THE OHIO STATE UNIVERSITY

YouTube Live and Twitch: A Tour of User-Generated Live Streaming System

Mengxue Zhang
Dingkang Wang
Xianxing Zhang

User Generated Content (UGC)

Any form of content created by users of a system or service and made available publicly on that system

Type:

Websites: Entertainment media publications include **Reddit**, **9Gag**, **4chan**, **Upworthy**
Video Games: mods, fan patches, etc.
Some games involed user-generated world: mincraft
Retailers: bargain hunting websites: eBay

Education: Wikipedia.

Live-video streaming: Youtube, Twitch, Steam, etc

Etc

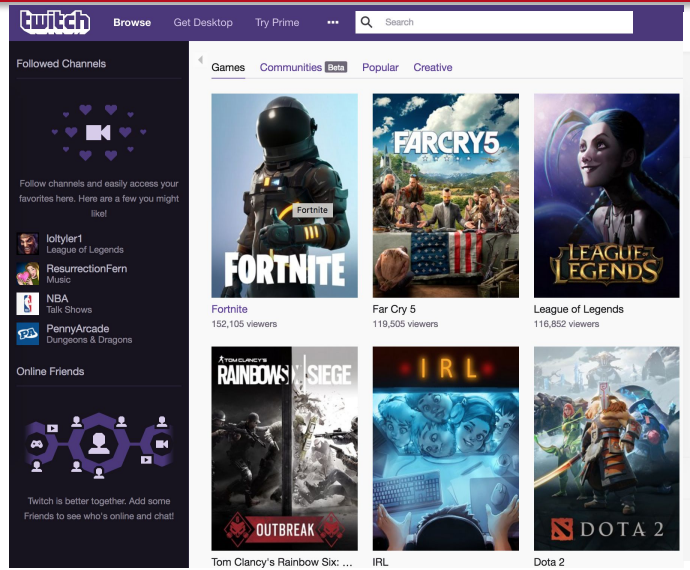
What is it about?

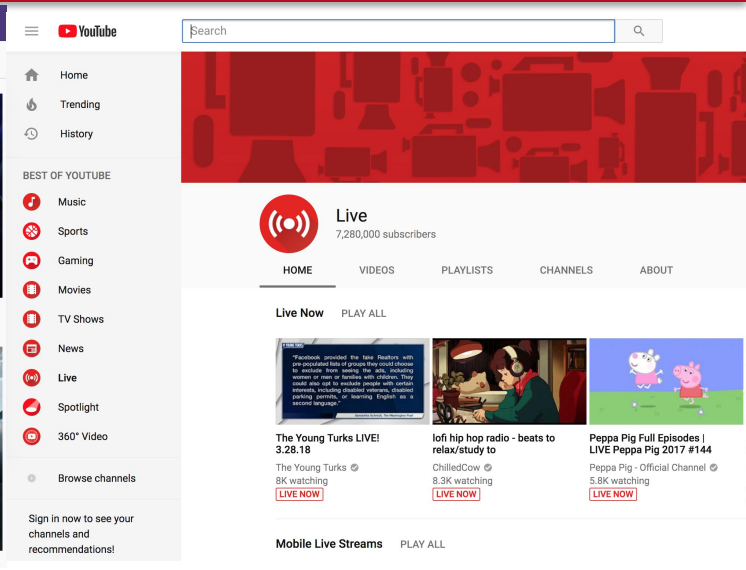
Karine Pires
Gwendal Simon

trace

Two User-Genereated Streaming System Services

2014.01-2014.04





Data

Observation:

1. Both generate traffic with frequent peaks at more than 1 Tbps.
2. Popularity of channels is more heterogeneous than other platforms

DATASET

Original Data

Fetch the following information at every 5 minutes

	Twitch	YouTube
channel id	yes	yes
session id	yes	yes
nb. of viewers	yes	yes
video bitrate	yes	no
video resolution	yes	no
uploader country	yes	no

session: time when the uploader is online, broadcasting live video
inter-session: uploader is offline

