Behind the Game: Exploring the Twitch Streaming Platform

Jie Deng, Felix Cuadrado, Gareth Tyson, Steve Uhlig

Queen Mary University of London, UK
email: {j.deng,felix.cuadrado,gareth.tyson,steve.uhlig}@qmul.ac.uk

***Abstract*—Twitch is a streaming platform that lets users broadcast their screen whilst playing games. People can share their game experience and interact with others in real time. Twitch has now become the fourth largest source of peak Internet traffic in the US. This paper explores the unique nature of this platform over a 11 month dataset. We find that Twitch is very different to existing video platforms, with a small number of games consistently achieving phenomenal dominance. We find a complex game ecosystem combining consistently popular games over years, newly released games enjoying bursts of popularity, and even old games appearing on the platform. Despite a strong skew of views across channels, the top ranked channels, although taking a significant share of the viewers, exhibit unexpectedly high churn. The reason behind this churn lies within another unique feature of this ecosystem, namely tournaments, live events that last for a limited amount of time but are capable of attracting a huge share of views when they take place, as well as dominate the views of the related games. Overall, our work reveals a complex and rich ecosystem, very different from existing user generated content platforms.**

I. Introduction

Video games are a form of entertainment enjoyed by a diverse, worldwide consumer base. Traditionally, gaming has been a pastime enjoyed by those who choose to *play*. Recently, however, gaming has become a spectator form of entertainment. Major eSports tournaments such as DreamHack are enjoyed by millions of online viewers, yet have seen little attention from mainstream media outlets: enter **Twitch**. Twitch is a large-scale video streaming platform used (almost) exclusively for live game broadcasting. It allows broadcasters to construct *channels* through which they can stream their gameplay to the world. Popular types of channels include amateurs broadcasting their gameplay, competitive eSports tournaments with commentaries, coaching game sessions, charity marathon events, and, finally, experimental large-scale cooperative events where games are played collectively. Twitch's popularity is undeniable, recently being reported as the 4^{th} largest website in the U.S. by peak traffic [3].

Despite this, we are yet to gain a comprehensive understanding of how users, games and broadcasters operate in this new environment. Twitch holds a number of novel features. Of most interest is the introduction of a new object of interest, the *game*. Whereas in traditional content services, channels tend to stream unique content, in Twitch the same games are streamed by many different people. This unusual property raises many questions in terms of how viewers are spread amongst these different games. In this paper we expose the Twitch platform by analyzing live viewing figures collected over an 11 month period. First, we detail our dataset and provide insight into the growing scale of Twitch (§ II). We witness significant viewing peaks, approaching one million simultaneous viewers. We show the rich ecosystem of games streamed through Twitch (§ III), exposing the impact of game features such as the release date and the genre. We discover a significant number of gaming events being broadcast through the platform (§ IV) bringing Twitch closer to a live sports platform than a user-generated content one. Finally we look at the individual broadcasters (§ V), revealing an *extremely* skewed popularity distribution, and significant dynamics in popularity.

II. Overview of Twitch and Dataset

Twitch is a live streaming video platform focusing on gaming. It allows users to broadcast themselves playing games to others, who tune in via a web interface. Streams include playthroughs of games by amateur users, and large-scale broadcasts of eSports competitions. Users in Twitch can take one of two roles, broadcaster or viewer. A *broadcaster* is somebody who streams their game play via a dedicated *channel*, whilst a *viewer* is somebody who watches the channel. Each streamer is limited to one live channel, which is only online for a fixed period of time when the player is broadcasting. To facilitate communication, the channels are enabled with an in-built chat room to allow the users (both broadcasters and viewers) to interact with each other.

To study Twitch, we have collected a dataset spanning 11 months from February 2014 to December 2014. This was collected using the public Twitch REST API, which we contacted repeatedly to extract available metadata. For each channel, we retrieved the game being played, the number of viewers, the title of the channel and the broadcaster's name. We repeated this every 15 minutes to build a time series of metadata for every channel in the system, totalling 323 million channel samples, which includes 5.2 million unique broadcasters.

