Journal of Retailing and Consumer Services 84 (2025) 104194

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser

Influential factors of streamer popularity in the live streaming platform

Shuhui Guo [a], Huan Chen [a], Bitao Dai [a], Mengning Wang [a], Shuo Liu [a], Cheng Zhang [b], Shan Liu [c], Xin Lu [a,*]

[a] College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
[b] School of Management, Fudan University, Shanghai 200433, China
[c] Logistics Science and Technology Integration Development Center, School of Management, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Keywords:
Live streaming
Streamer behavior
Popularity prediction
Influential factor
Social network

ABSTRACT

The rapid growth of online live streaming platforms has generated vast interaction data that presents opportunities for a quantitative analysis of streamer behavior and the dynamics of streamer popularity. Despite this, there remains a significant gap in understanding how streaming behaviors and co-playing relationships impact streamer popularity. This study addresses this gap by utilizing a comprehensive dataset encompassing over 450 thousand streamers from January 2020 to April 2023. We demonstrate that features derived from streaming and co-playing behavior can effectively forecast both short- and long-term popularity, achieving an AUC of 0.93 to 0.99. Our analysis indicates that streamer popularity is significantly impacted by the average number of followers and viewers, as well as their strategic positioning within the co-playing network, notably the number of weakly connected components. These findings elucidate strategies for streamers to attract and retain followers, enhancing their presence within the community. This research is the first to explore the influence of the live streamer co-playing network on popularity dynamics, revealing how interactions among online content creators can drive popularity beyond individual behaviors. Moreover, the insights gained can assist brands and businesses in collaborating with streamers more effectively, maximizing their influence and creating mutual value.

1. Introduction

Live (video) streaming is the real-time audio and video transmission of user-generated content (UGC) over the Internet (Chen and Lin, 2018). In a live stream, viewers can watch content while interacting with the broadcaster and other viewers (Kim et al., 2020). They have the opportunity to send virtual gifts to the broadcaster, and engage in free likes and chats (Lin et al., 2021). In recent years, live streaming has become a global economic and social phenomenon, along with the expanding number of users (Ye et al., 2024). Many streaming platforms such as Twitch and YouTube Live have been founded and are experiencing unprecedented growth around the world (Hu et al., 2017). For example, with over 7.7 million monthly active streamers and 1.2 billion visits in August 2022 (StreamScheme, 2023), a popular US-based live streaming platform Twitch has become one of the largest online communities for streamers and live streaming fans. Douyu, a Chinese game-centric live streaming platform, has 471.8 million registered users and 57.4 million mobile monthly active users (MAU) in the fourth quarter of 2022 (Douyu, 2022). Similarly, by the first quarter of 2023, the average DAU of Kuaishou has reached 374 million, and its average MAU has reached 654 million (Kuaishou, 2023). Douyin (Chinese version of TikTok) allows people to express themselves creatively, enjoy entertaining content, and connect with a diverse global community (TikTok, 2023). In 2022 it provided services for nearly 23.9 billion e-commerce orders, benefiting over 1 billion users (TrendInsight, 2023). In short, live streaming has become a career and livelihood for many people, and it has also become an active and emerging field for individuals, enterprises, governments, and other organizations to promote culture, disseminate information, develop e-commerce and so on.

In the live streaming landscape, streamers, as the core creators and disseminators of content, have emerged as pivotal figures. A significant number of popular streamers have emerged as influencers, attracting a loyal fan base, fostering unique personal brands and leveraging their influence to drive related industries forward (Kim and Kim, 2022; Ye et

* Corresponding author.
E-mail address: xin.lu.lab@outlook.com (X. Lu).

https://doi.org/10.1016/j.jretconser.2024.104194
Received 19 August 2024; Received in revised form 25 November 2024; Accepted 6 December 2024

Available online 12 December 2024

al., 2024). It is crucial to focus on the dynamic pattern of streamers' popularity and explore the factors that influence it. By understanding streamers' popularity dynamics, brands and businesses can more effectively collaborate with streamers, leverage their influence, and drive meaningful results for all parties involved (Zhang et al., 2024; Sun et al., 2024; Ma et al., 2024). Furthermore, it can also provide guidance for streamers' professional growth. In a broader context, it also presents tailored behavioral advice to online content creators, empowering them navigate online engagement and maximize impact (Netzorg et al., 2021; Guo et al., 2022; Tian and Frank, 2024).

For social media platforms, considerable research has focused on understanding user popularity based on social network theory (Strogatz, 2001), finding that the characteristics of users' behavior and users' social network structures are both indicative of their popularity (Lesser et al., 2017; Abbas et al., 2018; Cao et al., 2020; Zhang et al., 2021; Gurjar et al., 2022). However, when it comes to live streaming, despite the burgeoning popularity of this medium, there has been relatively limited exploration in applying social network theory to understand and predict the popularity dynamics of live streamers (Netzorg et al., 2021). The majority of existing literature has qualitatively analyzed influential factors of live streamers' performance from the perspective of emotional cognition (Guo et al., 2022; Dang-Van et al., 2023; Li et al., 2024; Luo et al., 2024) or game theory (Zhu and Liu, 2023; Huang et al., 2024). Additionally, some research has utilized the streaming behavior characteristics of streamers to understand their popularity (Jia et al., 2016; Netzorg et al., 2021). There remains a significant gap in understanding how streaming behaviors and co-playing relationships—defined as the social network among streamers—impact the popularity of streamers.

To address this knowledge gap, this paper delves into a vast and long-term dataset of real-time live streaming content sourced from Douyu platform, representing the third basic form of live streaming e-commerce. In this context, e-commerce transaction functionalities are introduced on the basis of the original production of live content, marking a unique and authentic approach to live streaming e-commerce (Daniel, 2022). Given the absence of a universal agreement on the metrics used to measure streamer popularity (such as the number of followers, viewers, and gifts received), this study initially defines streamer popularity from both absolute and relative perspectives, with a focus on the trajectory of change in follower amount. Specifically, we represent streamer popularity through two indicators (Netzorg et al., 2021) P_a and P_r: ranking within the top 10% of follower amount (streamers with a large follower base), and exceeding the median growth rate of follower amount (the fast-growing streamers).

Inspired by latest methods of influential factor analysis, such as regression model (Lu et al., 2021), dynamic simulation (Yang et al., 2023), factor analysis (Yang et al., 2024), this study utilizes predictive features extracted from both streaming behaviors (e.g., the number of viewers, followers, streaming days) and streamers' co-playing relationships (e.g., indegree, PageRank, eigenvector centrality), and constructs a series of machine learning models to accurately predict streamers' popularity at both short- and long-term levels by adjusting the prediction interval, thereby identifying the salient features that significantly influence the dynamics of streamers' popularity.

The rest of this paper is organized as follows: Section 2 reviews social network theory's applications in social media analytics, live streaming platform development, data analysis, and popularity prediction. Section 3 introduces the unique, large-scale live streaming dataset, and explains methods of streamer behavior feature extraction and streamer popularity prediction. Section 4 presents network analysis and illustrates prediction performance of streamer popularity, showing the influential factors of short-term and long-term growth of streamer popularity. Section 5 summarizes key findings and limitations of this research.

2. Related works

2.1. Social network theory and its applications in social media analytics

In recent years, the emergence and rapid proliferation of diverse social applications and media platforms have significantly reshaped the landscape of online interactions. As the backbone of these systems, network structure has evolved into a complex and influential force driving the dynamics of various social processes. Social network theory (Wasserman and Faust, 1994; Freeman, 2004; Liu et al., 2017) provides a theoretical foundation and an analytical framework to study such phenomena, offering a multidimensional lens to explore relationships, behaviors, and interactions. Especially in the context of big data and social media, it demonstrates powerful insights and application potential when integrated with other analytical methods.

Extensive research shows that the socioeconomic characteristics of individuals or communities are closely related to their network positions (Blumenstock et al., 2015; Bollen et al., 2017; Gao et al., 2019; Jusup et al., 2022). For instance, individuals with high degree or betweenness centrality often act as opinion leaders, significantly influencing information dissemination and behavior trends (Peng et al., 2018; Rehman et al., 2023). Therefore, it is essential to identify these influential nodes and implement targeted interventions to mitigate the spread of misinformation (Vosoughi et al., 2018; Aïmeur et al., 2023). Similarly, community detection methods, such as modularity maximization, identify tightly-knit groups within networks, offering insights into patterns of shared interests and collective behaviors (Newman, 2006; Su et al., 2022). In addition, content diffusion research emphasizes the role of weak ties in spreading information across diverse communities, while strong ties reinforce ideas within close-knit networks (Granovetter, 1973; Aubert et al., 2012). Other applications of social network theory include sentiment and opinion analysis, which examines how attitudes and emotions spread through networks, shaping public discourse and contributing to polarization (Lee et al., 2014; Xing et al., 2022). Studies on network evolution explore the dynamics of relationship formation and dissolution, uncovering socioeconomic drivers such as education, occupation, and shared interests that influence network structures (Zhou et al., 2007; Hu et al., 2024).

Overall, social network theory provides a framework for understanding the intricate relationships and dynamics that govern various social processes. By integrating social network theory with advanced analytical methods, researchers gain valuable insights into how individuals' network positions influence their behaviors, shedding light on the complexities of online networks and their societal impact.

2.2. Development of live streaming platforms

In recent years, live streaming platforms have emerged and proliferated rapidly. The large-scale interaction data of the online live streaming platform provides experimental datasets for the quantitative analysis of human behavior, and offers a new opportunity for the mining of the online interaction mechanism with collective dynamics. With the continuous development and popularization of technology, the live streaming industry will continue to maintain a high-speed development trend, bringing more opportunities and possibilities to more people. Table 1 shows the development of some major live streaming platforms.

2.3. Statistical analysis of live streaming data

The statistical analysis research of live streaming data mainly focuses on mining workload patterns (Claypool et al., 2015; Farrington and Muesch, 2015; Pires and Simon, 2015), behavioral characteristics of viewers and streamers (Li et al., 2016; Nascimento et al., 2014; Tang et al., 2006; Zhao et al., 2017; Arnett et al., 2019; Pires and Simon, 2014; Zhang and Liu, 2015), community of interactive networks (Churchill and Xu, 2016; Clauset et al., 2004; Hamilton et al., 2014; Lykousas et

Table 1
Development of several major live streaming platforms.

Platform	Launch Year	Users (MAU)	Streaming Categories	Publicly Listed	Parent Company
TikTok	2016	1 billion +	E-commerce; Sports; Games; Entertainment.	No	Douyin
Douyu	2014	51.7 million	Games; Sports; Entertainment.	Yes (NASDAQ)	Douyu
Huya	2014	239 million	Games; Entertainment; Sports.	Yes (NYSE)	Tencent
Kuaishou	2016	230 million	Lifestyle; Entertainment; Games.	Yes (SEHK)	Kuaishou
Bilibili	2009	169 million	Games; Films; Animation.	Yes (NASDAQ)	Bilibili
Twitch	2011	182 million	Games; E-Sports; Creativity.	No	Amazon
YouTube	2005	2 billion +	Films; Music; Games.	No	Google
Facebook	2004	2.4 billion	Social; News; Entertainment.	Yes (NASDAQ)	Meta
Periscope	2015	-	Lifestyle; News; Entertainment.	No	Twitter

al., 2018), and so on. In this field, heavy tail distributions of the viewing frequency and duration of the audience, the number of rewards and the number of comments, the frequency and duration of broadcasting, and the ability to attract audience have already been discovered, which show the heterogeneity and burstiness of human behavior.

