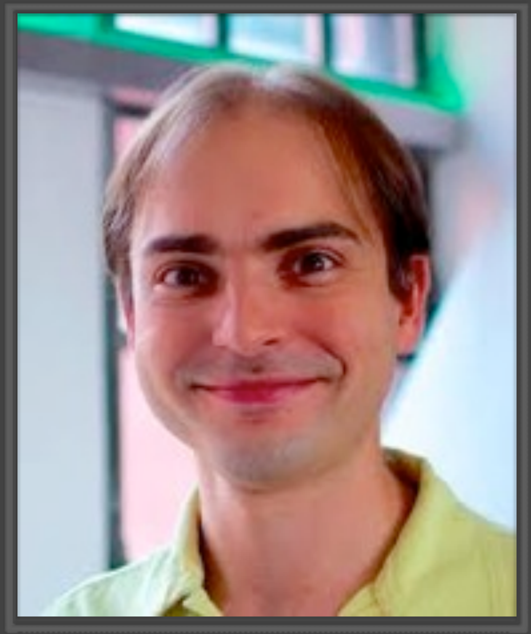
The Dynamics of Repeat Consumption

Ashton Anderson
Stanford University

Ravi Kumar, Andrew Tomkins, Sergei Vassilvitskii
Google



repeat consumption

a lot of consumption is *repeat* consumption

what factors determine what we reconsume?

given a set of previously-consumed candidates, predict which item a user will choose to reconsume

consumption data

BrightKite: location checkins

G+: public location checkins

MapClicks: clicks on Google Maps businesses

MapClicks-Food: clicks on Google Maps restaurants

consumption data

WikiClicks: all clicks on English Wikipedia pages by Google users

YouTube: last 10K video watches of users

YouTube-Music: YouTube restricted to music videos

baselines

Yes: radio playlists from hundreds of US radio stations*
(to compare against non-individual consumption data)

Shakespeare: full text of Shakespeare's works, with each letter considered an item
(to compare against data with repetitions)