We begin our analysis by inspecting the overall trends of Twitch as a service. On average across all 15 minute snapshots, there are 9100 broadcasters feeding 362k viewers.

We have observed significant growth over the measurement period, with viewers increasing by 25% and channels by 45% when comparing monthly views. The service shows significant dynamic behaviour, with August 2014 reaching a peak of 934k simultaneous viewers.

III. Exploring Games

The previous section has presented the Twitch platform and highlighted its overall growth. As a unique aspect of Twitch, we first investigate *games*.

A. Which games are played?

A basic and fundamental question to understand Twitch is what games are actually broadcast? To explore this, we use the 'game' field returned from the Twitch API. We observe a total number of 127,497 different name strings. The reason for this high number is that the field can be set manually by broadcasters, making it unreliable. To obtain something close to a ground truth, we collected game data from 'TheGamesDB.net' [2], an open, online database for video games, and 'GiantBomb' [1], an American video game and wiki. We match exact names (in lower case, removing spaces) with these two databases to filter out bogus 'game' field entries. This leaves a surprisingly large number (21k) of unique games played on Twitch, accounting for 95% of all views. We decided against more sophisticated name matching techniques (such as Levenshtein distance) because of the predominance of game sequels that would be wrongly resolved.

We next inspect the respective viewing and broadcast figures attained by these 21k games. We therefore begin by inspecting which games are most frequently broadcast and viewed in Twitch. Much like music artists vie for airtime on radio, game developers may wish to see their games effectively promoted on Twitch. We observed 41 games that were streamed in Twitch prior to their release date (often through official channels from the developers). This suggests that developers are using Twitch to gather feedback and generate interest on upcoming games.

Table I presents the top 10 (0.04%) games found in Twitch. It can be seen that there is a strongly uneven distribution of viewers and broadcasters across these games. The top 10% of games collect 95% of all viewers (far higher than seen with other types of content objects, *e.g.,* Youtube [7] or VoD [4]). Alone, the top 10 covers 64% of all viewers. This leaves a large number of games to languish with very little attention: 92.8% (≈19.6k) of the games have fewer than 100 viewers on average. Twitch has therefore become an ecosystem largely driven by several extremely popular games. Table I also presents the percentage of time each game spent in the top 10. We find that these ranks are remarkably stable; the most notable example is League of Legends, which held the top rank for 90.4% of our dataset (staying in the top 10 for 99.5%).

Rank	Name	% of Viewers	% of Channels	% of Uniq Streamers	% in Top 10
1	League of Legends	29.1	14.9	11.9	99.5
2	DOTA 2	11	3.2	2.8	99.1
3	Hearthstone: Heroes of Warcraft	8	2	3.3	98.9
4	Counter Strike: Global Offensive	6.2	3.9	5.2	91.67
5	Minecraft	3.5	4.2	8.5	86.2
6	Starcraft II: HeartoftheSwarm	3.0	1.0	0.6	68.3
7	WorldofWarcraft: Mists of Pandaria	2.2	2.8	1.9	71.3
8	DiabloIII: ReaperofSouls	1.9	1.9	1.7	33.5
9	DayZ	1.5	1.5	1.6	28.7
10	Call of Duty: Ghosts	1.0	2.2	4.1	20.7

TABLE I: Top 10 games in Twitch.

B. Game features

The previous section has highlighted that a few prominent games dominate Twitch. A key question is what are their characteristics? To explore this, we augment our games with metadata obtained by the API from GamesDB [2] and GiantBomb [1]. We focus on release date and genre.

1) Release date: A common property held across many content repositories is user preference for recent releases, as consumers tend to constantly seek out new stimulation [16], [19]. We investigate if Twitch shares this property. Specifically, we inspect if new game releases gain popularity and manage to steal viewers away from older and more established games.