Data Visualization

show the data in figure or table for analyzing

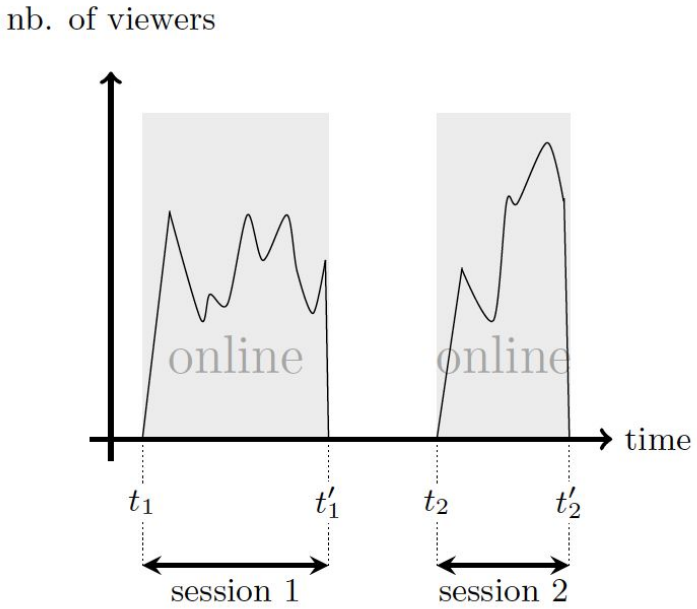


Figure 1: A life in a channel

Pre-Processing

Filtering Out the “Actual Uploaders” vs “Testers”

1.Uploaders who launched a channel for only one session with less ten minutes

2.Channels from uploaders have no viewers at all

	Twitch	YouTube
total nb. of channels	1,536,492	236,957
total nb. of sessions	6,242,609	737,233
10 min. channels	25%	27%
no viewers	11%	40%
filtered nb. of channels	1,068,138 (69%)	120,097 (51%)
filtered nb. of sessions	5,221,208 (83%)	527,677 (71%)

Table 2: Filtering testers from the traces

Testers may do harm to delivery infrastructure on the uplink.

UGC live system should be able to detect and prevent testers from harming the serverce.

Size of these systems

- Overall bandwidth
- Number of concurrent sessions
- Number of different channels

Overall bandwidth

- Both peaks more than 1Tbps
- Twitch has higher peaks (1.6 Tbps)

