling Tales: Nostalgia, Collective Identity and an Ex-Mining Village. In *Emotion, Place and Culture*. Routledge, London, 1 ed., 2009. 3
- [4] Z. Deng, P. Benckendorff, and J. Wang. Blended Tourism Experiencescape: A Conceptualisation of Live-Streaming Tourism. In J. Pesonen and J. Neidhardt, eds., *Information and Communication Technologies in Tourism 2019*, pp. 212–222. Springer International Publishing, Cham, 2019. doi: 10.1007/978-3-030-05940-8_17 3
- [5] K. Devine. Removing the rough edges? Nostalgia and storytelling at a UK museum. *Consumption Markets & Culture*, 17(2):208–214, Mar. 2014. 3
- [6] A. C. S. Genay, A. Lecuyer, and M. Hachet. Being an Avatar "for Real": A Survey on Virtual Embodiment in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5071–5090, Dec. 2022. 3
- [7] V. M. González and L. A. Castro. Keeping Strong Connections to the Homeland via Web-based Tools: The Case of Mexican Migrant Communities in the United States. *Journal of Community Informatics*, 3(3), 2007. 1, 3
- [8] O. L. Haimson and J. C. Tang. What Makes Live Events Engaging on Facebook Live, Periscope, and Snapchat. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 48–60. Association for Computing Machinery, New York, NY, USA, May 2017. 1, 2
- [9] W. A. Hamilton, O. Garretson, and A. Kerne. Streaming on twitch: Fostering participatory communities of play within live mixed media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pp. 1315–1324. Association for Computing Machinery, New York, NY, USA, Apr. 2014. 1, 3
- [10] M. Hu, M. Zhang, and Y. Wang. Why do audiences choose to keep watching on live video streaming platforms? An explanation of dual identification framework. *Computers in Human Behavior*, 75:594–606, Oct. 2017. 1
- [11] D. Jarratt. An exploration of webcam-travel: Connecting to place and nature through webcams during the COVID-19 lockdown of 2020. *Tourism and Hospitality Research*, 21(2):156–168, Apr. 2021. 1, 2
- [12] J. Kim, K. Bae, E. Park, and A. P. del Pobil. Who will Subscribe to My Streaming Channel? The Case of Twitch. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, pp. 247–251. Association for Computing Machinery, New York, NY, USA, Nov. 2019. 1, 2, 3
- [13] H. Koskela. 'The other side of surveillance': Webcams, power and agency. In *Theorizing Surveillance*. Willan, London, 1 ed., 2006. 1, 2
- [14] F. Lamberti and A. Sanna. A Streaming-Based Solution for Remote Visualization of 3D Graphics on Mobile Devices. *IEEE Transactions on Visualization and Computer Graphics*, 13(2):247–260, Mar. 2007. 2
- [15] Y. Li, C. Wang, and J. Liu. A Systematic Review of Literature on User Behavior in Video Game Live Streaming. *International Journal of Environmental Research and Public Health*, 17(9):3328, Jan. 2020. 1, 2
- [16] C. Liu, S. Cao, W. Tse, and X. Xu. Augmented Reality-assisted Intelligent Window for Cyber-Physical Machine Tools. *Journal of Manufacturing Systems*, 44:280–286, July 2017. 2
- [17] E. Lu, S. Bharadwaj, M. Dasari, C. Smith, S. Seshan, and A. Rowe. RenderFusion: Balancing Local and Remote Rendering for Interactive 3D Scenes. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 312–321, Oct. 2023. 2
- [18] Z. Lu, H. Xia, S. Heo, and D. Wigdor. You Watch, You Give, and You Engage: A Study of Live Streaming Practices in China. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 1–13. Association for Computing Machinery, New York, NY, USA, Apr. 2018. 1, 2, 3
- [19] I. Szentandrás, M. Zachariá, R. Kajan, J. Tinka, M. Dubská, J. Sochor, and A. Herout. [POSTER] INCAST: Interactive Camera Streams for Surveillance Cams AR. In *2015 IEEE International Symposium on Mixed and Augmented Reality*, ISMAR '15, pp. 80–83. IEEE Computer Society, USA, Sept. 2015. 2
- [20] Timothy S. Y. Lam Museum of Anthropology. Chinese New Year Traditions, 2021. 3
- [21] W.-F. Tung. Augmented Reality for Mobile Service of Film-Induced Tourism App. In P. C. K. Hung, ed., *Mobile Services for Toy Computing*, International Series on Computer Entertainment and Media Technology, pp. 129–140. Springer International Publishing, Cham, 2015. 2
- [22] Y. Wang, X. Deng, K. Zhang, and Y. Lang. The Intangible Cultural Heritage Show Mode Based on AR Technology in Museums - Take the Li Nationality Non-Material Cultural Heritage as an Example. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 936–940. IEEE, Chongqing, June 2018. 2
- [23] A. M. Webb, C. Wang, A. Kerne, and P. Cesar. Distributed Liveness: Understanding How New Technologies Transform Performance Experiences. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pp. 432–437. Association for Computing Machinery, New York, NY, USA, Feb. 2016. 2, 3
- [24] T. Wulf, F. M. Schneider, and S. Beckert. Watching Players: An Exploration of Media Enjoyment on Twitch. *Games and Culture*, 15(3):328–346, May 2020. 1
- [25] R. Zhang and K. Taylor. Cultural landscape meanings. The case of West Lake, Hangzhou, China. *Landscape Research*, 45:1–15, Mar. 2019. doi: 10.1080/01426397.2019.1589438 2
- [26] J. Zhou, J. Zhou, Y. Ding, and H. Wang. The magic of danmaku: A social interaction perspective of gift sending on live streaming platforms. *Electronic Commerce Research and Applications*, 34:100815, Mar. 2019. 1, 2, 3