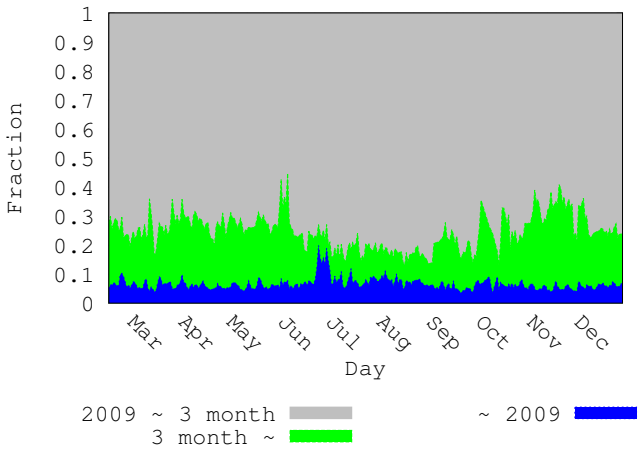

Fig. 1: Fraction of views across games with different release times: between 2009 to three months before the observation point, within three months and games released before 2009.

Figure 1 breaks down the viewings across games split into 3 bins of release date. For the games released every year, we calculate the number of viewers each game collects in each snapshot. We set our main boundary in 2009, as all the top games in Table I were released between that year and today. The popularity of these games is shown clearly in the data; games released between 2009 and 3 months before the data point consistently attain ≈75% of the viewer share. Thus,

game popularity is not as ephemeral as seen in other domains, which frequently see newly released content at the top of the rankings [16], [19]. In Twitch, fresh games (*i.e.,* <3 months) manage to attract only 18% of viewers; this can be contrasted with YouTube, in which new videos constitute 80% of the top list [7]. Surprisingly, we also witness a notable share (7%) for very old games released prior to 2009. This fraction of the share is bursty, with peaks of 40% of the total platform views, and significant variation in the individual games that contribute to these shares. We will revisit this burstiness of recent games later in the paper.

2) Genres: We further explore the games in Twitch by classifying them into different genres (based on information from the GiantBomb API). Figure 2 shows the fraction of views taken by the top genres across the measurement period. We find that MOBA (*i.e.,* multiplayer online battle arena) is the dominant genre, gathering 40% of the views (as expected with the top 2 games belonging to that genre. Nonetheless, high variance can be seen with significant changes across the measurement period. One of the most abrupt changes is the big drop in the aggregate share of these popular genres at the beginning of June. This corresponds to the broadcasting of the "E3 press conferences" (the main video games festival) on June 9th 2014, which shifted viewers away from the usual content genres enjoyed by users. It can also be seen that notable shifts occur with views transferring between these popular genres; for example, role playing games lost a significant share of their viewing figures to other genres in mid-April (this was caused by the fading popularity of a viral phenomenon, Twitch Plays Pokemon [11]). This is just one example of the complex interactions between simultaneous events occurring in Twitch.

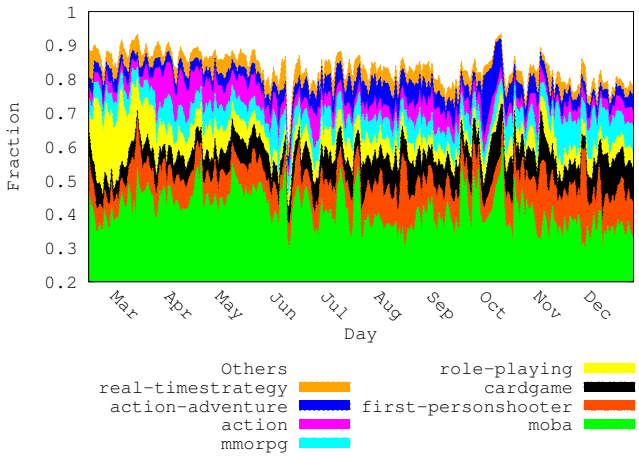

Fig. 2: Fraction of daily views for each genre.

IV. Exploring Tournaments

The above exploration has revealed a particularly important type of channel: tournament broadcasts. The real time nature of Twitch makes it ideal for broadcasting live events taking place anywhere in the world. This is the case for the burgeoning eSports competitive scene.

We have manually extracted the key events streamed via Twitch throughout 2014. In total, we have identified 56 events, some lasting multiple days. 53 are eSports tournaments (we left out of this analysis competitive leagues and preliminary phases); the remaining 3 are a charity event, the famous E3 press conference, and the TwitchPlaysPokemon phenomenon [11]. The total number of days from our dataset with an event running are 150 (47% of the considered days).