Some studies explore the viewing habits, interaction preferences, and purchase intentions of live streaming users in different situations with data collected through questionnaires (Lim et al., 2020; Park and Lin, 2020; Wongkitrungrueng and Assarut, 2020), or analyze how the live streaming affect sale performance, audience engagement and platform development (Luo et al., 2021; Zhao et al., 2023). Some conclusions are drawn, such as that live streaming can increase sales and customer loyalty, that the number of followers and viewers has a positive effect on sales, and that viewers' wishful identification and emotional engagement have indirect effects on behavioral loyalty.

Other research focuses on exploring the relationship among live streaming platforms, streamers, viewers, and retailers based on game models and simulation experiments, to regulate the behaviors of these agents (Liu et al., 2022; Lv et al., 2022). By analyzing the complex dynamics among them, researchers find that there is a unique optimal decision on the goods price and live streaming effort; the reputation environment in viewers' social networks have an impact on information dissemination in live streaming e-commerce.

2.4. Live streaming popularity prediction

In the context of live streaming, streamer popularity generally pertains to streamer's follower and viewer count, as well as the value of virtual gifts received (Guo and Lu, 2020). Several studies have focused on live streaming popularity. For example, Kaytoue et al. (Kaytoue et al., 2012) have found a high correlation between early and late popularity of a stream in terms of viewer numbers, and proposed a log-linear regression model with a Pearson correlation score varying from 0.75 to 0.9. Netzorg et al. (Netzorg et al., 2021) have proposed a popularity prediction model based on streamers' behavior with AUC (Ling et al., 2003) varying from 0.6 to 0.9, and found streamer efforts, such as increasing the frequency and regularity of live streams, as well as streaming for longer durations, are positively correlated with follower growth. Zhu et al. (Zhu et al., 2017) have conducted a linear correlation analysis between the total number of viewers and the total value of gifts, yielding a regression coefficient of 0.64. Jia et al. (Jia et al., 2016) have found a high correlation score (over 0.83) between streamer popularity and the number of broadcasting times, suggesting that streamers who are more active in broadcasting may be more popular.

Table 2 exhibits the problem definition, methods and metrics in recent research on predicting live streaming popularity. It is evident from the above review that a number of studies have paid attention to live streaming popularity prediction using different methods, indicating the significant role it plays in the digital media and entertainment industry. However datasets used in previous studies were relatively small, limiting the applicability of these findings. In addition, the variables considered in statistical or predictive models were relatively simplistic and incomplete, overlooking streamer interactions. There is an urgent need in live streaming popularity research to explore user behavior characteristics and the structure of user interaction networks.

3. Data and methods

3.1. Dataset

To quantitatively analyze streamer popularity dynamics and to investigate its influential factors, we have continuously collected and maintained a comprehensive live streaming dataset. The dataset is the primary repository of our online open data collection effort initiated since 2018, and continuously updated with daily information from major streaming platforms in China, including Douyu (Douyu, 2023a), Huya (Huya, 2023) and Bilibili (Bilibili, 2023) utilizing APIs provided by each platform. Note that all the collected data are based on public information on the platform.

The complete dataset covers the period from January 2020 to April 2023. For this study, we have extracted two datasets of the Douyu live streaming platform from this repository.

The first dataset (D1) contains of long-term streaming data for all streamers at 30-minute intervals, spanning from January 1^{st} 2020 to April 30^{th} 2023. The streaming data provided by API for each streamer contains 5 attributes: number of followers, number of online viewers, live streaming category, start time of a stream, and the time the data was collected. The live streaming category refers to the category of live broadcast content of streamers. Douyu platform has set up over 3,000 secondary live streaming categories, which belong to 10 primary live streaming categories, such as online games streaming, mobile games streaming, single player games streaming, entertainment streaming, and appearance-based streaming.

To explore both short-term and long-term streamer popularity dynamics, we split all streaming data in D1 into 40 monthly streaming

Table 2
Research on live streaming popularity prediction.

Research	Target	Method	Metric
(Kaytoue et al., 2012; Zhu et al., 2017; Wang et al., 2018; Tu et al., 2018)	Regression on the number of viewers or followers of streamers	Log-linear regression model, Regression tree, Random forests, CART, Adaboost, GBDT	MSE; NMSE; Pearson correlation score; Regression coefficient score; Significance value
(Netzorg et al., 2021; Chen et al., 2021; Xi et al., 2023)	Binary classification on the number of followers, concurrent viewers, cumulative views, cheers, gift value	Logistic regression model, SVM, LASSO, multimodal time-series methods	AUC, F1-score, Precision, Recall, Time consumption
(Arnett et al., 2019; Jia et al., 2016)	Influential factors analysis on the number of views of live streaming	Correlation analysis, popularity distribution comparison	Pearson correlation score
(Guo et al., 2022; Zhao et al., 2019)	Correlation analysis of influential factors on live streaming popularity consumers' watching and consuming intention	Structural equation model	Coefficient score

Table 3
Brief statistics of D1-34 to D1-40.

	D1-34	D1-35	D1-36	D1-37	D1-38	D1-39	D1-40
$No.$ streamers	0.31M	0.31M	0.31M	0.30M	0.29M	0.28M	0.28M
$No.$ viewers	2.82B	2.89B	2.84B	2.79B	2.78B	2.83B	2.96B
$No.$ continued streamers	-	0.20M	0.16M	0.12M	0.11M	0.10M	0.10M
α	1.99	1.99	1.98	1.97	1.97	1.96	1.97
Popular category	OG	OG	OG	OG	OG	OG	OG

Note: "OG" is the abbreviation of live streaming category "online games". OG is short.

series, D1-1 ($t = 1$, from January 1^{st} to 31^{th} 2020) to D1-40 ($t = 40$, from April 1^{st} to 30^{th} 2023). As an example to illustrate the dynamics of the live streaming platform, we have computed five statistical indicators for D1-34 to D1-40 ($34 \leq t \leq 40$) in Table 3. $No.$ streamers shows the number of unique streamers who broadcast in each month is around 0.3 million, and $No.$ viewers shows the average number of viewers is over 2.75 billion. $No.$ continued streamers at D1-t refers to the number of streamers who have broadcasted on both D1-t and D1-$t-1$, and it is around 0.1 million (30% of $No.$ streamers). α means that the power law coefficient of follower count (similar to Fig. 5) is around 1.95, and the most popular category of live streaming for streamers is online games, especially League of Legends, which has not changed from D1-34 to D1-40.

The second dataset (D2) captures streamers' co-playing behavior, collected on October 14^{th}, 2022. Douyu provides a special function called 'neighborhood', allowing streamers to add other streamers as friends (Douyu, 2023b). This tool facilitates directing traffic to other streamers and showcases their co-playing relationships. This function is particularly suitable for streamers within unions with shared interests, reflecting streamers' teamwork, organization, and business planning. Because all streamers are free to add any others (up to 54), the co-playing relationships are directed. As illustrated in Fig. 1, when A adds B and C adds A as a friend, we designate C as A's **source friend**, and B as A's **target friend**.

We included all streamers from December 2018 to October 2022 in the sample streamer list, and iteratively crawled each sample streamer's target friends. Any new streamer found in the sample streamer's target friend list but not in the sample streamer list were added to the sample streamer list. Once all sample streamers in the sample streamer list were iterated, D2 was completely collected. Ultimately, D2 includes all streamers' neighbors.

Building on social network theory, we construct a network to represent the co-playing relationships among streamers on the live streaming platform. By analyzing this network, we investigate how a streamer's structural position—such as centrality or connectivity—affects their popularity, offering insights into the influence of network dynamics on individual success in live streaming environments. A friend relationship

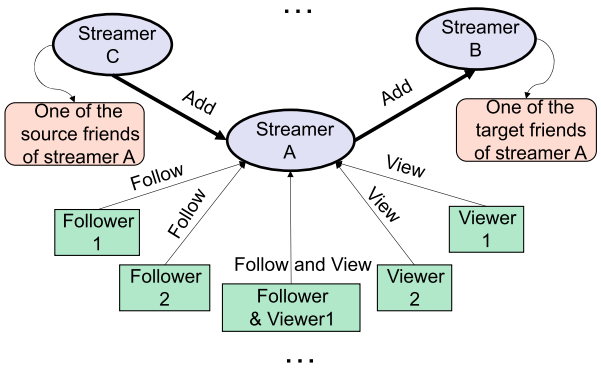

Fig. 1. Various relationships of a streamer and other users on the live streaming platform.

can be described as a directed edge from the source streamer to the target streamer, and D2 can be formed as a network $G(V, E)$ with V nodes and E edges. Here we exclude components only have one streamer in $G(V, E)$ and get the streamers' co-playing network G_{SCN}, containing 456,717 nodes and 1,317,275 directed edges. The giant weakly connected component of G_{SCN}, called G_{SCN-GC}, contains 396,164 nodes and 1,275,506 directed edges.

We illustrate the co-playing relationships among streamers in Fig. 2. The nodes represent streamers in the giant strongly connected component of G_{SCN} who have also streamed in D1-34, and the edges represent the co-playing relationships among streamers. To clearly display the streamers with high connectivity, only nodes whose $k_{in} \geq 10$ are retained in the figure. A total of 623 nodes (6.63%) and 3,552 edges (11.0%) are visible. The colors of nodes mean the live streaming categories of streamers. The larger the node, the more followers the streamer has. The ten nodes that display each streamer's nickname are the top ten nodes in terms of k_{in} in the network.