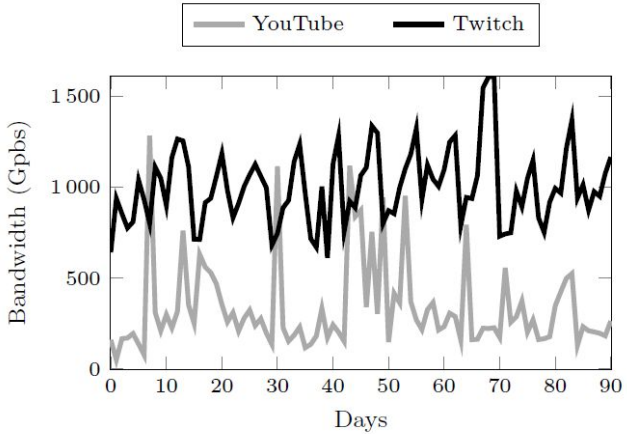


Figure 2: Bandwidth usage for live video delivery

Number of Concurrent Sessions

- Youtube: 300-700
- Twitch: more than 6000

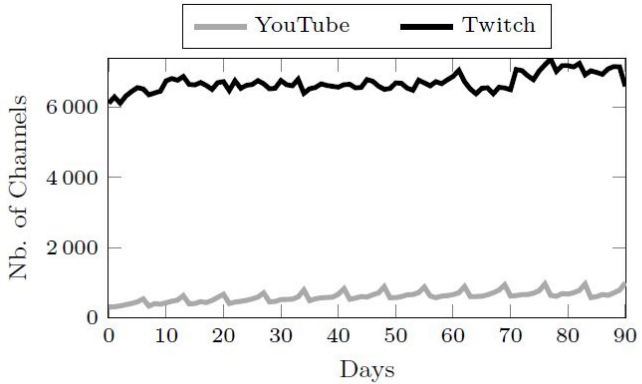


Figure 3: Number of simultaneous online channels

Number of Different Channels

- Twitch has far more living channels
- Seems the numbers are still increasing

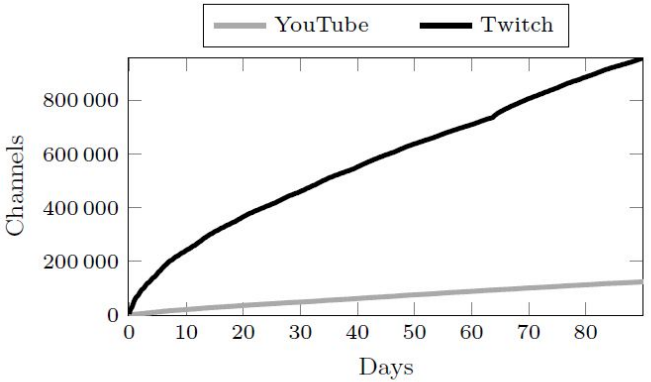


Figure 4: Cumulative number of unique channels

Diurnal & weekly patterns

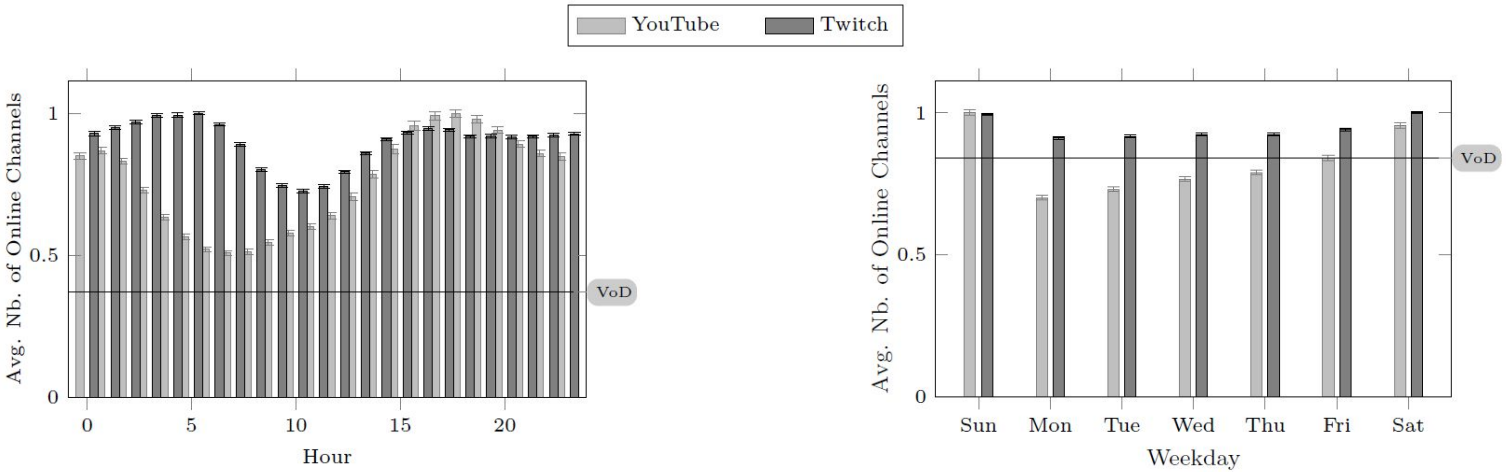


Figure 5: Average number and confidence interval of simultaneous online channels by hour and weekday

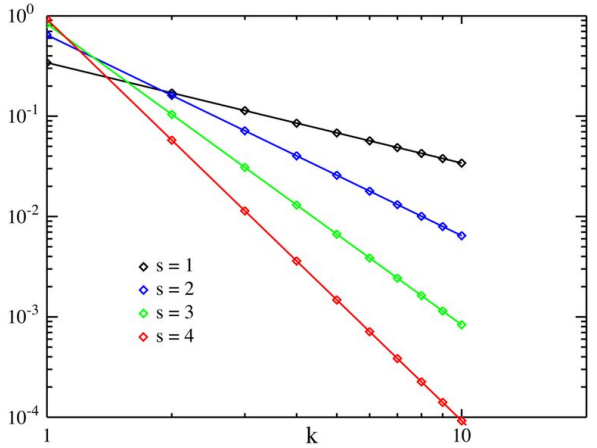
Twitch is less sensitive

- Service starts earlier, uploaders from the whole world
- Twitch is more related to e-sport, which is popular in asia, so it is more balanced during the day & week

Zipf's law

$$f(k; s, N) = \frac{1/k^s}{\sum_{n=1}^{N} (1/n^s)}$$

Larger s (***a***), fewer channels have high popularity.