Figure 3 presents each event's share of the overall viewing figures over time (each bar is an event colour coded by the game played[1]). The huge impact of event broadcasters is undeniable; many exceed 20% of the daily viewers, making them the top ranked channel (Figure 7). Whereas the average top ranked channel's viewing share is 8.9% (daily), this can increase up to 30.5% during events. In absolute terms we observe close to a million simultaneous viewers in the whole platform. Thus, Twitch resembles other live sports platforms, with spikes during key events [20].

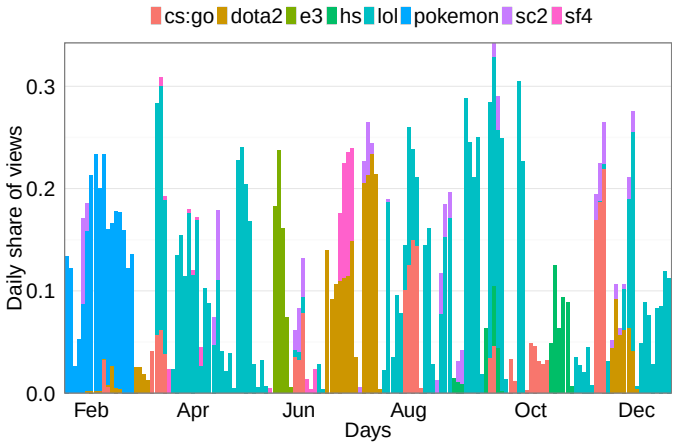

Fig. 3: Fraction of viewing figures collected by tournaments per day. Tournaments are grouped per game: Counter Strike: Global Offensive, DOTA 2, E3 (gaming trade show), Hearthstone, League of Legends, Pokemon, Starcraft 2, and Street Fighter 4.

We also inspect the relationship between events and games (see color code in Figure 3). Tournaments playing League of Legends (LoL) achieve the highest and most frequent peaks. Numerous events playing other less popular games also reach comparable levels of popularity over time. Again, we see that event broadcasters gather the attention of Twitch viewers, even if they are streaming not so popular games.

Figure 4 shows a CDF of the share of viewers garnered by the tournaments (on a per game basis); we take samples from every 15 minutes. Overall, we observe that tournaments have a substantial impact on every game (with a lesser extent for DOTA2, which incorporates and incentivises watching

[1]We treat E3 as a separate game, although this is actually a trade show.

tournaments through their own platform). In the most extreme case, for Street Fighter 4, the tournament can reach 100% of all viewers of that game. The sometimes large share of viewers captured by tournaments for a given game explains why the companies behind these games provide so much support for the tournaments, *e.g.,* in the form of prize money.

Overall, tournaments nicely illustrate the complexity of the Twitch platform, that lies at the crossing between viewers, broadcasters, as well as different types of companies (*e.g.,* game development and advertising).

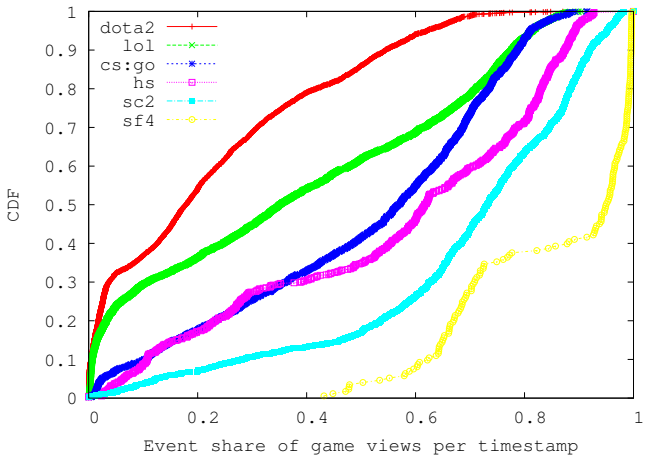

Fig. 4: CDF of the share of game views gathered by events in each snapshot.