There are several communities formed by nodes with the same color, such as the orange community (OVERWATCH2), the blue community (World of Warcraft) and the black community (CS: GO). The top ten

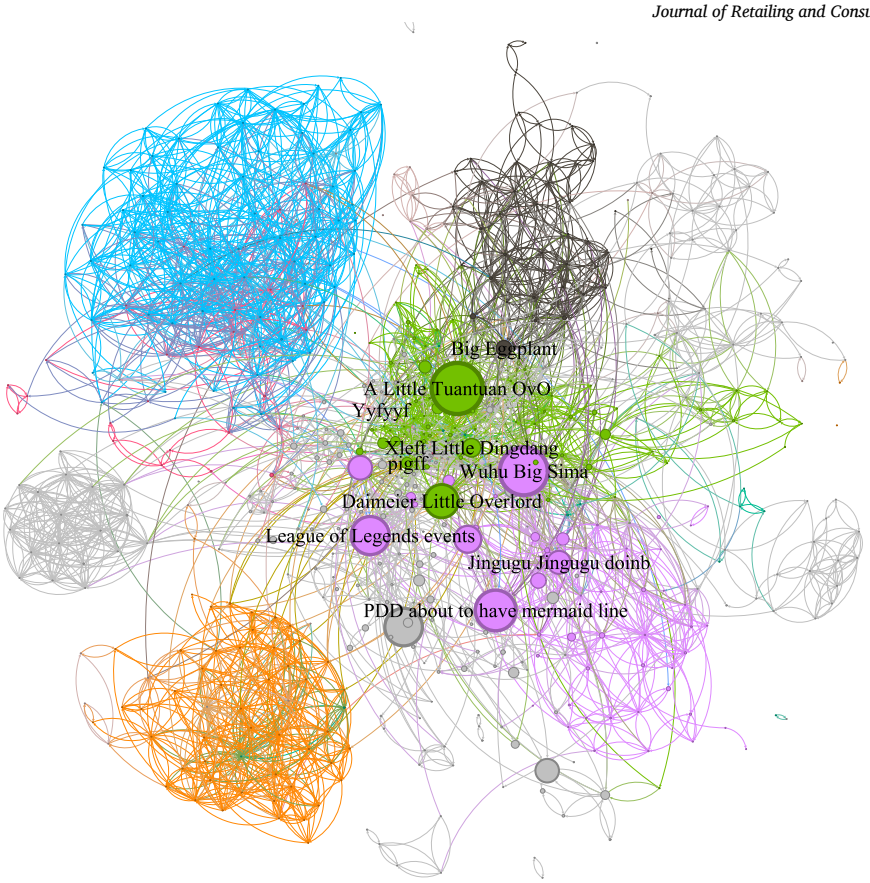

Fig. 2. Streamers' co-playing network.

nodes in k_{in} basically have a large number of followers, and distribute at the center of the graph. For example, the streamer "PDD about to have mermaid line" has more than 19 million followers, whose $k_{in} = 259$ is also very high. But streamer "pigff" and streamer "Yyfyyf" respectively have 5.1 million and 2.7 million followers, lower than some streamers with smaller k_{in}. From Fig. 2 we observe that streamers are likely to have co-playing relationships with streamers in the same stream category. Streamers with more followers are more likely to attract streamers to co-play with them.

We also visualize the streamers' live streaming categories in G_{SCN} (Fig. 3), where streamers with the same live streaming category are integrated into one node, with the node size positively correlated with the number of streamers in that category. All nodes are distributed in a coordinate system with indegree (the number of edges whose target friend uses the live streaming category) as the horizontal axis and outdegree (the number of edges whose source friend uses the live streaming category) as the vertical axis. Nodes with the same color belong to the same community, detected by Infomap algorithm (Rosvall and Bergstrom, 2008). The color of edges is determined by the source node, and the thickness of edges is determined by the number of streamers.

There are 1,054 nodes, 33,862 edges and 13 communities in the diagram. In addition to the edges between nodes, there are also many self-loops, indicating that streamers in the same category have numerous co-playing relationships. Some streaming categories with large amount of streamers have a significantly large indegree and outdegree. For example, League of Legends has 50,758 streamers with an indegree of 806 and an outdegree of 479. Many streamers from streaming categories with small amount of streamers choose to interact with streamers from streaming categories with large amount of streamers, such as League of Legends, Hornor of Kings and Top games.

We also demonstrate the basic statistic information of G_{SCN} and G_{SCN-GC} in Table 4. From Table 4 we can tell that the average indegree

Table 4
Brief statistics of G_{SCN} and G_{SCN-GC}.

	G_{SCN}	G_{SCN-GC}
V	456,717	396,164 (86.7%)
E	1,317,275	1,275,506 (96.8%)
$d(G)$	34	34
$\langle d \rangle$	8.65	8.65
$D(G)$	6.32×10^{-6}	8.13×10^{-6}
N	393,509	1
$Modularity$	0.06	0.05
k^{in}	2.88	3.22
k^{out}	2.88	3.22

k^{in} and outdegree k^{out} are equal (around 3) in G_{SCN} and G_{SCN-GC}, and the network density $D(G)$ are quite sparse when there are over 300 thousand disconnected components ($N = 393,509$) in G_{SCN}. The connectivity efficiency is relatively low when the network diameter $d(G)$ is around 30, the average shortest path distance $\langle d \rangle$ is around 8 and the network modularity is around 0.05.

3.2. Feature extraction

Using the broadcasting and co-playing behavior dataset, we design and calculate a series of **streaming features (SF)** from D1, and **co-playing network features (NF)** of streamers from D2, respectively, to reveal the key performance indicators (KPIs) for streamers' live streaming popularity. These features are then fed into a number of machine learning models to evaluate the effectiveness of predictive powers on live streaming popularity.

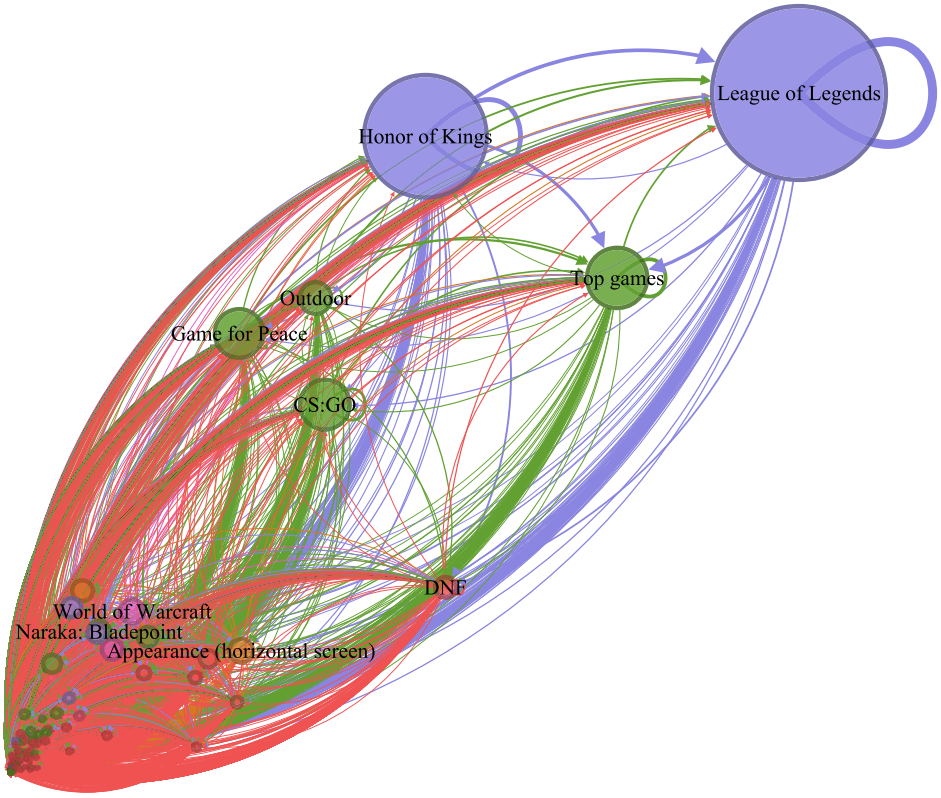

Fig. 3. Live streaming categories in streamers' co-playing network.

3.2.1. Streaming features

As mentioned above, we split all streaming data in D1 into 40 monthly streaming series D1-1 to D1-40. Each streamer who had streamed during a month has 3 kinds of data sequence: the number of online viewers, the number of followers and the streaming category. All streamers' data sequences constitute the data matrices of D1-1, M_{viewer}, $M_{follower}$, and $M_{category}$:

$$M_{viewer} = \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} & \cdots & V_{1,d} \\ V_{2,1} & V_{2,2} & V_{2,3} & \cdots & V_{2,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{n,1} & V_{n,2} & V_{n,3} & \cdots & V_{n,d} \end{bmatrix} \tag{1}$$

$$M_{follower} = \begin{bmatrix} F_{1,1} & F_{1,2} & F_{1,3} & \cdots & F_{1,d} \\ F_{2,1} & F_{2,2} & F_{2,3} & \cdots & F_{2,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F_{n,1} & F_{n,2} & F_{n,3} & \cdots & F_{n,d} \end{bmatrix} \tag{2}$$

$$M_{category} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & \cdots & C_{1,d} \\ C_{2,1} & C_{2,2} & C_{2,3} & \cdots & C_{2,d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & \cdots & C_{n,d} \end{bmatrix} \tag{3}$$

$V_{i,j}$, $F_{i,j}$ and $C_{i,j}$ respectively represent the number of online viewers, the number of followers and the streaming category of streamer i in the j^{th} day of the month. d means that a month can have a maximum of d days. If the streamer i don't stream in the j^{th} day, the value of matrix in the position i, j will be Nan. Based on the matrices above, we calculate 5 streaming features for each streamer in D1-1 to D1-40 to reveal how streamers' live streaming behaviors influence their popularity.

1) The number of streaming days of streamer i:

$$SSD_i = \sum_{j=1}^{d} \begin{cases} 1, & \text{if } C_{i,j} \neq Nan \\ 0, & \text{if } C_{i,j} = Nan \end{cases} \tag{4}$$

2) The average amount of online viewers of streamer i:

$$SAV_i = \frac{\sum_{j=1}^{d} V_{i,j}}{SSD_i} \tag{5}$$

3) The average amount of followers of streamer i:

$$SAF_i = \frac{\sum_{j=1}^{d} F_{i,j}}{SSD_i} \tag{6}$$

4) Considering the scale of the number of followers of streamers may impact the streamer popularity dynamics, we artificially divide SAF_i into three levels:

$$SFL_i = \begin{cases} 1, & SAF_i < 100 \\ 2, & 100 \leq SAF_i < 10000 \\ 3, & SAF_i \geq 10000 \end{cases} \tag{7}$$

5) The category entropy of streaming category of streamer i:

$$SCE_i = -\sum_{c \in C_i} p(c)\log(p(c)),\ C_i = \{C_{i,1}, C_{i,2}, ..., C_{i,d}\} \tag{8}$$

Table 5 shows the brief introduction and statistics of each feature for streamers in D1-34 (October 2022). The mean value of all streamers' SSD is 8 days, with a standard error of 9. The distributions of SAV and SAF are both non-uniforms, with a significant difference between the mean and median values. The distribution of SCE shows a smaller difference, with both the mean and median values being close to 1, and a standard deviation of 0.45. SFL is a classification display of SAF, where the proportions of streamers with SFL values of 1, 2, and 3 are 27.78%, 68.72%, and 3.50%, respectively.

3.2.2. Co-playing network features

In the streamer co-playing network $G_{SCN}(V, E)$, a node $v \in V$ represents a streamer v, and a directed edge from node u to node v is $(u, v) = e \in E$, which represents the co-playing relationship from streamer u to

Table 5
Statistics of streaming features (SF) and co-playing network features (NF).

Feature	Mean	Median	std.	Range
SSD	8	4	9	[1, 31]
SAV	9,194	350	69,023	[0, 7,094,554]
SCE	1.12	1.00	0.45	[1.00, 11.00]
SAF	4,833	336	126,951	[0, 24,618,243]
SFL	-	-	-	1, 2, 3
NID	3	0	129	[0, 42,072]
NOD	3	1	5	[0, 54]
NWC	2	0	81	[0, 29,894]
NSC	3	0	128	[0, 41,852]
NPK	2.19×10^{-6}	5.21×10^{-7}	3.82×10^{-5}	[5.21×10^{-7}, 0.01]
NRE	0.12	0	0.29	[0, 1]
NCL	0.32	0.23	0.25	[0.07, 1]
NBT	2.71×10^{5}	0	6.91×10^{6}	[0, 1.82×10^{9}]
NCR	3	2	4	[1, 29]
NEC	7.33×10^{-5}	1.39×10^{-20}	6.75×10^{-3}	[0, 1]

streamer v. To investigate the influence of streamers' positional features within the co-playing network on their popularity, we also calculate 10 network features of all nodes in G_{SCN}.