Probability mass function on log-log scale

Both systems follow zipf's law

- Both youtube and twitch follows zipf's law, NRMSD < 5%.
- Youtube has larger ***α*** values, and not that stable.

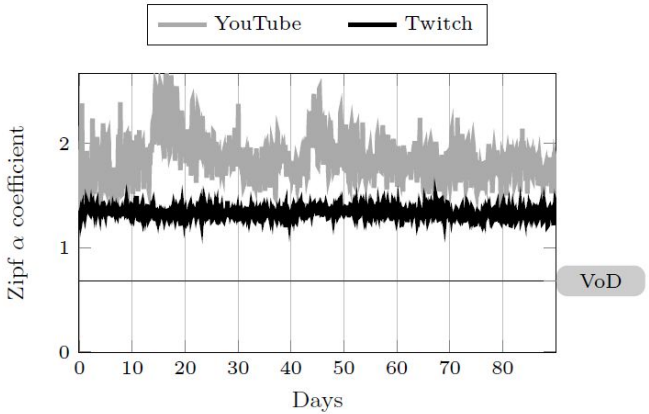


Figure 6: Zipf α coefficient evolution over time

Dataset Usage

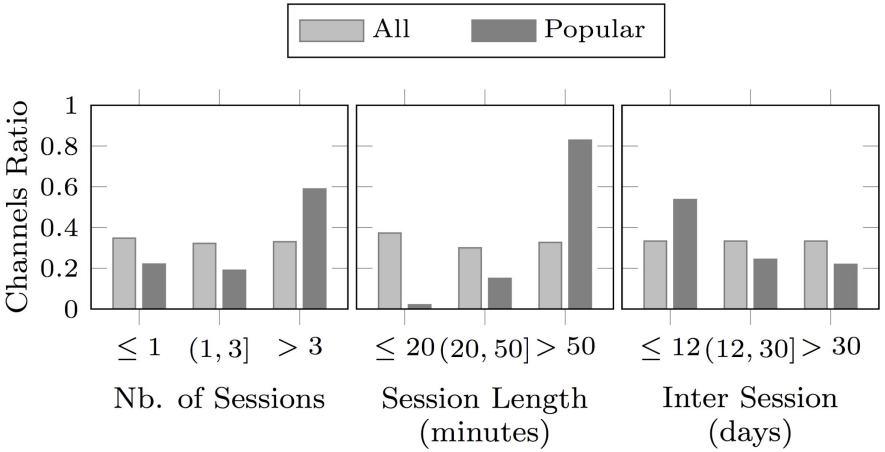
1. **Forescast of popular Channels**

Goal: identify the most popular channels as early as possible.

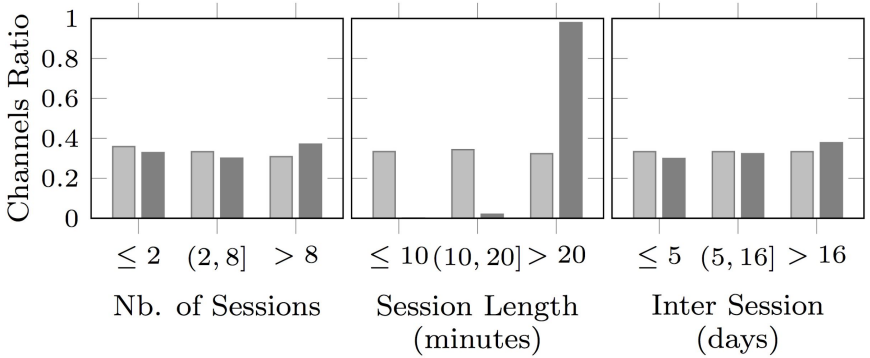
Selected characteristics of Channels: the length of their sessions, the interval between sessions and the # of sessions.

preliminary Method: compare the distribution with existing top 1% popular channels

Dataset Usage: Forescast of popular **Channels**(some results)



(a) YouTube



(b) Twitch

Dataset Usage

2. study of delivery methods on live streaming sercies

Goal: to reduce the delivery bandwidth cost and increase QoE (quality of experience) of viewers

Method:

Treat as management probelm

Design strategies for deciding which online chanels should be delivered by adaptive

Resulting paper:

K. Pires and G. Simon. ***Dash in twitch: Adaptive bitrate streaming in live game streaming platforms***. In VideoNext CoNEXT Workshop. ACM, 2014.