V. Exploring Channels

In this section, we explore the relationship between viewers, games and channels.

A. Channels ≠ Games

We have previously shown that some games (*e.g.,* League of Legends) are extremely popular in Twitch. Hence, it is natural to expect a similar skew in channels. Figure 5 presents the distribution of viewers across channels (x-axis provides the rank of the channel), confirming that much like with games, channel popularity is highly skewed towards a few prominent broadcasters. The top 10 broadcasters alone collect 16% of all views. In each 15 minute snapshot, the top 1% of channels collect 70% of the viewers, whilst the top 10% collect 93% of all viewers. In contrast, the remaining channels get very poor viewing figures (62% of them below 1 viewer on average). The majority of broadcasters therefore make little impact on the overall system — Twitch is a system dominated by a tiny minority. Figure 5 also provides the distribution of viewers across subscription channels (whitelisted by Twitch under the "partners" scheme). Subscription channels provide the opportunity for viewers to support their favourite broadcasters by paying an optional monthly fee. Subscription is shared between Twitch and the broadcaster. We observe from Figure 5 that subscription channels are a representative subset of the most popular channels in terms of viewing figures and skew. Consistent with their purpose, we also observe that no low ranked channel is among the subscription channels.

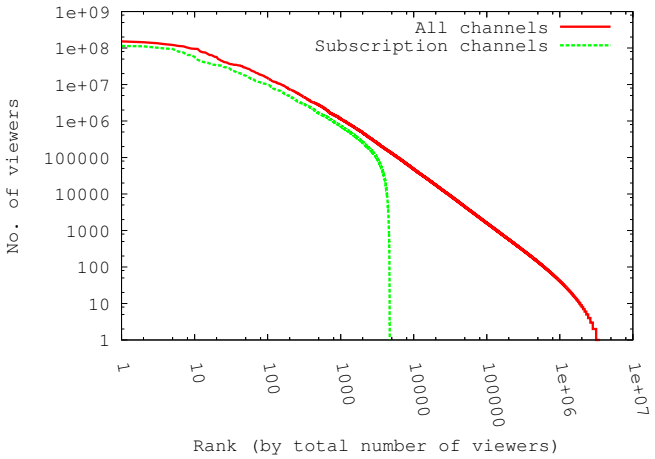

Fig. 5: Distribution of viewers per channel.

Now we come to the question of the "relationship" between games and channels. To answer this, Figure 6 shows how many games have been played by each channel over its lifespan. Substantial diversity can be observed, with some broadcasters playing in excess of 100 games. As such, channels ≠ games: often channels broadcast many different games. Figure 6 also breaks down broadcasters into different popularity groups; it can be seen that broadcasters with high viewing figures tend to play many games. For channels that garner between 100 and 10k viewers, only 30.4% play a single game. Moreover, roughly half of the popular subscription channels play at least ten games. That said, curiously, extremely popular channels (>10k viewers) tend to play fewer games. This indicates that there is no direct relationship between gaining viewers and the number of games played. Instead, *popularity appears inherent to the broadcaster, rather than the games they play*. This is a key observation when trying to understand Twitch, particularly when considering companies wishing to advertise games (*e.g.,* by getting popular broadcasters to promote it by playing).

B. Churn in top channels

So far, we have shown that the top 10 channels garner a significant share of the viewers. We next see how this share varies over time. Figure 7 shows, for each week, the share of views accumulated by the top 10 channels. Despite seeing a significant concentration of views, the total share rarely surpasses 30% of the weekly views. While Figure 7 might give the impression that the share of viewers captured by the channels showed in this top 10 is stable, this is actually not the case. Indeed, there is significant churn in which actual channels are those that are ranked within this top 10 over time. Figure 8 shows a visual representation of the top 10 channels every week. Each coloured symbol depicts a different channel, with lines showing how they move up and down the rankings. The top ranked channels (*i.e.,* ranked first) are extremely consistent; for example, although it only streams for 80% of

