Just One More: Modeling Binge Watching Behavior
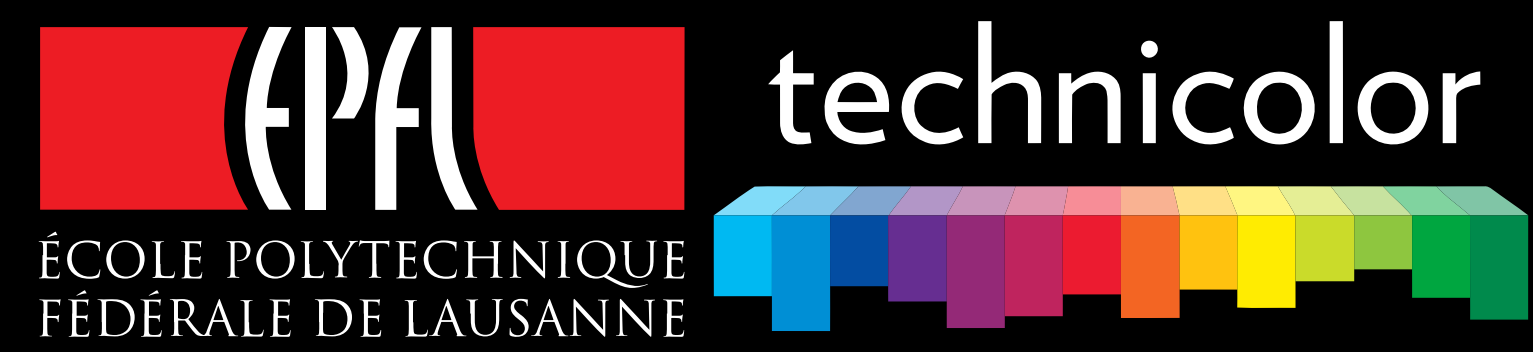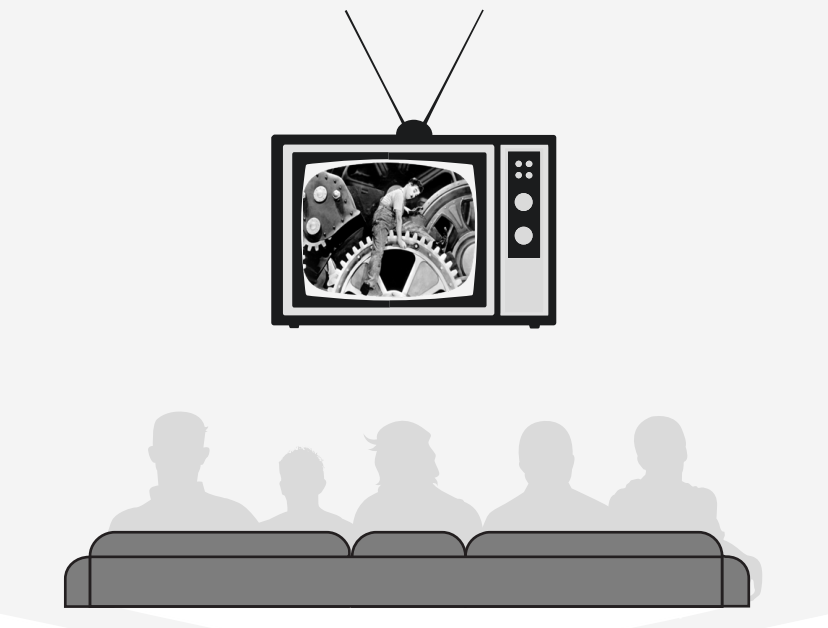
William Trouleau[†], Azin Ashkan[‡*], Weicong Ding[‡] and Brian Eriksson[‡]