1) The indegree of v, i.e., the number of source friends of a streamer:

$$NID = |u \in V|(u,v) \in E| \tag{9}$$

2) The outdegree of v, i.e., the number of target friends of a streamer:

$$NOD = |u \in V|(v,u) \in E| \tag{10}$$

3) The number of weakly connected components (WCC) (Zhang et al., 2021) of v's source friends:

$$NWC = |WCC_1, WCC_2, \ldots, WCC_k| \tag{11}$$

4) The number of strongly connected components (SCC) (Zhang et al., 2021) of v's source friends:

$$NSC = |SCC_1, SCC_2, \ldots, SCC_k| \tag{12}$$

5) PageRank value (Brin and Page, 1998) of v:

$$NPK = (1-c) + c\sum_{u\in M(v)}\frac{PR(u)}{L(u)} \tag{13}$$

where c is the damping factor. $M(v)$ is the set of all nodes that link to v. $PR(u)$ represents the PageRank value of u, which is a node that links to v. $L(u)$ represents the outdegree of u.

6) Reciprocity value (Newman, 2003) of v:

$$NRE = |(u,v) \in E|(v,u) \in E|/|(v,u) \in E| \tag{14}$$

7) Closeness centrality (Freeman, 1979) of v:

$$NCL = \frac{n-1}{\sum_{u\in V} d(v,u)} \tag{15}$$

where n is the number of nodes in G_{SCN}, $d(v,u)$ is the length of the shortest path from v to u. $\sum_{u\in V} d(v,u)$ is the sum of the distances from v to all other nodes in G_{SCN}.

8) Betweenness centrality (Freeman, 1979) of v:

$$NBT = \sum_{s\neq v\neq t\in V}\frac{\sigma_{st}(v)}{\sigma_{st}} \tag{16}$$

where V is the set of all nodes in G_{SCN}, s and t are two distinct nodes in G_{SCN}, different from v. σ_{st} denotes the number of all shortest paths from s to t. $\sigma_{st}(v)$ denotes the number of shortest paths from s to t that pass through v.

9) Coreness value (Batagelj and Zaversnik, 2003) of v:

$$NCR = \max\{k \,|\, \text{if all nodes } u \text{ with } d(u) < k \text{ and their incident edges are removed, then } v \text{ still belongs to } G_{\text{SCN}}\} \tag{17}$$

where $d(u)$ is the degree (indegree and outdegree) of u.

10) Eigenvector centrality (Bonacich, 2007) of v:

The eigenvector centrality formula for directed graphs is based on the largest eigenvalue λ of the adjacency matrix A and its corresponding eigenvector x. The formula can be represented as:

$$Ax = \lambda x \tag{18}$$

where A is the adjacency matrix for G_{SCN}. λ is the largest eigenvalue of the adjacency matrix A. x is the eigenvector associated with λ. The eigenvector centrality value of a node (NEC) is the corresponding element's value in x.

Table 5 also provides a concise overview and statistics of each introduced feature. The average value of NID and NOD is 3, which means that each streamer has an average of 3 source friends and 3 target friends. Due to the restriction of maximum number of target friends that can be added, maximum NOD is 54 while maximum NID is 42,072. The NWC and NSC of most streamers are small (about 3), and the NWC and NSC of a small number of streamers are large (about 30,000). The other features show the structural importance of streamers in G_{SCN} from different aspects.

3.3. Problem definition and prediction settings

3.3.1. Popularity: definition and dynamics

Online streaming popularity prediction is typically modeled as a binary classification problem (Netzorg et al., 2021; Chen et al., 2021). This study defines streamer popularity from two dimensions: absolute popularity (P_a) and relative popularity (P_r).

1) P_a

Assuming a streamer has a follower count of SAF_t at time t ($1 \le t \le 39$), and a follower count of $SAF_{t+\delta}$ at time $t+\delta$ ($2 \le t+\delta \le 40$), if $SAF_{t+\delta}$ ranks within the top 10% among all streamers, P_a of the streamer is assigned as 1; otherwise, it is assigned as 0. $P_a = 1$ represents streamers who rank within the top 10% of follower amount. These streamers possess a large and dedicated follower base, indicating their widespread appeal and established presence in the streaming community. By focusing on this metric, we aim to capture the most influential and recognizable streamers in the field.

2) P_r

Assuming a streamer has a follower count of SAF_t at t ($1 \le t \le 39$) and a follower count of $SAF_{t+\delta}$ at $t+\delta$ ($2 \le t+\delta \le 40$), the growth rate of the streamer's follower count is calculated as:

$$SGR = \frac{SAF_{t+\delta} - SAF_t}{SAF_t} \tag{19}$$

If SGR of the streamer exceeds the median SGR of all streamers, P_r of the streamer is assigned as 1; otherwise, it is assigned as 0. $P_r = 1$ represents streamers who exceed the median growth rate of follower amount. These streamers demonstrate rapid and sustained growth in their follower count, indicating their ability to attract and retain new audiences over time. By including this metric, we aim to identify emerging talents and trends within the streaming industry.

By considering both P_a and P_r simultaneously, we can gain a comprehensive understanding of a streamer's popularity and potential. While P_a highlights established streamers with a large and dedicated follower base, P_r uncovers emerging talents who are growing rapidly. This dual focus allows us to not only recognize the current leaders in the streaming community but also anticipate future trends and rising stars. Therefore, this dual-measurement approach significantly improves the ability to predict popularity dynamics accurately and comprehensively.

On the other hand, the prediction interval parameter δ, which ranges from 1 to 39, plays a crucial role in capturing both short- and long-term trends in streamer popularity. Given the dataset D1 covers a time span of 40 months from D1-1 to D1-40, we leverage predictable features available at any given time t ($1 \le t \le 39$) to forecast the streamer popularity

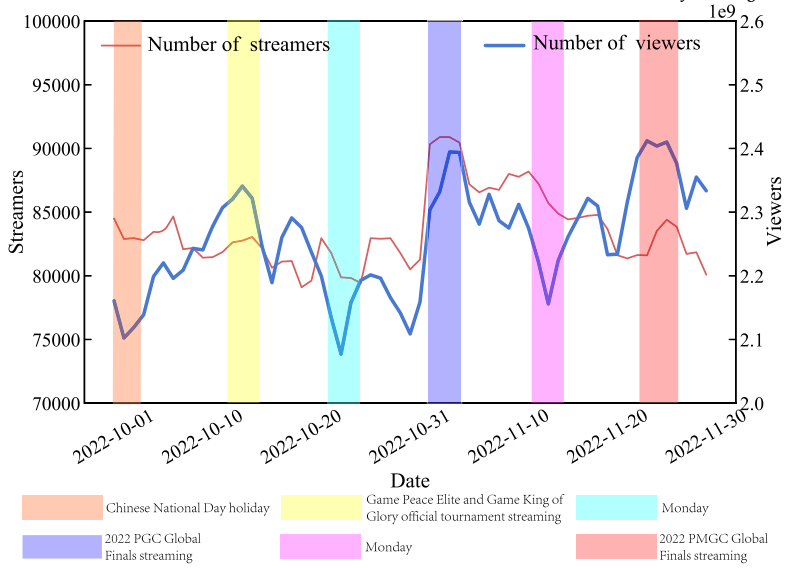

Fig. 4. The temporal pattern of live streaming workload.

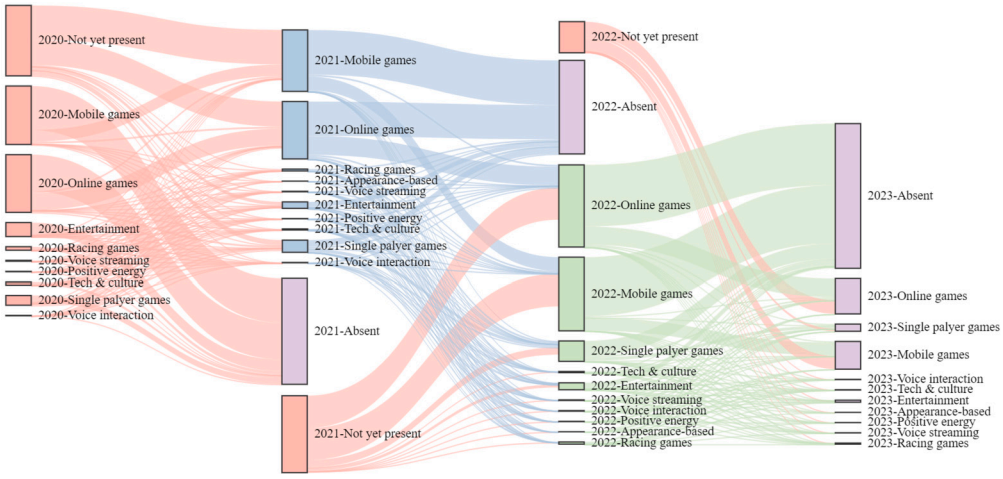

Fig. 5. Temporal dynamics of live streaming category's streamer count.

P_a and P_r, at a future time $t+\delta$ ($2 \le t+\delta \le 40$). As δ increases from 1 to 39, it reflects the dynamic evolution of streamer popularity, transitioning from short-term popularity gains to long-term popularity.

First, we observe the temporal patterns of the number of streamers and viewers which implying the workload level of the live streaming platform during D1-35 to D1-36 as an example (Fig. 4). It can be seen that the number of viewers of Douyu TV from October to November of 2022 has a clear and fluctuating rise trend. The number of viewers increases from the beginning of 210 million to the ending of 233 million. On the other hand, the number of streamers has been relatively stabilized between 80 thousand and 90 thousand.

Fig. 5 shows the temporal dynamics of the streamer count for each live streaming category from the year of 2020 to 2023. As we can see, a significant proportion of streamers both joined the live streaming platform (Not yet present) and left it (Absent), with the respective percentages reaching as high as 32% and 42%. There are two main categories, Mobile games and Online games, attracted more than 26% streamers over years. From 2020 to 2022, many streamers opted to switch streaming categories, while a consistent proportion of streamers (14% to 17%) remained in the same category for live streaming.

3.3.2. Prediction settings

In the prediction process, we choose six commonly used binary classification models: LightGBM (Ke et al., 2017), Random forest (Biau and Scornet, 2016), Adaboost (Freund and Schapire, 1997), Logistic (LaValley, 2008), Decision tree (Myles et al., 2004) and GBDT (Friedman, 2002) as candidate prediction models. We use two classical evaluation metrics for binary classification: F1-score (Chicco and Jurman, 2020) and AUC as performance metrics.

To evaluate the importance of features in popularity prediction, we conduct prediction experiments on three feature combinations, i.e., SF, NF, SF&NF. All categorical features are encoded by one-hot method. All features in the training dataset and test dataset are Min-Max standardized to reduce the impact of extreme values. The size of test dataset is set as 20% with dataset shuffling to avoid overfitting.