Dataset Usage (other work)

3. evaluation of scenarios that are based on different service providers competitors.
In other word check the content delivery network(CDN)'s fairness.

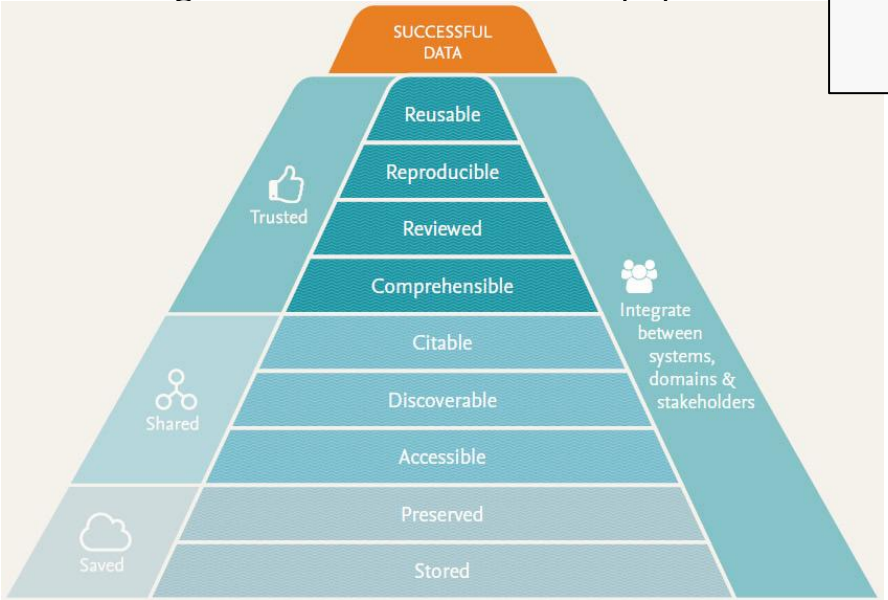
---

4. ??? What's more in data mining

	Twitch	YouTube
channel id	yes	yes
session id	yes	yes
nb. of viewers	yes	yes
video bitrate	yes	no
video resolution	yes	no
uploader country	yes	no

Discussion and suggestion

What is a good dataset and a dataset paper.



Inspiration

Discussion

1. Better have more the attributes for channels :
   For example:
   - The key word or content for the video
   - Viewers' information (with privacy protection)
   - comments under stream

2. It's unfair to compare Youtube live stream and Twich stream. (different domain)
3.
4.
5.
6. More

Thank you!

Project Update

1. Finish sampling by using Steam API
   Sampling strategies using currently

   a. Start with a list of seed steamID selected from different country (famous steam user). Get their friends list and expanding until the networks size reach the maximize value(10,000 now)

   b. Start with a list of seed steamID selected from US (famous steam user, but playing different types of games). Get their friends list and expanding until the networks size reach the maximize value(10,000 now)

Project Update

2. finish building attribute network

ID	Recently_played_games	recently palyed time	User Countrycode	User StateCode
123234843	"total_count": 13, "games": [{ "appid": 250820, "name": "SteamVR", "playtime_2weeks": 1488, "playtime_forever": 1641, "img_icon_url": "7f286d5ee22905b51663e0cea505e1c2f7500f36", "img_logo_url": "cb78fa6183c3c876ad90ed6d377cc2e98f8dfb59" }, { "appid": 646570, "name": "Slay the Spire", "playtime_2weeks": 652, "playtime_forever": 1141, "img_icon_url": "33ea124ea8c03a9ce7012d34c3b348a351612fca", "img_logo_url": "6ecbf741b482a476d7d809ba3e3fea028b805b67" }, { "appid": 377160, "name": "Fallout 4", "playtime_2weeks": 158, "playtime_forever": 163, "img_icon_url": "779c4356ebe32af2af7c9f0bbba595dfe872cd7f", "img_logo_url": "8977a8e98acbbdd3c0ff905afb7e0a6e2eb555ea" },	9802 sec	US	WA

Project Update

3. Next week

a. run comunity detection algorithm

b. analize attribute distribution

c. Add weight to the network