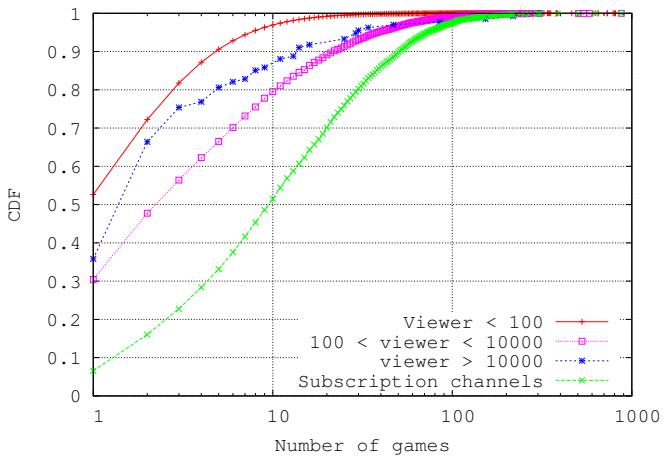

Fig. 6: Number of games played by each channel. Channels are separated into popularity groups by average number of viewers.

the observed period, 'riotgames' is the top channel 60% of the time. This consistency is unusual considering the short-lived nature of popularity in other domains, *e.g.,* YouTube. However, as can be seen thanks to the solid lines that connect every coloured symbol, most of the channels exhibit a significant amount of churn both in terms of their presence within the top 10, as well as their actual rank. The "randomised" look of Figure 8 visualises well the amount of churn taking place within top ranked channels.

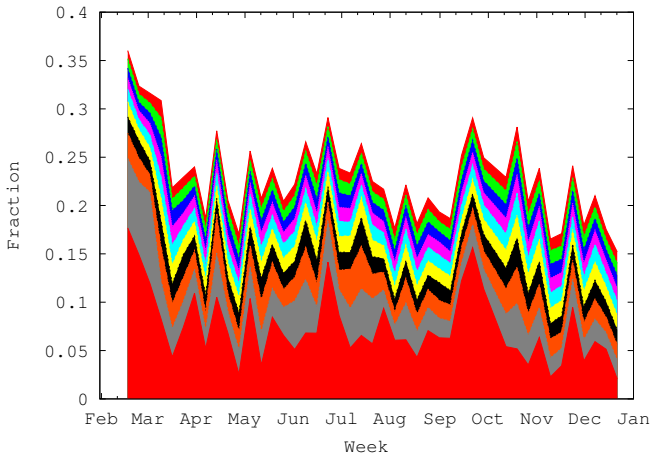

Fig. 7: Platform share of the top 10 channels each week by number of viewers.

A closer inspection of the weekly top channel over the measurement period reveals that 95% of the time the top ranked channel is *riotgames*, the official channel for League of Legends, dedicated to broadcasting official eSports tournaments. Individual broadcasters (as opposed to corporations) regularly appear in the top list of channels, although clearly the most popular channels are broadcasted live events. Indeed, Twitch is a natural platform for this type of content (as viewers can interact while they watch, and live streaming allows to broadcast the event as it happens).

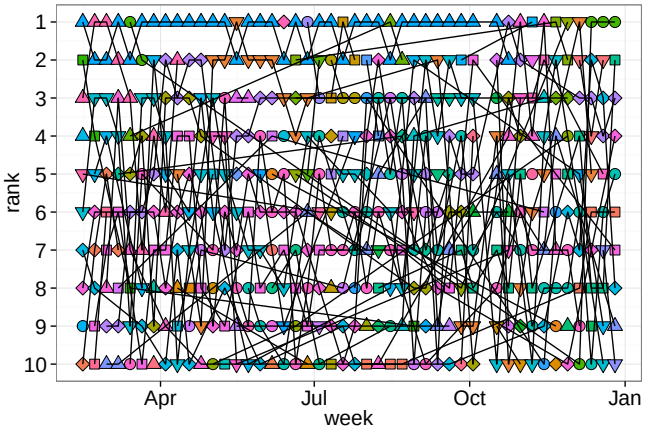

Fig. 8: Top 10 channels each week by number of viewers. Lines connect reoccurring instances of the same channel.