† LCA 4, School of Computer and Communication Sciences, EPFL
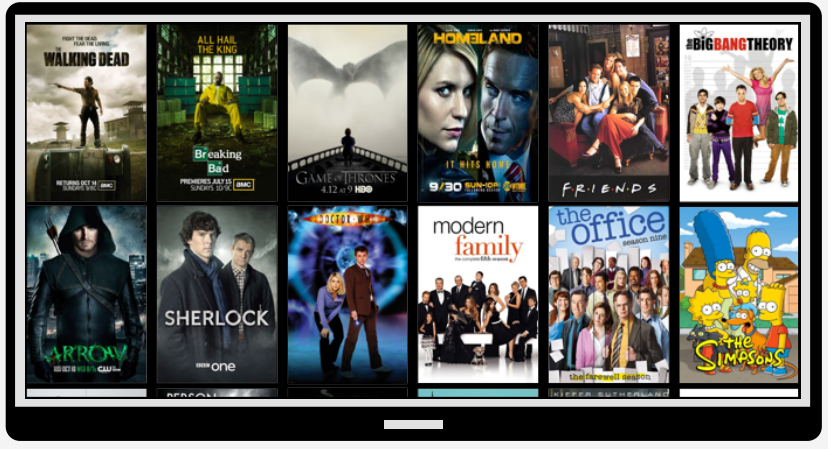‡ AI Lab, Technicolor Research

KDD2016

EPFL — ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE — technicolor

---

I. History of TV show consumption schemes

Past TV show consumption scheme:
TV broadcast

· No choice of device... Only TV
· No choice of content
· No choice of time

New TV show consumption scheme:
VOD services

· Any device you want (TV, tablet, cellphone...)
· Any content you want
· Any time you want

VOD enabled **Binge Watching!**

II. Popular-culture definition of binge watching

Definition

Popular press and market research define binge watching as

Watching **multiple episodes** of a television program
in a single sitting (or **session**)

Limitations

· External factors ignored
· Based on survey data... And reports do not agree
on the episode threshold:
· TiVo: **3** episodes
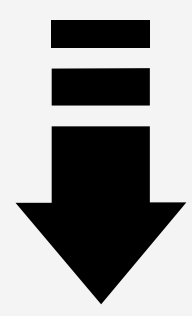· Netflix: **2** episodes

Questions

· Using a data-driven approach, is there a clear
episode threshold?
· Are there other factors to consider?

III. Objective

Step 1

Model viewer behavior with a
data-driven approach using
real-world VOD data.

Step 2

Characterize binge watching
behavior through this model.
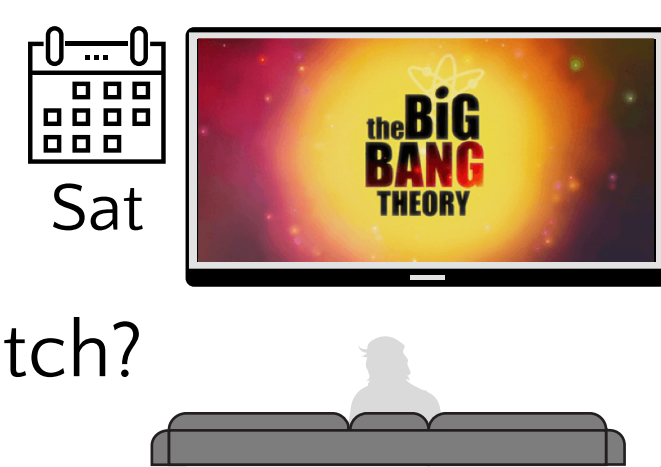

IV. Modeling TV show viewing behavior

Goal

Model the **number of episodes** viewed in a **session**

Example:
It is **Saturday** at **4pm**, you start watching
"The Big Bang Theory" on **TV**.

How many episodes are you likely to watch?

Model

Censored Poisson Regression with **Latent Factors**

Latent factors

Observation: There are different **types** of viewing behaviors.