When δ changes from 1 to 39, the size of dataset may change because the streamers may leave live streaming platform at any time. To avoid this situation, we set the size of dataset of all prediction experiments to be 10,000 random samples with P_a or P_r equals to 1 and 10,000 random samples with P_a or P_r equals to 0.

By observing the changes in prediction performance under different time intervals and feature combinations, we can identify the influential factors of short-term and long-term live streaming popularity, as well as determine the feature combinations that yield the best prediction performance.

3.3.3. SHapley Additive exPlanations (SHAP) values

SHAP is an additive method to overhaul each feature contribution to the prediction results in the learning model (Lundberg, 2017). Let f be the original prediction model to be explained and g the explanation model. The output value is given as:

$$g\left(z'\right) = \phi_0 + \sum_{j=1}^{M} \phi_j z_i' \tag{20}$$

where g is the explanation model; $z' \in \{0,1\}^M$ indicates whether each feature exists; M is the number of input features; ϕ_i is the Shapley regression value of feature i for model f in the presence of multicollinearity; summing the effects of all feature attributions approximates the output $f(x)$ of the original model f.

$$\phi_i = \sum_{S \subseteq F\{i\}} \frac{|S|! * (|F| - |S| - 1)!}{|F|!} \left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S\left(x_S\right)\right] \tag{21}$$

where F is the set of all features; The model f is trained on all possible feature subsets $S \subset F, S \subseteq F/\{i\}$. f_S is the model retrained on S. x_S means the values of input features in S. Finally, the feature importance is ranked by calculating the total SHAP values magnitude ϕ_i over all predictions.

4. Results

4.1. Network analysis of broadcasting and co-playing activities

4.1.1. Degree distribution of streamers

We illustrate probability distributions of NID and NOD in Fig. 6(a) and Fig. 6(b). The distributions of NID and NOD follow a power-law distribution, with coefficient α of 1.57 ($r^2 = 0.99$) and 2.30 ($r^2 = 0.99$), respectively, indicating significant heterogeneity in the number of streamer friends, both source and target.

The majority of streamers (65%) either have no source friends or just 1 source friend (22%). Only 11 out of 456,717 streamers have more than 10 thousand source friends, all of whom are famous and popular streamers in Douyu. This result implies that majority streamers are not easy to be added as target friends. Streamers may add other streamers as target friends for 2 reasons: 1) to direct their own viewers and followers to a few streamers and 2) to seek recognition or reciprocation from popular streamers. We also conducted a single factor analysis of variance (ANOVA) to assess the significant differences among different groups of source friend amount on the number of follower and viewers, and found that the differences to be significant ($p < 0.001$).

Most streamers (88%) tend to add fewer than 5 streamers as target friends. The first 3 commonly used target friend amount is 1 (44%), 0 (14%) and 2 (14%). This phenomenon imply that streamers are inclined to direct their viewers and followers to a few streamers to prevent excessive resource decentralization. We also conducted ANOVA to assess the significant differences among different groups of target friend amount on the number of followers and viewers, and found that the differences to be significant ($p < 0.001$).

4.1.2. Second order degree distribution of streamers

We then explore the distribution of the number of source friends of streamers' source friends, i.e. the second order indegree (Fig. 7(a)). For streamers with only 1 source friend (22%), their source friends are likely to have an average of 5 source friends with variance of approximately 101. Streamers with 2 to 10 source friends (11%) have source friends with an average of 9 source friends and a variance of approximately 110. Streamers with 11 to 100 source friends (2%) have source friends with an average of 15 source friends and a variance of approximately 82. Streamers with over 10 thousand source friends (fewer than 0.01%) with an average of 3 source friends and a variance of approximately 2.

Similarly, we also conduct ANOVA to assess significant differences among different groups of average number of source friends of each streamer's source friends on the number of followers and viewers, and found that the differences to be significant ($p < 0.001$).

Fig. 7(b) illustrates the distribution of the number of target friends of streamers' target friends, i.e., the second order outdegree. Streamers with fewer than 5 target friends are more likely to have target friends who add fewer than 6 streamers as target friends. On the other hand, streamers with over 5 target friends are likely to have target friends who add 7 to 9 streamers as target friends. The number of target friends of streamers' target friends does not show significant difference among 10 groups, from (5-10] to (50, 55].

We also conduct ANOVA to assess significant differences among different groups of the number of target friends of streamers' target friends on the number of followers and viewers, and found that the differences to be significant ($p < 0.001$).

4.1.3. Friendship preference analysis

In this section, we explore the preferences of source streamers (those who add others as target friends) and target streamers (those who are added by other streamers as target friends) in G_{SCN}. The Pearson correlation scores (r) of SAF, SAV, SSD, NID and NOD between source streamers and target streamers are shown in Table 6.

We can find that source streamers are weakly correlated with target streamers in SAF, SAV, SSD, NID and NOD ($-0.15 < r < 0.15$). Target streamers tend to have more followers (more than 4 million) and viewers (more than 1 million) compared to source streamers. Target streamers tend to be more active in the streaming platform, with average active streaming days in a month is 22 days. Additionally, the average number of source friends of source streamers is 10 but the average number of source friends of target streamers is 6,008.

Fig. 8 shows the preferences for streaming category among source streamers and target streamers. Each row and column represent a stream category, and the darker the blue color, the greater the number of streamers. Streamers are more likely to add streamers from the same streaming category as target friends (56%). For example, 387,861 of 484,604 (80%) online games source streamers have target streamers in online games streaming category and 96,743 (20%) have target streamers in other nine streaming categories.

4.2. Popularity prediction

4.2.1. Prediction model comparison

We use all 6 candidate models for streamer popularity prediction. Table 7 shows the average AUC and F1-score (when δ varies from 1 to 39) of 6 candidate models, using the feature combination SF, NF, SF&NF, respectively.

By comparing prediction performances of 6 candidate models, it can be clearly found that LightGBM has the best average AUC (0.95) and F1-score (0.95), so we finally choose LightGBM as the streamer popularity prediction model.

4.2.2. Prediction performance

Fig. 9 shows the prediction results of streamer popularity using LightGBM. The best prediction performance of P_a and P_r are both gained by SF&NF. When δ varies from 1 to 39, SF&NF achieves AUC values ranging from 0.93 to 0.99 in P_a prediction and 0.78 to 0.94 in P_r prediction, as well as F1-scores ranging from 0.93 to 0.99 in P_a prediction and 0.79 to 0.94 in P_r prediction. SF&NF demonstrates a superior ability to predict streamers' popularity, achieving the highest AUC values of 0.99 for P_a and 0.94 for P_r, as well as the best F1-scores of 0.99 for P_a and 0.94

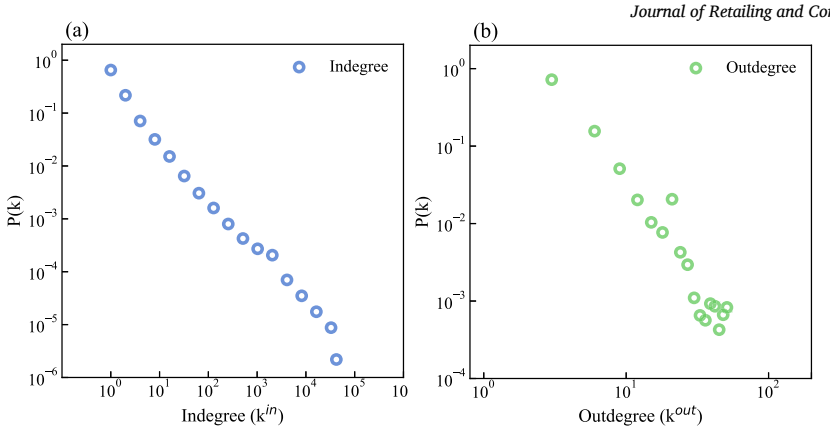

Fig. 6. NID and NOD distributions of streamers (log-bin).

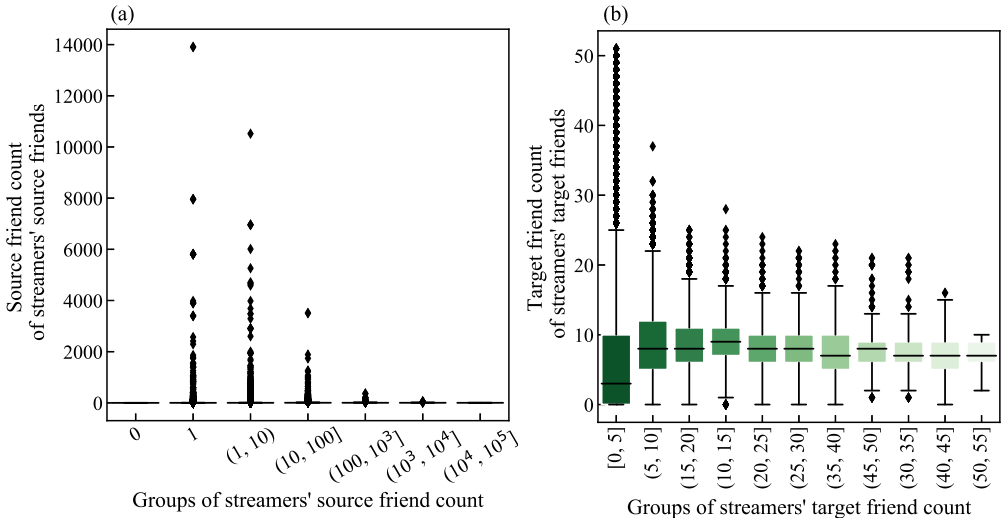

Fig. 7. Second order degree distribution of streamer. (a) Distribution of the number of target friends. (b) Distribution of the number of source friends.

Table 6
Friendship preference in G_{SCN}.

	Mean value of source streamers	Mean value of target streamers	r
SAF	10,600	4,214,391	−0.02**
SAV	12,227	1,757,547	−0.05**
SSD	7	22	−0.11**
NID	10	6,008	−0.12**
NOD	11	8	0.02**

** represents the significant value is $p < 0.01$.

Table 7
Candidate model comparison in streamer popularity prediction.