VI. Implications

We have observed several unique characteristics that differentiate Twitch from other large-scale streaming services. Twitch constitutes a novel form of multimedia, introducing the concept of *game* as a new multimedia object, possessing unique popularity characteristics (distinct from traditional channels/broadcasters). Exploiting the knowledge introduced by these games is a ripe area of further exploration. For instance, use of game information for things like recommender engines would be highly beneficial. Similarly, the observation that many users watch the same games (potentially played by different streamers) may indicate that such recommender engines could easily guide users to nearby (low-cost) streamers, rather than distant (high-cost) streamers playing the same game [14]. A key finding of our work is the preference users have for particular users, which suggests that this would have to be done with caution: players are certainly not all equal.

Unlike many other streaming services, Twitch also benefits from highly predictable popularity trends regarding its flash crowds. This is because Twitch flashcrowds are almost always driven by scheduled tournaments. In contrast, open video services (e.g., YouTube) are often left unaware of new up and coming events that may generate flashcrowds. Naturally, this is driven by the existence of many different unknown video providers and genres. This could make Twitch's infrastructural provisioning much more straightforward than other platform's. Similarly, outside of tournaments, a relatively small number of extremely popular broadcasters share $> 90\%$ of online viewers. This means that predicting the behaviour of this small number of broadcasters could lead to similar benefits when provisioning infrastructure. Such broadcasters could even be asked to release schedules, so Twitch could know in advance how to provision its infrastructure. The value of this predictability cannot be underestimated.

Twitch *delay* settings allow streamers to configure the delay of a stream before transmission. This feature was introduced to prevent multiplayer cheating, but could also be exploited by the infrastructure. Specifically, the delay parameter could be used to enable caching and staggering of delivery across users. This would be, in essence, consensual buffering.

Another key implication is the impact that Twitch has on the games industry more generally. Curiously, we find little correlation between traditional games ratings and the popularity of games on Twitch. This suggests that users watching and playing games may have different needs. This has clear ramifications for the games industry. Most notably, it introduces a powerful means to gather rapid feedback. This is clearly something that is already being explored, as we found 41 games pre-released on Twitch, as part of the promotion activity. We envisage games will increasingly be designed with Twitch-like broadcast in-mind. We already see this with many prominent games having in-built support for Twitch. Finally, the growing popularity of Twitch should be treated as a wake-up call for TV broadcast outlets. Broadcast is an extremely efficient medium for popular content and, as such, we argue it is only a matter of time before gameplay content becomes commonplace on TV broadcast, much like other sports.

VII. Related work

Video gaming has a history spanning decades. Recently, researchers have turned their attentions to online gaming, looking at its evolution [10]. Furthermore, online social gaming has recently emerged as a hot topic, as it integrates the fields of gaming with that of social networking [9], [6], [13]. This work is very different to our own, though, as we focus on the nature of *streaming* games, rather than playing them.

It could be argued that Twitch is more closely related to general video streaming platforms, particularly user generated content (UGC) repositories. Many studies have been devoted to other UGC platforms [5], [7], [19]. Our work is orthogonal to these, as we show the live broadcasting nature of Twitch makes it fundamentally different to these platforms.

A small set of researchers have started look into Twitch from different perspectives. Most notably, [12] explored the early stages of Twitch (in 2012), finding channel viewing figures were highly predictable by looking at its early viewing figures. We have explored the actual content (*i.e.,* games and tournaments) being broadcast by channels. This has allowed us to shed light on the underlying driving factors of popularity on Twitch. Other researchers have recently explored the delivery infrastructure of Twitch; for instance, measuring the traffic generated by certain channels [8], [17], [18], mapping the video delivery infrastructure [21] and building models of user chat interactions [15].

VIII. Conclusion

In this paper we have explored the most popular game streaming platform in the world, Twitch. We have uncovered in the paper that the popularity of game streams possesses unique characteristics distinct from traditional channels, broadcasters and games. We have uncovered a complex ecosystem, with multiple types of content competing for the attention of users. Newly released games gather a relatively small part of the views, particularly when compared to the ephemeral nature of content in other UGC platforms.