Method: Model number of episodes v_i viewed in
session i as a random drawing from a
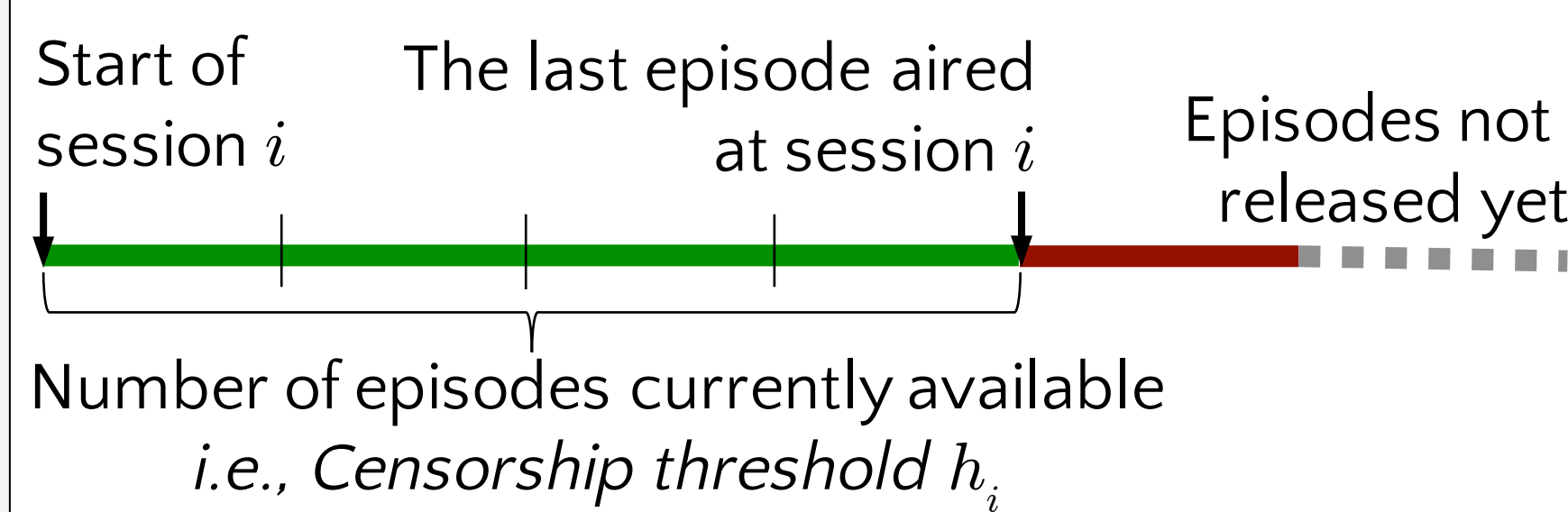**mixture** of K Poisson distributions.

Insights: Account for the **heterogeneous** behavior
between sessions.

Censorship

Observation: "censored" session, *i.e.*, the user
have consumed all available content.

Method: · Model each mixture component as
a censored Poisson distribution.

· A session is "censored" ($c_i=1$) if the
latest episode is consumed in the
session.

Insights: Account for the case when the user
cannot possibly consume more but
may desire to.

Start of session i — The last episode aired at session i — Episodes not released yet

Number of episodes currently available
i.e., Censorship threshold h_i

Poisson Regression

Observation: Context of a session influences the
rate of consumption.

Method: Model the consumption rate of each
session as function (log-linear) of the
covariates of each session.

Covariates: · Device used   · Day of Week
· Hour of Day   · TV Series Title

Parameter Estimation and Inference

Inference: EM algorithm

Model Selection: Set K=3 using predictive log-likelihood
via 5-fold cross-validation.

Predictive Tasks: (1) Predicting the number of episodes in a
session: 9% improvement in MAE compared
to other **regression** methods.

(2) Predicting if the user will watch the next
episode: 7% improvement in AUC
compared to other **classification** methods.