* available at http://www.cs.cornell.edu/~shuochen/

the dynamics of repeat consumption

1. empirical analysis
2. models
3. experiments

the dynamics of repeat consumption

1. empirical analysis
2. models
3. experiments

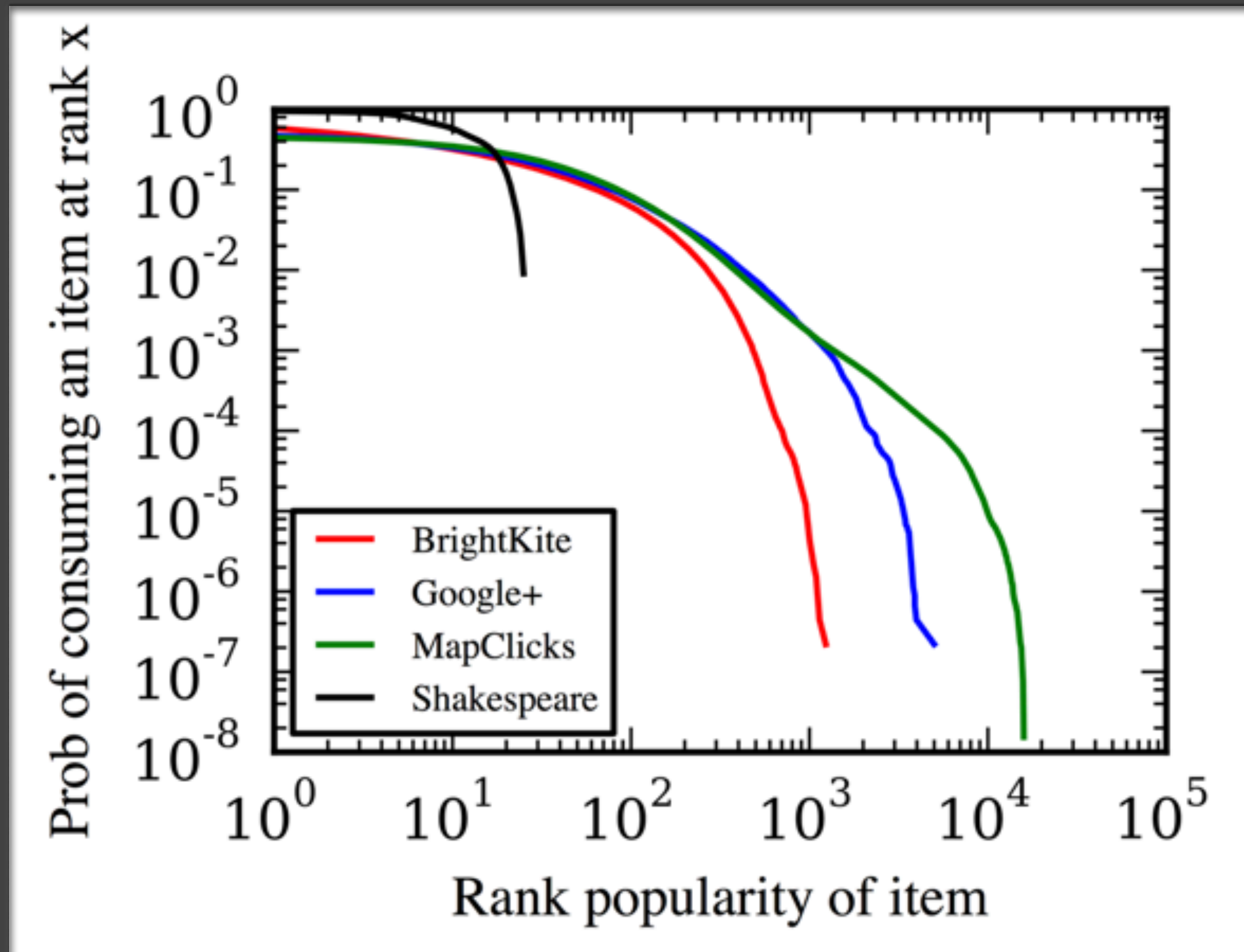
empirical analysis

what are the empirical traits of reconsumed items?

popularity

individual popularity: are users generally exploiting or exploring?

popularity



more frequently consumed items are more likely to be reconsumed

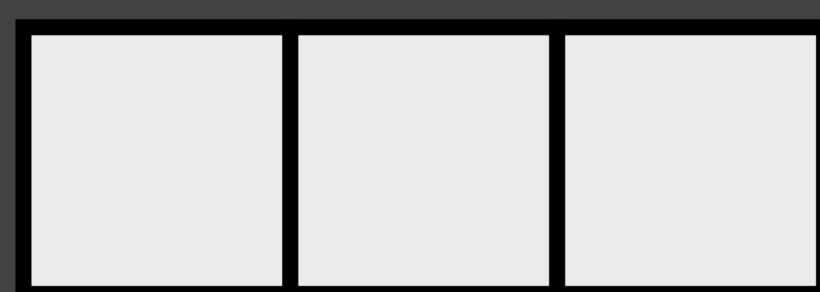
recency

how does the recency of consumption affect the likelihood of reconsumption?

to answer this question, we use a cache-based analysis technique

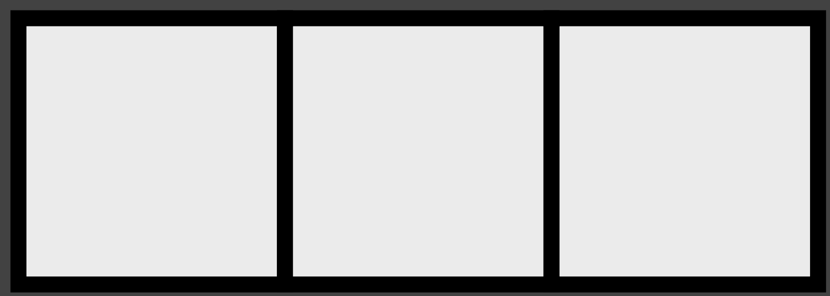
recency

consider a cache of size k=3:



recency

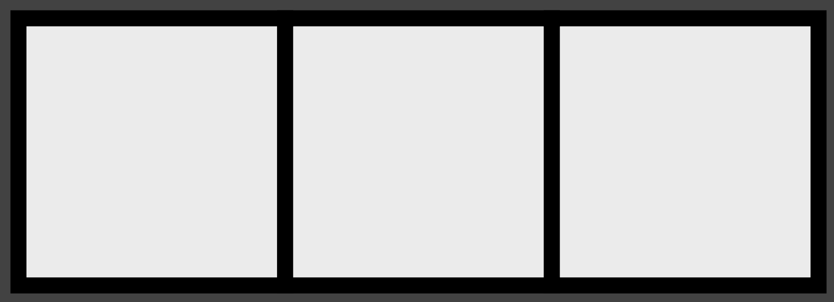
process a consumption history using optimal offline caching (replace item that occurs furthest in the future)



recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c



recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a		

Hits: 0 Misses: 1

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	b	

Hits: 0 Misses: 2

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	b	

Hits: 1 Misses: 2

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	b	c

Hits: 1 Misses: 3

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	b	d

Hits: 1 Misses: 4

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

e	b	d

Hits: 1 Misses: 5

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

e	b	d

Hits: 2 Misses: 5

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

e	b	d

Hits: 3 Misses: 5

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	b	d

Hits: 3 Misses: 6

recency

consumption history:

a b b c d e b d a c d c

a c d

Hits: 3 Misses: 7

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

a	c	d

Hits: 4 Misses: 7

recency

consumption history:

a	b	b	c	d	e	b	d	a	c	d	c

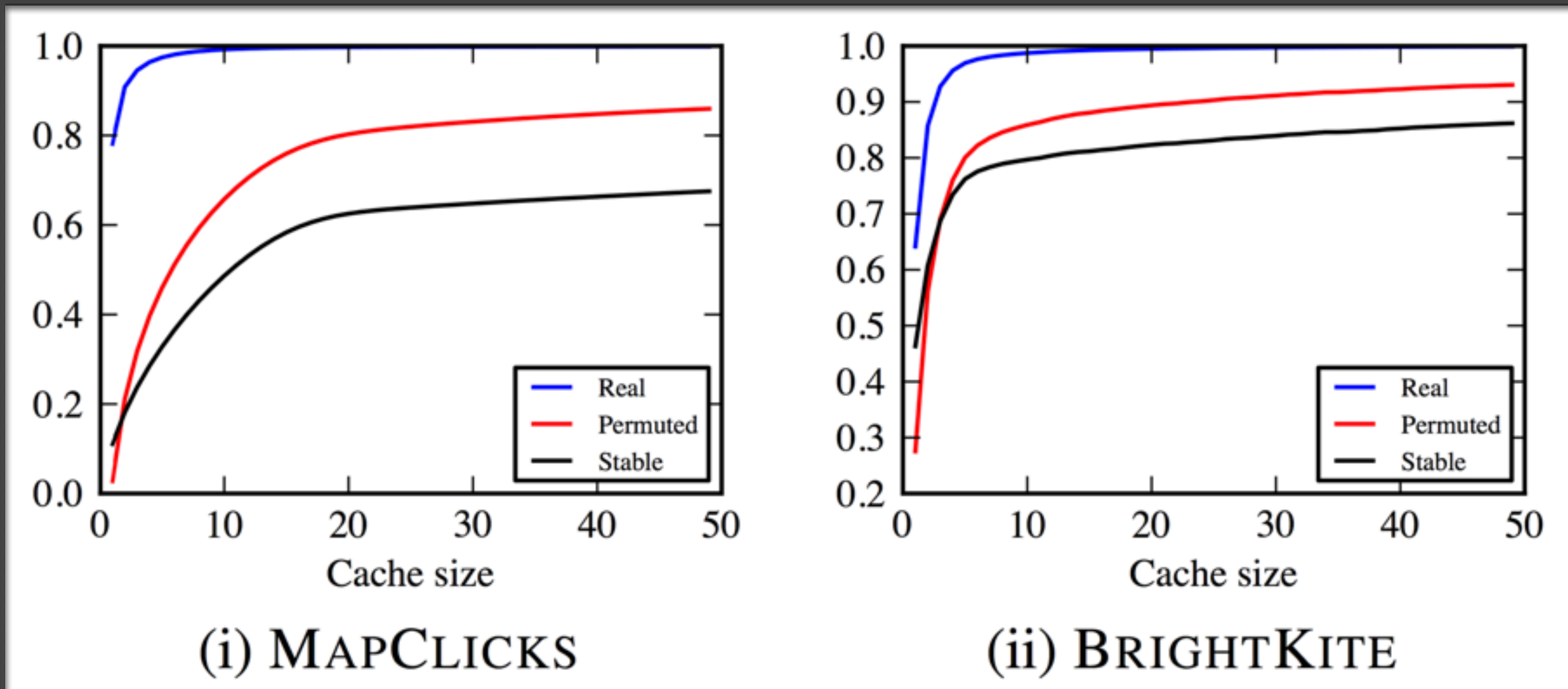
a	c	d

Hits: 5 Misses: 7

recency

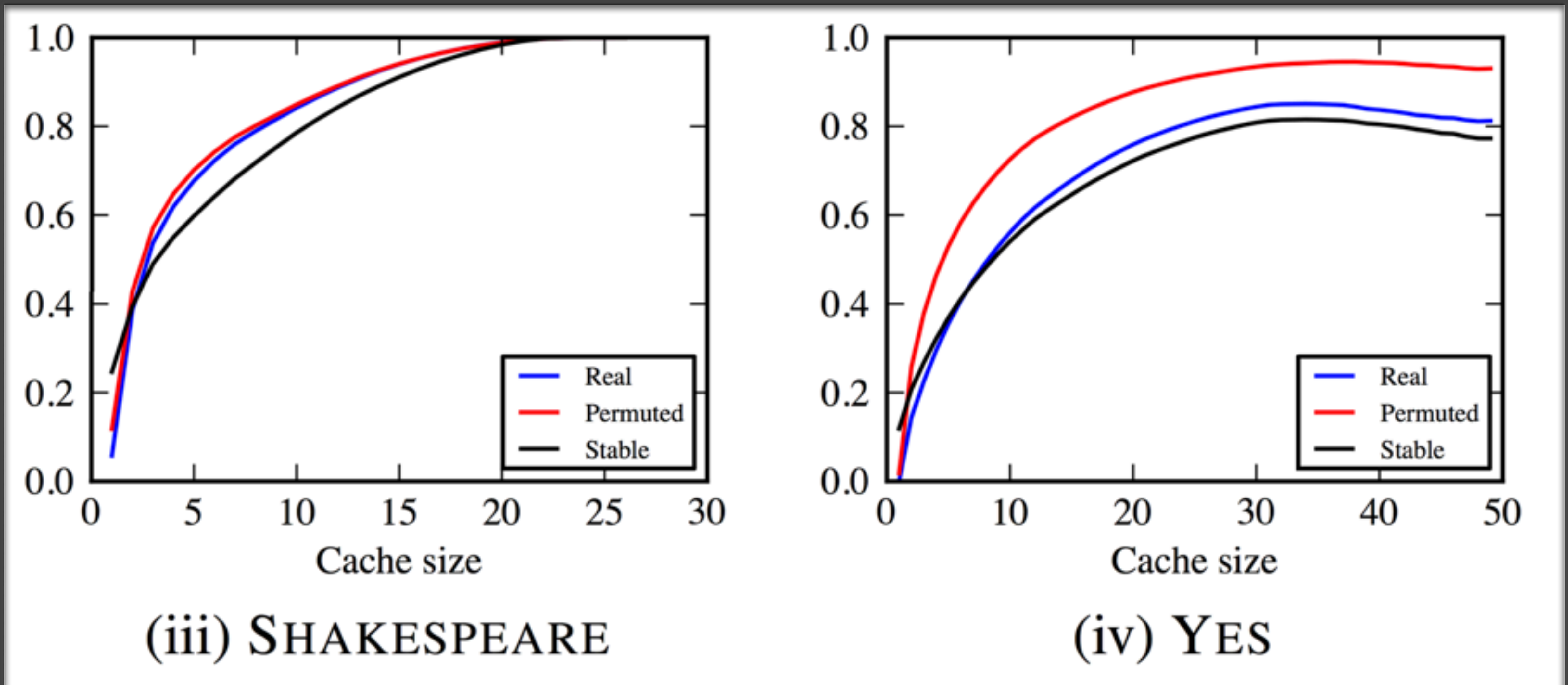
the hit ratio is an indication of the degree to which recency is displayed in a consumption history