P_a prediction	AUC (SF)	F1-score (SF)	AUC (NF)	F1-score (NF)	AUC (SF&NF)	F1-score (SF&NF)
Adaboost	0.91	0.90	0.79	0.77	0.91	0.91
Decision tree	0.91	0.91	0.80	0.78	0.92	0.92
GBDT	0.92	0.92	0.81	0.80	0.94	0.93
LightGBM	**0.93**	**0.92**	**0.83**	**0.82**	**0.95**	**0.95**
Logistic	0.90	0.90	0.56	0.69	0.52	0.68
Random forest	0.92	0.91	0.80	0.79	0.92	0.92
P_r prediction	AUC (SF)	F1-score (SF)	AUC (NF)	F1-score (NF)	AUC (SF&NF)	F1-score (SF&NF)
Adaboost	0.75	0.75	0.63	0.67	0.75	0.76
Decision tree	0.76	0.77	0.64	0.67	0.75	0.77
GBDT	0.78	0.78	0.66	0.69	0.79	0.80
LightGBM	**0.78**	**0.78**	**0.68**	**0.71**	**0.81**	**0.82**
Logistic	0.63	0.53	0.54	0.15	0.58	0.37
Random forest	0.76	0.77	0.64	0.67	0.76	0.77

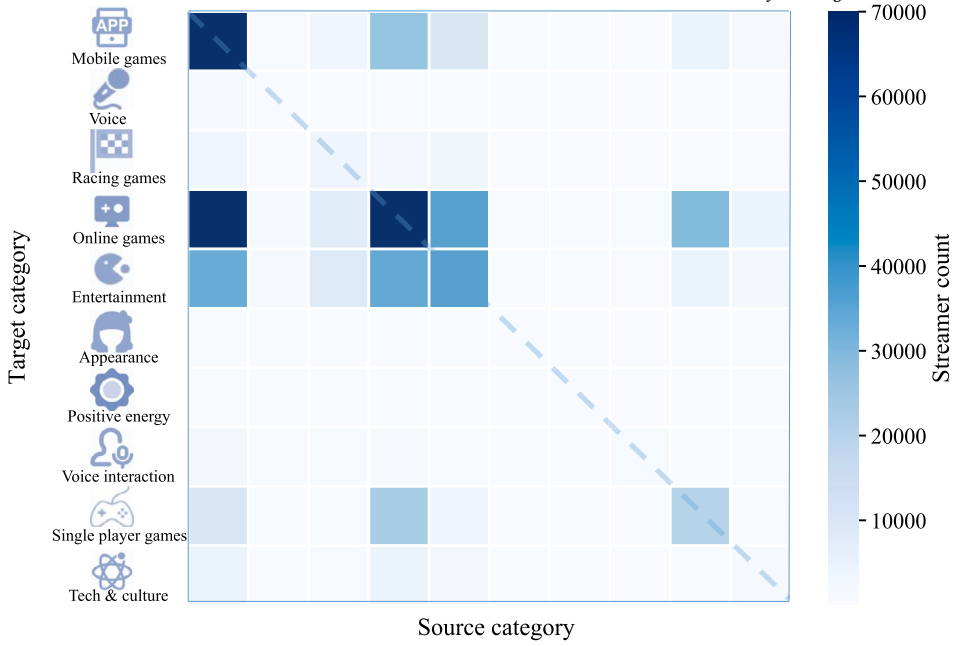

Fig. 8. Preference on streaming category.

for P_r. These results markedly surpass the popularity prediction performances reported in previous studies, with AUC value approximately 0.90 (Netzorg et al., 2021) and F1-score approximately 0.60 (Chen et al., 2021; Xi et al., 2023).

When δ is the same, using SF&NF to predict P_a can gain the maximum AUC value improvement of 5% and 31% compared to using SF and NF alone, respectively. Similarly, for predicting P_r, the maximum improvement is 9% and 25%. Compared with NF, the advantage of SF&NF is more prominent when δ is small. Compared with SF, the advantage of SF&NF is more prominent when δ is large.

Comparing prediction performances in P_a and P_r, it is evident that P_a prediction is superior. The best AUC of P_a and P_r are 0.99 and 0.95, respectively, both in the situation with SF&NF feature combination and $\delta = 39$. Importantly, our prediction model (regardless of the feature combination) can perform better in long-term popularity prediction, indicating that the extracted features can capture streamer popularity in the long-term future.

4.2.3. Feature importance analysis

The importance of streaming behaviors and streamers' co-playing relationships in predicting streamer popularity P_a and P_r is confirmed by SHAP values. Fig. 10(a) and Fig. 10(b) respectively depict the mean SHAP value contributed by each predictive feature in P_a and P_r predictions. The inner panels illustrate SHAP value distributions of top 5 features. This analysis pinpoints three crucial features for predicting both P_a and P_r: the average amount of followers (SAF), the number of weakly connected components (NWC) and the average amount of online viewers (SAV). Notably, streamer popularity is profoundly influenced not only by the sheer numbers of followers and viewers but also by their strategic positioning within the interconnected network of co-playing relationships.

We then assess the impact of important features on streamer popularity. For features SAF, SAV, SSD, NPK, NCL and NEC, we divide streamers into five groups based on ascending order of feature values, with an equal number of samples in each group. By observing the patterns of popularity variation with feature values and time intervals, we analyze the impact of these features on P_a (Fig. 11(a)) and P_r (Fig. 11(b)). The average values of P_a and P_r achieved by each feature group are included in Table A.1.

As we can see, when values of SAF, SAV, SSD, NPK and NEC increase from the group with the lowest value to the group with the highest value, there is a significant pattern of P_a increases and P_r decreases. This pattern illustrates the promoting effect of these features on P_a and weakening effect on P_r. Then, we focus on the optimal group of each feature on P_a and P_r when δ increases from 1 to 39. As δ increases, the optimal group of SAF, SAV, SSD, NPK and NEC all perform worse and worse on P_a, but better and better on P_r. This pattern means that if the feature can achieve a high P_a at the current month, its future performance on P_a will become worse. While if the feature can achieve a high P_r, its future performance on P_r will be better. Unfortunately, the impact of feature NCL on P_a and P_r is not very clear.

5. Discussion and conclusion

This study underscores the potential of utilizing streaming behaviors and streamers' co-playing relationships to make accurate streamer popularity prediction. Our approach bridges the gap in understanding the streamer popularity dynamics and provides a comprehensive disclosure of influential factors that shape streamer popularity. Through rigorous quantitative analysis and prediction experiments based on a long-time real live streaming dataset, we have discovered that the SF&NF feature combination proposed in this study significantly predicts both short- and long-term popularity of streamer. Notably, the average number of followers, viewers, and weakly connected components in the co-playing network, profoundly affect streamer's popularity dynamics.

Beyond merely applying previous methods of streaming behaviors analysis, this study delves into the initial exploration of the co-playing network among streamers on live streaming platforms, recognizing it as a intricate weave of social and interest-based connections. By applying social network theory to this context, we have demonstrated the theory's relevance and applicability in this new domain. Our analysis has shown that the streamer co-playing network holds significant predictive power for popularity, emphasizing how the synergy between collaboration and individual efforts can amplify visibility and success. This underscores the pivotal role of personalized strategies in bolstering a streamer's influence.

While our approach in predicting streamer popularity indeed achieves significant performance, it is important to acknowledge several limitations. The predictive features for streamers were derived from

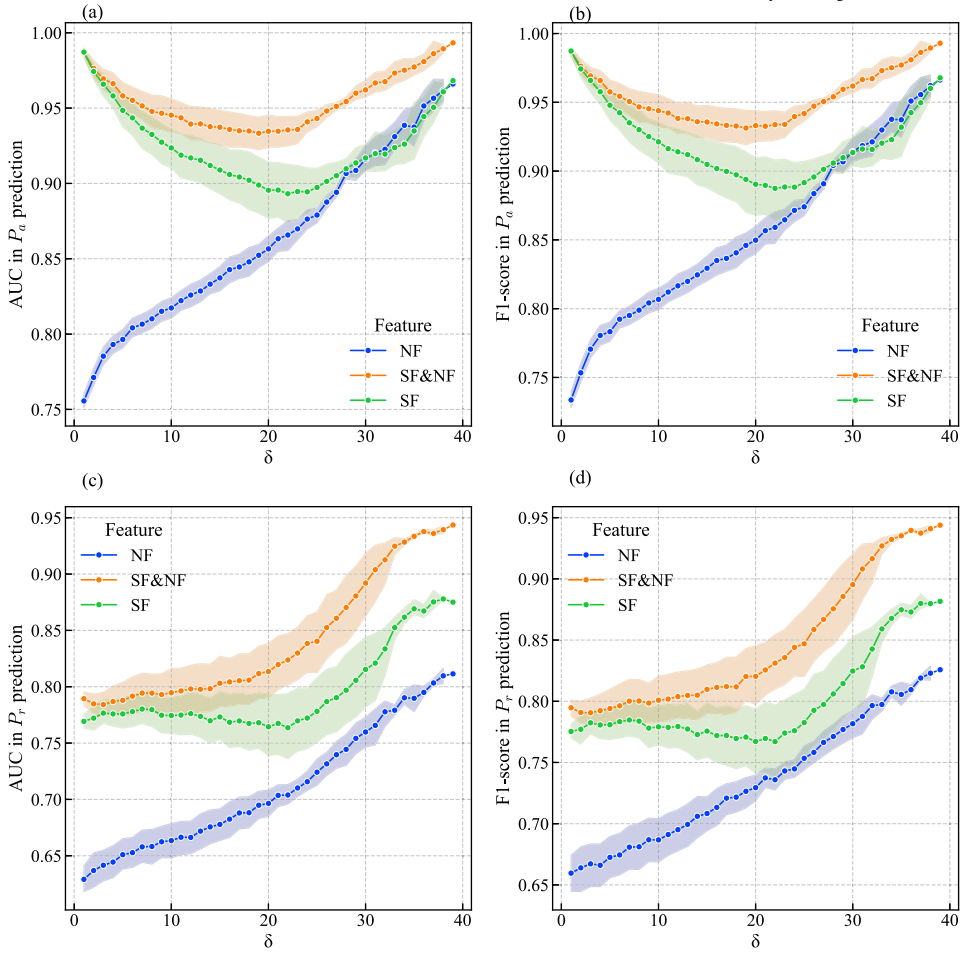

Fig. 9. Prediction results of streamer popularity. (a) AUC metric for P_a. (b) F1-score metric for P_a. (c) AUC metric for P_r. (b) F1-score metric for P_r.

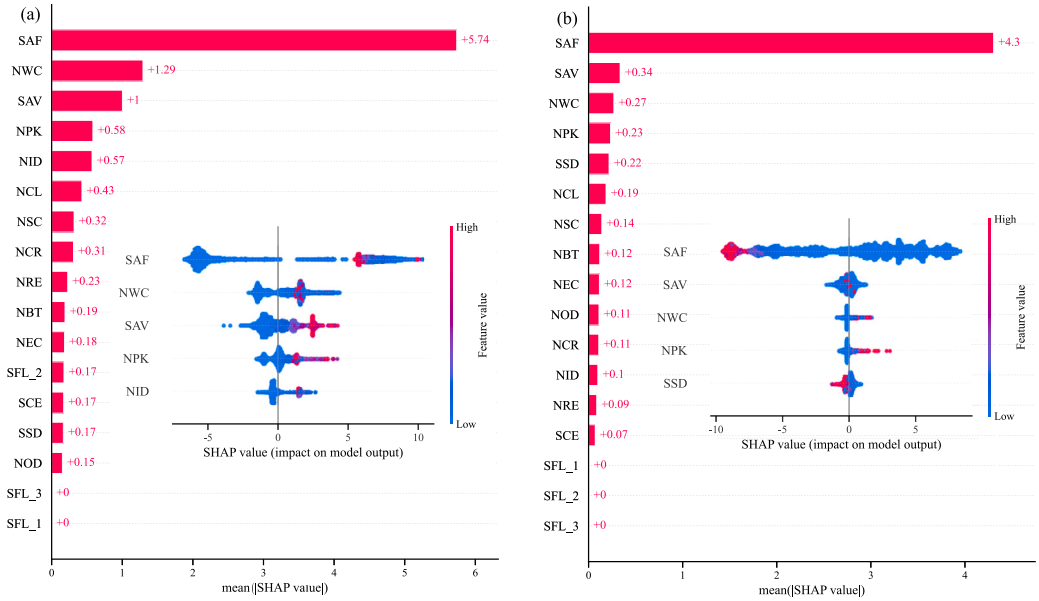

Fig. 10. SHAP values contributed by each feature of SF&NF in P_a and P_r predictions (SFL_1, SFL_2 and SFL_3 are one-hot forms of SFL).