V. Characterizing binge watching

Differences between the mixture components

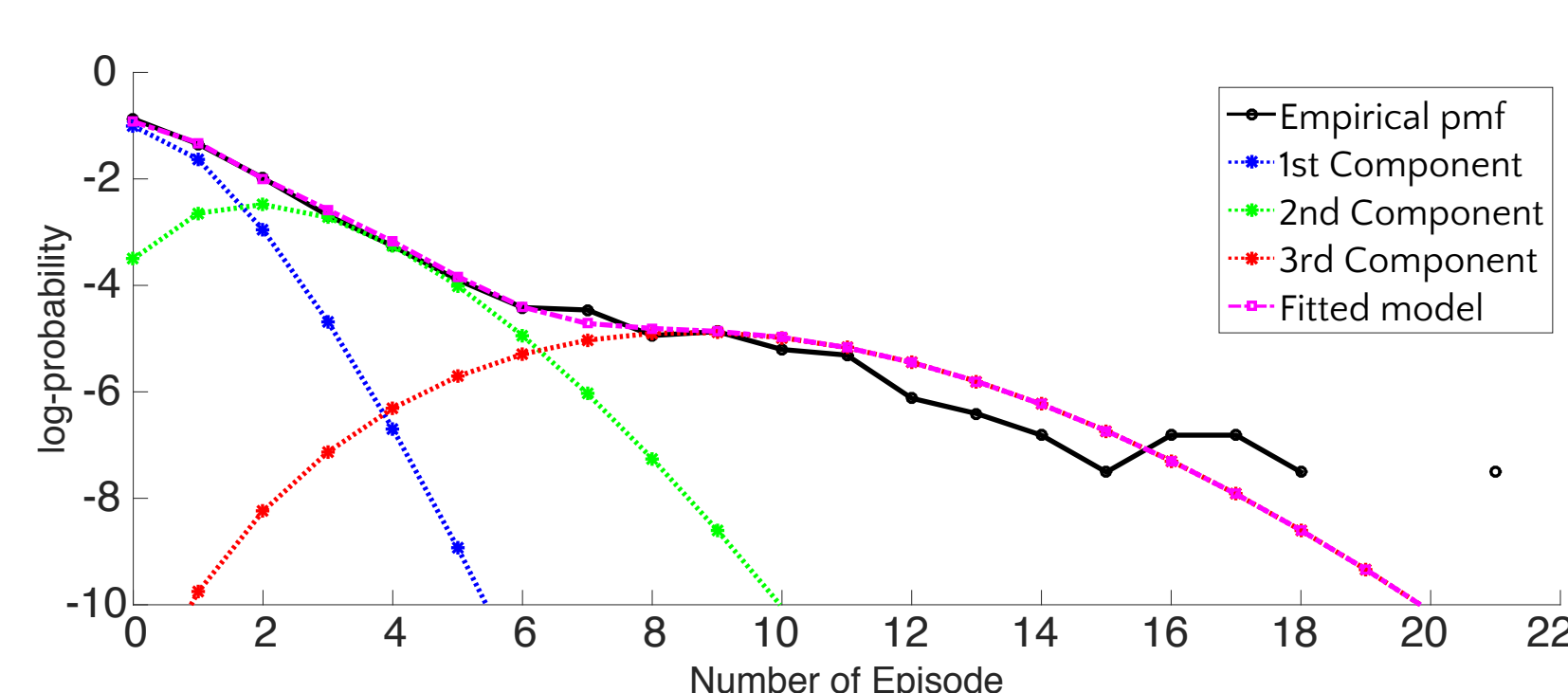From our viewership model, we can
observe the differences between
the mixture components.