A significant part of Twitch activity is centered on live gaming events, some of which dominate the views when they are taking place. From this angle, Twitch resembles traditional TV sports broadcasting. These events generate predictable flash crowds (as they follow a schedule), gathering millions of concurrent viewers.

References

[1] Giantbomb.com. www.giantbomb.com/api.

[2] Thegamesdb.net. thegamesdb.net.

[3] Twitch ranked 4th in peak internet traffic. http://blog.twitch.tv/2014/02/twitch-community-4th-in-peak-us-internet-traffic/.

[4] H. Abrahamsson and M. Nordmark. Program popularity and viewer behaviour in a large tv-on-demand system. pages 199–210, New York, NY, USA, 2012. ACM.

[5] P. Ameigeiras. Analysis and modelling of YouTube traffic. *Transactions on Emerging Telecommunications Technologies*, (2):360–377, 2012.

[6] J. Blackburn and H. Kwak. STFU NOOB!: predicting crowdsourced decisions on toxic behavior in online games. *Proc. of WWW*, pages 877–887, 2014.

[7] M. Cha, H. Kwak, P. Rodriguez, Y.-y. Ahn, and S. Moon. I tube, you tube, everybody tubes: analyzing the world's largest user generated content video system. *Proc. of ACM IMC*, 2007.

[8] F. Chen, C. Zhang, F. Wang, and J. Liu. Crowdsourced live streaming over the cloud. *arXiv preprint arXiv:1502.06314*, 2015.

[9] T. Chung, J. Han, D. Choi, and T. Kwon. Unveiling group characteristics in online social games: a socio-economic analysis. *Proc. of WWW*, pages 889–900, 2014.

[10] S. Gallagher and S. Park. Innovation and competition in standard-based industries: a historical analysis of the US home video game market. *IEEE Engineering Management*, 49(1):67–82, 2002.

[11] A. Hern. Twitch plays pokémon: live gamings latest big hit. *The Guardian*, 2014.

[12] M. Kaytoue, A. Silva, L. Cerf, W. Meira Jr, and C. Raïssi. Watch me playing, i am a professional: a first study on video game live streaming. In *Proc. of WWW Companion*, pages 1181–1188, 2012.

[13] R. Kowert and J. A. Oldmeadow. 965-976. *Computers in Human Behavior*, June 2014.

[14] D. K. Krishnappa, M. Zink, C. Griwodz, and P. Halvorsen. Cache-centric video recommendation: An approach to improve the efficiency of youtube caches. *ACM Trans. Multimedia Comput. Commun. Appl.*, 11(4):48:1–48:20, June 2015.

[15] G. Nascimento, M. Ribeiro, L. Cerf, N. Cesario, M. Kaytoue, C. Raissi, T. Vasconcelos, and W. Meira. Modeling and analyzing the video game live-streaming community. In *Proc. of IEEE LA-WEB*, pages 1–9. IEEE, 2014.

[16] G. Nencioni, N. Sastry, J. Chandaria, and J. Crowcroft. Understanding and decreasing the network footprint of over-the-top on-demand delivery of tv content. In *Proc. of WWW*, pages 965–976, 2013.

[17] K. Pires and G. Simon. Dash in twitch: Adaptive bitrate streaming in live game streaming platforms. In *Proc. of VideoNext*, pages 13–18, 2014.

[18] K. Pires and G. Simon. Youtube live and twitch: a tour of user-generated live streaming systems. In *Proc. of ACM MMSys*, pages 225–230, 2015.

[19] G. Tyson, Y. Elkhatib, N. Sastry, and S. Uhlig. Demystifying porn 2.0: a look into a major adult video streaming website. In *Proc. of ACM IMC*, pages 417–426, 2013.

[20] H. Yin, X. Liu, F. Qiu, N. Xia, C. Lin, H. Zhang, V. Sekar, and G. Min. Inside the bird's nest: measurements of large-scale live vod from the 2008 olympics. In *Proc. of ACM IMC*, pages 442–455, 2009.

[21] C. Zhang and J. Liu. 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*, pages 55–60. ACM, 2015.