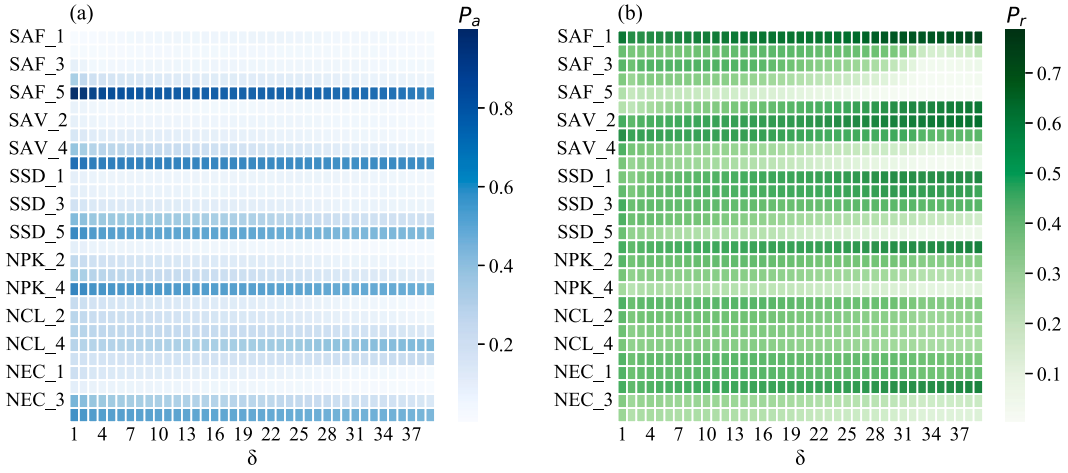

Fig. 11. Heatmap of relationships between feature groups and streamer popularity. (a) P_a (b) P_r.

a long-term, real-world dataset in the Douyu live streaming platform, which falls within the third category of live streaming. Consequently, the generalizability of our findings may be restricted to other live streaming categories. Additionally, the underlying causal mechanisms that drive these predictive features and streamer popularity are not fully revealed in this study. Furthermore, P_a and P_r are tailored to capture the dynamics of streamer popularity by focusing on the evolution of the number of followers over time, rather than the static total count of followers or viewers. This focus may limit their applicability to static metrics, such as the total count of followers or viewers, across different contexts.

Future studies may delve into the formation and evolutionary mechanisms of the streamer co-playing network, offering deeper insights into the dynamics of online content creation and promotion. By doing so, we can further understand the multifaceted nature of streamer popularity and the strategic pathways to achieving it.

CRediT authorship contribution statement

Shuhui Guo: Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Huan Chen:** Writing – review & editing, Investigation. **Bitao Dai:** Writing – review & editing, Conceptualization. **Mengning Wang:** Writing – review & editing, Methodology. **Shuo Liu:** Writing – review & editing, Supervision. **Cheng Zhang:** Writing – review & editing, Supervision. **Shan Liu:** Supervision, Funding acquisition. **Xin Lu:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (72025405, 72421002, 72088101, 72301285, 72001211, 72401039, 72030226, 21&ZD119), the National Social Science Foundation of China (22ZDA102), the Natural Science Foundation of Hunan Province (2023JJ40685, 2024JJ6069), the Humanities and Social Sciences Research Project of the Ministry of Education (24YJC630128), the Hunan Province Graduate Research Innovation Project (CX20230067), and the Key Research and Development Program of Shanxi Province (2023GXLH-036).

Appendix A. Popularity of groups divided by different features

Table A.1 shows the average values of P_a and P_r achieved by each feature group. SAF, SAV, SSD and NCL were divided into 5 groups, while NPK and NEC were divided into 4 groups. Due to the large number of small value samples for NPK and NEC, the samples from the first two groups were merged.

Table A.1
The average values of P_a and P_r in different feature groups.

Group name	Group range	P_a	P_r
SAF_1	[0, 81]	0.04	0.67
SAF_2	[82, 326]	0.02	0.37
SAF_3	[327, 1,220]	0.04	0.32
SAF_4	[1,221, 4,875]	0.12	0.22
SAF_5	[4,876, 25,808,670]	0.81	0.11
SAV_1	[0, 120]	0.03	0.46
SAV_2	[121, 569]	0.05	0.56
SAV_3	[570, 2,752]	0.08	0.50
SAV_4	[2,753, 19,545]	0.19	0.25
SAV_5	[19,546, 38,466,581]	0.66	0.19
SSD_1	[1, 3]	0.06	0.50
SSD_2	[4, 7]	0.06	0.49
SSD_3	[8, 14]	0.10	0.43
SSD_4	[14, 26]	0.30	0.30
SSD_5	[27, 31]]	0.51	0.18
NPK_1	[0, 5.4×10^{-7}]	0.04	0.53
NPK_2	(5.4×10^{-7}, 1.2×10^{-6}]	0.12	0.40
NPK_3	(1.2×10^{-6}, 3.6×10^{-6}]	0.21	0.31
NPK_4	(3.6×10^{-6}, 1.1×10^{-2}]	0.56	0.19
NCL_1	(7.3×10^{-2}, 1.9×10^{-1}]	0.11	0.42
NCL_2	(1.9×10^{-1}, 2.2×10^{-1}]	0.15	0.37
NCL_3	(2.2×10^{-1}, 2.4×10^{-1}]	0.23	0.34
NCL_4	(2.4×10^{-1}, 2.7×10^{-1}]	0.37	0.32
NCL_5	(2.7×10^{-1}, 1.0]	0.20	0.42
NEC_1	[0, 1.2×10^{-20}]	0.09	0.43
NEC_2	(1.2×10^{-20}, 1.4×10^{-20}]	0.04	0.52
NEC_3	(1.4×10^{-20}, 4.9×10^{-9}]	0.29	0.26
NEC_4	(4.9×10^{-9}, 1.0]	0.52	0.21

Data availability

Data will be made available on request.

References

Abbas, K., Shang, M., Abbasi, A., Luo, X., Xu, J.J., Zhang, Y.X., 2018. Popularity and novelty dynamics in evolving networks. Sci. Rep. 8, 6332.

Aïmeur, E., Amri, S., Brassard, G., 2023. Fake news, disinformation and misinformation in social media: a review. Soc. Netw. Anal. Min. 13, 30.

Arnett, L., Netzorg, R., Chaintreau, A., Wu, E., 2019. Cross-platform interactions and popularity in the live-streaming community. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–6.

Aubert, B., Léger, P.M., Larocque, D., 2012. Differentiating weak ties and strong ties among external sources of influences for enterprise resource planning (erp) adoption. Enterp. Inf. Syst. 6, 215–235.

Batagelj, V., Zaversnik, M., 2003. An o(m) Algorithm for Cores Decomposition of Networks, pp. 1–10. arXiv preprint cs/0310049 1.

Biau, G., Scornet, E., 2016. A random forest guided tour. Test 25, 197–227.

Bilibili, 2023. Homepage of Bilibili. https://live.bilibili.com.

Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science 350, 1073–1076.

Bollen, J., Gonçalves, B., van de Leemput, I., Ruan, G., 2017. The happiness paradox: your friends are happier than you. EPJ Data Sci. 6, 4.

Bonacich, P., 2007. Some unique properties of eigenvector centrality. Soc. Netw. 29, 555–564.

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. 30, 107–117.

Cao, Q., Shen, H., Gao, J., Wei, B., Cheng, X., 2020. Popularity prediction on social platforms with coupled graph neural networks. In: Proceedings of the 13th International Conference on Web Search and Data Mining, pp. 70–78.

Chen, C.C., Lin, Y.C., 2018. What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement. Telemat. Inform. 35, 293–303.

Chen, W.K., Chen, L.S., Pan, Y.T., 2021. A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. Appl. Soft Comput. 111, 107704.

Chicco, D., Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21, 1–13.

Churchill, B.C., Xu, W., 2016. The modem nation: a first study on Twitch.tv social structure and player/game relationships. In: 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), pp. 223–228.

Clauset, A., Newman, M.E., Moore, C., 2004. Finding community structure in very large networks. Phys. Rev. E 70, 066111.

Claypool, M., Farrington, D., Muesch, N., 2015. Measurement-based analysis of the video characteristics of Twitch.tv. In: 2015 IEEE Games Entertainment Media Conference (GEM), pp. 1–4.

Dang-Van, T., Vo-Thanh, T., Vu, T.T., Wang, J., Nguyen, N., 2023. Do consumers stick with good-looking broadcasters? The mediating and moderating mechanisms of motivation and emotion. J. Bus. Res. 156, 113483.

Daniel, D., 2022. Basic operating principles of an e-commerce system. In: E-Commerce: Concepts, Principles, and Application, pp. 327–388.

Douyu, 2022. Douyu 2022 annual report. https://www.annualreports.com/Company/douyu-international-holdings-limited.

Douyu, 2023a. Homepage of Douyu. https://www.douyu.com.

Douyu, 2023b. Introduction of neighbor function in douyu. https://www.douyu.com/cms/gong/201901/30/9759.shtml.

Farrington, D., Muesch, N., 2015. Analysis of the characteristics and content of Twitch live-streaming. BS qualifying project report. Worcester Polytechnic Institute 1, pp. 1–64.

Freeman, L., 2004. The development of social network analysis. In: A Study in the Sociology of Science 1, pp. 159–167.

Freeman, L.C., 1979. Centrality in social networks: conceptual clarification. Soc. Netw. 1, 215–239.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55, 119–139.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38, 367–378.

Gao, J., Zhang, Y.C., Zhou, T., 2019. Computational socioeconomics. Phys. Rep. 817, 1–104.

Granovetter, M.S., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380.

Guo, S., Lu, X., 2020. Live streaming: data mining and behavior analysis. Acta Phys. Sin. 69, 088908.

Guo, Y., Zhang, K., Wang, C., 2022. Way to success: understanding top streamer's popularity and influence from the perspective of source characteristics. J. Retail. Consum. Serv. 64, 102786.

Gurjar, O., Bansal, T., Jangra, H., Lamba, H., Kumaraguru, P., 2022. Effect of popularity shocks on user behaviour. In: Proceedings of the International AAAI Conference on Web and Social Media, pp. 253–263.

Hamilton, W.A., Garretson, O., Kerne, A., 2014. Streaming on Twitch: fostering participatory communities of play within live mixed media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1315–1324.

Hu, F., Qiu, L., Wei, S., Zhou, H., Bathuure, I.A., Hu, H., 2024. The spatiotemporal evolution of global innovation networks and the changing position of China: a social network analysis based on cooperative patents. R & D Manag. 54, 574–589.

Hu, M., Zhang, M., Wang, Y., 2017. Why do audiences choose to keep watching on live video streaming platforms? An explanation of dual identification framework. Comput. Hum. Behav. 75, 594–606.