Given K=3, we find three different
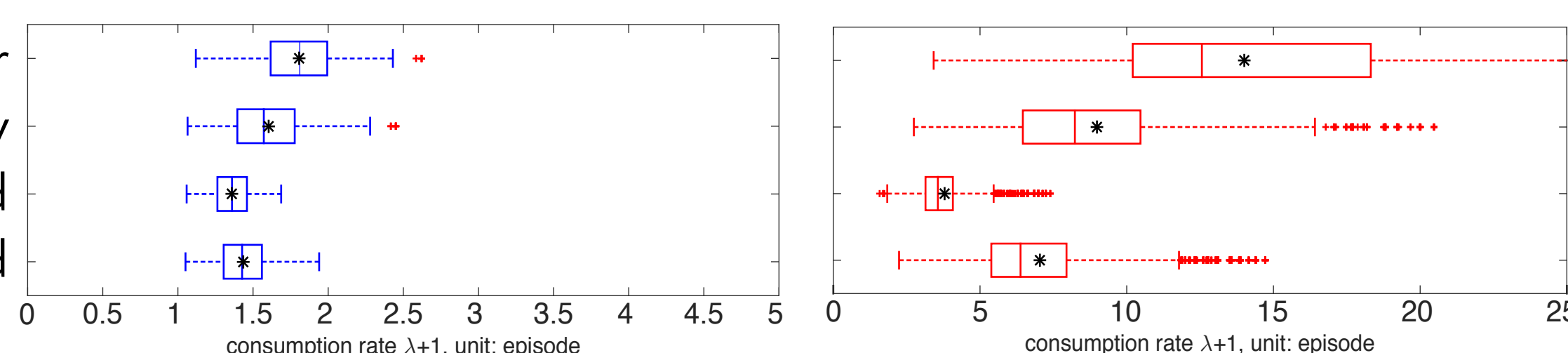types of viewer behavior:
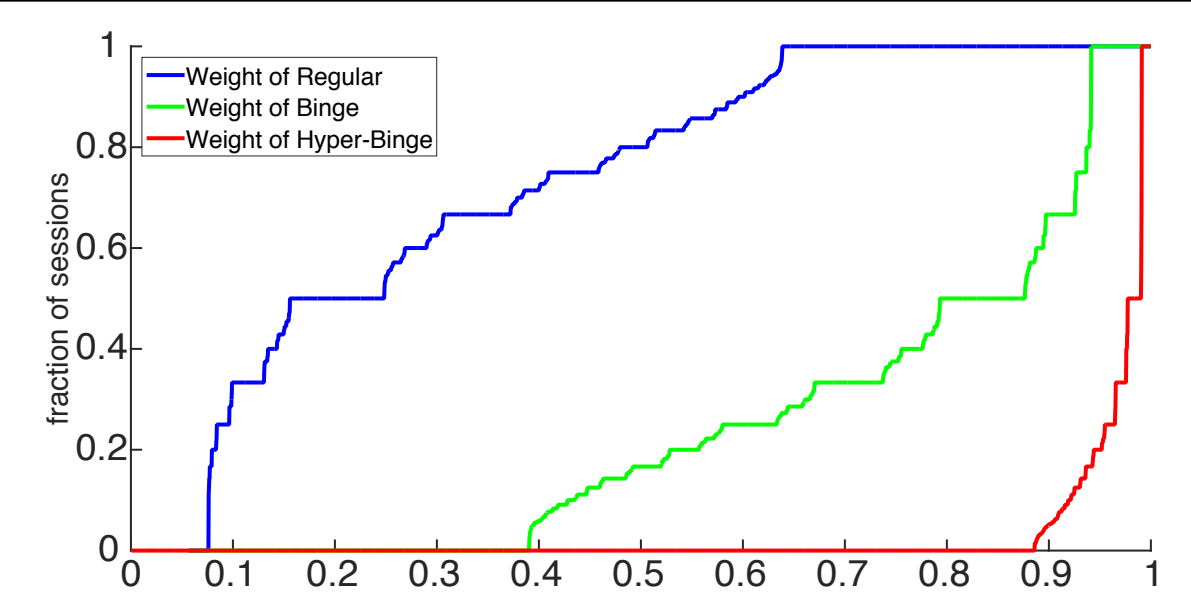
Regular

Binge

Hyper-binge

Model fit to "The Big Bang Theory" sessions

Consumption changes greatly depending on content

Regular   **Hyper-binge**

A majority of users binge

· Over **64%** binge at leat once
· **20%** binge in half their sessions
· **8%** binge all the time
· **11%** Hyper-binge at least once

Not all binge sessions are trying to complete a season

· Dramas such as "The Walking Dead" and
"Homelad" are watched **in-order**.

· Comedies such as "The Big Bang Theory"
and "Modern Family" are **non-sequential** in
over 40% of binge sessions.

Title	Regular	Binge	Hyper-binge
All Series	97%	84%	76%
Walking Dead	98%	87%	85%
Homeland	98%	94%	96%
HIMYM	97%	88%	77%
Big Bang Theory	87%	56%	52%
Modern Family	83%	59%	78%
NCIS	98%	71%	67%

Percentage of sessions wiewing episodes in order

Binge watching is irregular

		Next session	
	Regular	Binge	Hyper-binge
Regular	0.82	0.16	0.02
Binge	0.66	0.30	0.04
Hyper-binge	0.59	0.32	0.09

User behavior transition matrix

A majority of binge watchers return
to regular consumption in their next
session.

VI. Take away message

Accurately modeling user viewing behavior
requires a model that considers both
inhomogeneity between sessions and
content censorship.

Our model reveals two distinct types of
binge watching behaviors:
· **binge** and
· **hyper-binge** watching.

· Binge watching is a **prominent** behavior adopted by
more than 64% of users.
· Binge watching is an **irregular** behavior as more than
65% of binge watchers will not binge in their next session.