Huang, L., Liu, B., Zhang, R., 2024. Channel strategies for competing retailers: whether and when to introduce live stream? Eur. J. Oper. Res. 312, 413–426.

Huya, 2023. Homepage of Huya. https://www.huya.com.

Jia, A.L., Shen, S., Epema, D.H., Iosup, A., 2016. When game becomes life: the creators and spectators of online game replays and live streaming. ACM Trans. Multimed. Comput. Commun. Appl. 12, 1–24.

Jusup, M., Holme, P., Kanazawa, K., Takayasu, M., Romić, I., Wang, Z., Geček, S., Lipić, T., Podobnik, B., Wang, L., et al., 2022. Social physics. Phys. Rep. 948, 1–148.

Kaytoue, M., Silva, A., Cerf, L., Meira Jr, W., Raïssi, C., 2012. Watch me playing, I am a professional: a first study on video game live streaming. In: Proceedings of the 21st International Conference on World Wide Web, pp. 1181–1188.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30, 1–9.

Kim, J., Park, K., Song, H., Park, J.Y., Cha, M., 2020. Learning how spectator reactions affect popularity on Twitch. In: 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), pp. 147–154.

Kim, M., Kim, H.M., 2022. What online game spectators want from their twitch streamers: flow and well-being perspectives. J. Retail. Consum. Serv. 66, 102951.

Kuaishou, 2023. Kuaishou Technology's performance announcement for the three months ended March 31, 2023. https://www1.hkexnews.hk/listedco/listconews/sehk/2023/0522/2023052200218_c.pdf.

LaValley, M.P., 2008. Logistic regression. Circulation 117, 2395–2399.

Lee, J.K., Choi, J., Kim, C., Kim, Y., 2014. Social media, network heterogeneity, and opinion polarization. J. Commun. 64, 702–722.

Lesser, O., Hayat, T., Elovici, Y., 2017. The role of network setting and gender in online content popularity. Inf. Commun. Soc. 20, 1607–1624.

Li, L., Chen, X., Zhu, P., 2024. How do e-commerce anchors' characteristics influence consumers' impulse buying? An emotional contagion perspective. J. Retail. Consum. Serv. 76, 103587.

Li, Z., Kaafar, M.A., Salamatian, K., Xie, G., 2016. Characterizing and modeling user behavior in a large-scale mobile live streaming system. IEEE Trans. Circuits Syst. Video Technol. 27, 2675–2686.

Lim, J.S., Choe, M.J., Zhang, J., Noh, G.Y., 2020. The role of wishful identification, emotional engagement, and parasocial relationships in repeated viewing of live-streaming games: a social cognitive theory perspective. Comput. Hum. Behav. 108, 106327.

Lin, Y., Yao, D., Chen, X., 2021. Happiness begets money: emotion and engagement in live streaming. J. Mark. Res. 58, 417–438.

Ling, C.X., Huang, J., Zhang, H., et al., 2003. AUC: a statistically consistent and more discriminating measure than accuracy. In: International Joint Conference on Artificial Intelligence, pp. 519–524.

Liu, J., Xu, Q., Sun, Z., 2022. Joint optimization decision of online retailers' pricing and live-streaming effort in the postepidemic era. Complexity 2022, 1–11.

Liu, W., Sidhu, A., Beacom, A.M., Valente, T.W., 2017. Social Network Theory, vol. 1. John Wiley & Sons, Inc., pp. 1–12.

Lu, S., Zhao, Y., Chen, Z., Dou, M., Zhang, Q., Yang, W., 2021. Association between atrial fibrillation incidence and temperatures, wind scale and air quality: an exploratory study for Shanghai and Kunming. Sustainability 13, 5247.

Lundberg, S., 2017. A unified approach to interpreting model predictions. preprint arXiv: 1705.07874. 1–10.

Luo, H., Cheng, S., Zhou, W., 2021. The factors influencing sales in online celebrities' live streaming. In: 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE), pp. 233–237.

Luo, X., Cheah, J.H., Hollebeek, L.D., Lim, X.J., 2024. Boosting customers' impulsive buying tendency in live-streaming commerce: the role of customer engagement and deal proneness. J. Retail. Consum. Serv. 77, 103644.

Lv, J., Yao, W., Wang, Y., Wang, Z., Yu, J., 2022. A game model for information dissemination in live streaming e-commerce environment. Int. J. Commun. Syst. 35, 1–14.

Lykousas, N., Gómez, V., Patsakis, C., 2018. Adult content in social live streaming services: characterizing deviant users and relationships. In: 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 375–382.

Ma, X., Chen, H., Lang, X., Li, T., Wu, N., Duong, B., 2024. Research on the impact of streamers' linguistic emotional valence on live streaming performance in live streaming shopping environments. J. Retail. Consum. Serv. 81, 104040.

Myles, A.J., Feudale, R.N., Liu, Y., Woody, N.A., Brown, S.D., 2004. An introduction to decision tree modeling. J. Chemom. 18, 275–285.

Nascimento, G., Ribeiro, M., Cerf, L., Cesário, N., Kaytoue, M., Raïssi, C., Vasconcelos, T., Meira, W., 2014. Modeling and analyzing the video game live-streaming community. In: 2014 9th Latin American Web Congress, pp. 1–9.

Netzorg, R., Arnett, L., Chaintreau, A., Wu, E., 2021. Popfactor: live-streamer behavior and popularity. In: Proceedings of the International AAAI Conference on Web and Social Media, pp. 432–442.

Newman, M.E., 2003. The structure and function of complex networks. SIAM Rev. 45, 167–256.

Newman, M.E., 2006. Modularity and community structure in networks. Proc. Natl. Acad. Sci. 103, 8577–8582.

Park, H.J., Lin, L.M., 2020. The effects of match-ups on the consumer attitudes toward Internet celebrities and their live streaming contents in the context of product endorsement. J. Retail. Consum. Serv. 52, 101934.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W., 2018. Influence analysis in social networks: a survey. J. Netw. Comput. Appl. 106, 17–32.

Pires, K., Simon, G., 2014. Dash in twitch: adaptive bitrate streaming in live game streaming platforms. In: Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming, pp. 13–18.

Pires, K., Simon, G., 2015. Youtube live and Twitch: a tour of user-generated live streaming systems. In: Proceedings of the 6th ACM Multimedia Systems Conference, pp. 225–230.

Rehman, A.U., Jiang, A., Rehman, A., Paul, A., Din, S., Sadiq, M.T., 2023. Identification and role of opinion leaders in information diffusion for online discussion network. J. Ambient Intell. Humaniz. Comput., 1–13.

Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. 105, 1118–1123.

StreamScheme, 2023. Twitch demographic & growth statistics. https://www.streamscheme.com/twitch-statistics/.

Strogatz, S.H., 2001. Exploring complex networks. Nature 410, 268–276.

Su, X., Xue, S., Liu, F., Wu, J., Yang, J., Zhou, C., Hu, W., Paris, C., Nepal, S., Jin, D., et al., 2022. A comprehensive survey on community detection with deep learning. In: IEEE Transactions on Neural Networks and Learning Systems.

Sun, J., Sarfraz, M., Ivascu, L., Han, H., Ozturk, I., 2024. Live streaming and livelihoods: decoding the creator economy's influence on consumer attitude and digital behavior. J. Retail. Consum. Serv. 78, 103753.

Tang, Y., Sun, L., Luo, J., Zhong, Y., 2006. Characterizing user behavior to improve quality of streaming service over P2P networks. In: Advances in Multimedia Information Processing-PCM 2006: 7th Pacific Rim Conference on Multimedia, pp. 175–184.

Tian, Y., Frank, B., 2024. Optimizing live streaming features to enhance customer immersion and engagement: a comparative study of live streaming genres in China. J. Retail. Consum. Serv. 81, 103974.

TikTok, 2023. Homepage of TikTok. https://www.tiktokus.info/.

TrendInsight, 2023. 2022 Douyin e-commerce intellectual property protection report. https://trendinsight.oceanengine.com/arithmetic-report/detail/913.

Tu, W., Yan, C., Yan, Y., Ding, X., Sun, L., 2018. Who is earning? Understanding and modeling the virtual gifts behavior of users in live streaming economy. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), pp. 118–123.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151.

Wang, X., Tian, Y., Lan, R., Yang, W., Zhang, X., 2018. Beyond the watching: understanding viewer interactions in crowdsourced live video broadcasting services. IEEE Trans. Circuits Syst. Video Technol. 29, 3454–3468.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press.

Wongkitrungrueng, A., Assarut, N., 2020. The role of live streaming in building consumer trust and engagement with social commerce sellers. J. Bus. Res. 117, 543–556.

Xi, D., Tang, L., Chen, R., Xu, W., 2023. A multimodal time-series method for gifting prediction in live streaming platforms. Inf. Process. Manag. 60, 103254.

Xing, Y., Wang, X., Qiu, C., Li, Y., He, W., 2022. Research on opinion polarization by big data analytics capabilities in online social networks. Technol. Soc. 68, 101902.

Yang, L., Dong, J., Yang, W., 2024. Analysis of regional competitiveness of China's cross-border e-commerce. Sustainability 16, 1007.

Yang, W., Pan, L., Ding, Q., 2023. Dynamic analysis of natural gas substitution for crude oil: scenario simulation and quantitative evaluation. Energy 282, 128764.

Ye, F., Ji, L., Ning, Y., Li, Y., 2024. Influencer selection and strategic analysis for live streaming selling. J. Retail. Consum. Serv. 77, 103673.

Zhang, C., Liu, J., 2015. On crowdsourced interactive live streaming: a Twitch.tv-based measurement study. In: Proceedings of the 25th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video, pp. 55–60.

Zhang, Q., Wang, Y., Ariffin, S.K., 2024. Consumers purchase intention in live-streaming e-commerce: a consumption value perspective and the role of streamer popularity. PLoS ONE 19, e0296339.

Zhang, Y., Wang, L., Zhu, J.J., Wang, X., Pentland, A.S., 2021. The strength of structural diversity in online social networks. Research 2021, 9831621.

Zhao, J., Ma, M., Gong, W., Zhang, L., Zhu, Y., Liu, J., 2017. Social media stickiness in mobile personal livestreaming service. In: 2017 IEEE/ACM 25th International Symposium on Quality of Service (IWQoS), pp. 1–2.

Zhao, K., Hu, Y., Hong, Y., Westland, J.C., 2019. Understanding characteristics of popular streamers on live streaming platforms: evidence from Twitch.tv. J. Assoc. Inf. Syst. 22, 1076–1098.

Zhao, K., Lu, Y., Hu, Y., Hong, Y., 2023. Direct and indirect spillovers from content providers' switching: evidence from online livestreaming. Inf. Syst. Res. 34, 847–866.

Zhou, L., Wu, W.p., Luo, X., 2007. Internationalization and the performance of born-global smes: the mediating role of social networks. J. Int. Bus. Stud. 38, 673–690.

Zhu, L., Liu, N., 2023. Game theoretic analysis of logistics service coordination in a live-streaming e-commerce system. Electron. Commer. Res. 23, 1049–1087.

Zhu, Z., Yang, Z., Dai, Y., 2017. Understanding the gift-sending interaction on live-streaming video websites. In: International Conference on Social Computing and Social Media, pp. 274–285.