recency



Real consumption sequences display a significant amount of recency

recency



Baseline datasets *don't* display recency

(Yes even shows anti-recency)

empirical analysis

user-level item popularity generally positive predictor

recency is the strongest effect

the dynamics of repeat consumption

1. empirical analysis
2. models
3. experiments

models

goal: develop a simple mathematical framework powerful enough to explain patterns of reconsumption we observe in real data

models

first, fix vocabulary E of items

a consumption history for user u is $X_u = x_1, \ldots$
where each $x_i \in E$

at each step, user picks next item to consume
using some function of consumption history

quality model

natural hypothesis: item quality dictates consumption behavior

associate score $s(e)$ for each $e \in E$, and at each step next item is chosen proportionally to its score:

$$P(x_i = e) = s(e) / \sum_{e' \in E} s(e')$$

recency model

since recency is the strongest empirical effect, we formulate a *copying* model based on it

at every step *i*, user copies item at position *i-j* proportional to weight *w(i-j)*

recency model

since recency is the strongest empirical effect, we formulate a *copying* model based on it

at every step *i*, user picks item at position *i-j* proportional to weight *w(i-j)*

consumption history

a	b	b	c	d	e	b	d	a	c	d	c	?

recency model

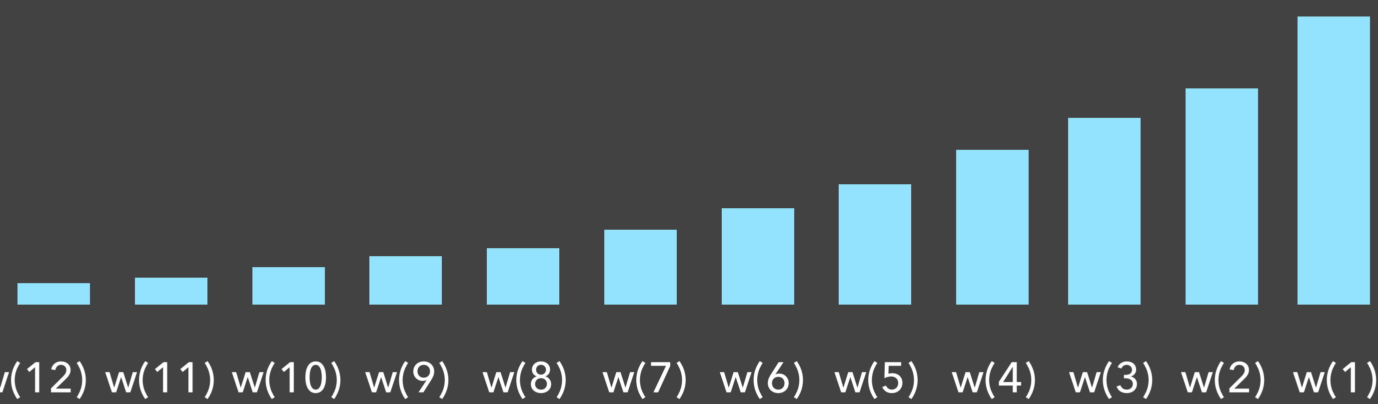
since recency is the strongest empirical effect, we formulate a *copying* model based on it

at every step *i*, user picks item at position *i-j* proportional to weight *w(i-j)*

consumption history

a	b	b	c	d	e	b	d	a	c	d	c	?

weights *w*



recency model

since recency is the strongest empirical effect, we formulate a *copying* model based on it

at every step *i*, user picks item at position *i-j* proportional to weight *w(i-j)*

consumption history

a	b	b	c	d	e	b	d	a	c	d	c	?

e.g.: $P(x_i = d) \sim$ w(8) $+$ w(5) $+$ w(2)

recency model

since recency is the strongest empirical effect, we formulate a *copying* model based on it

at every step *i*, user picks item at position *i-j* proportional to weight *w(i-j)*

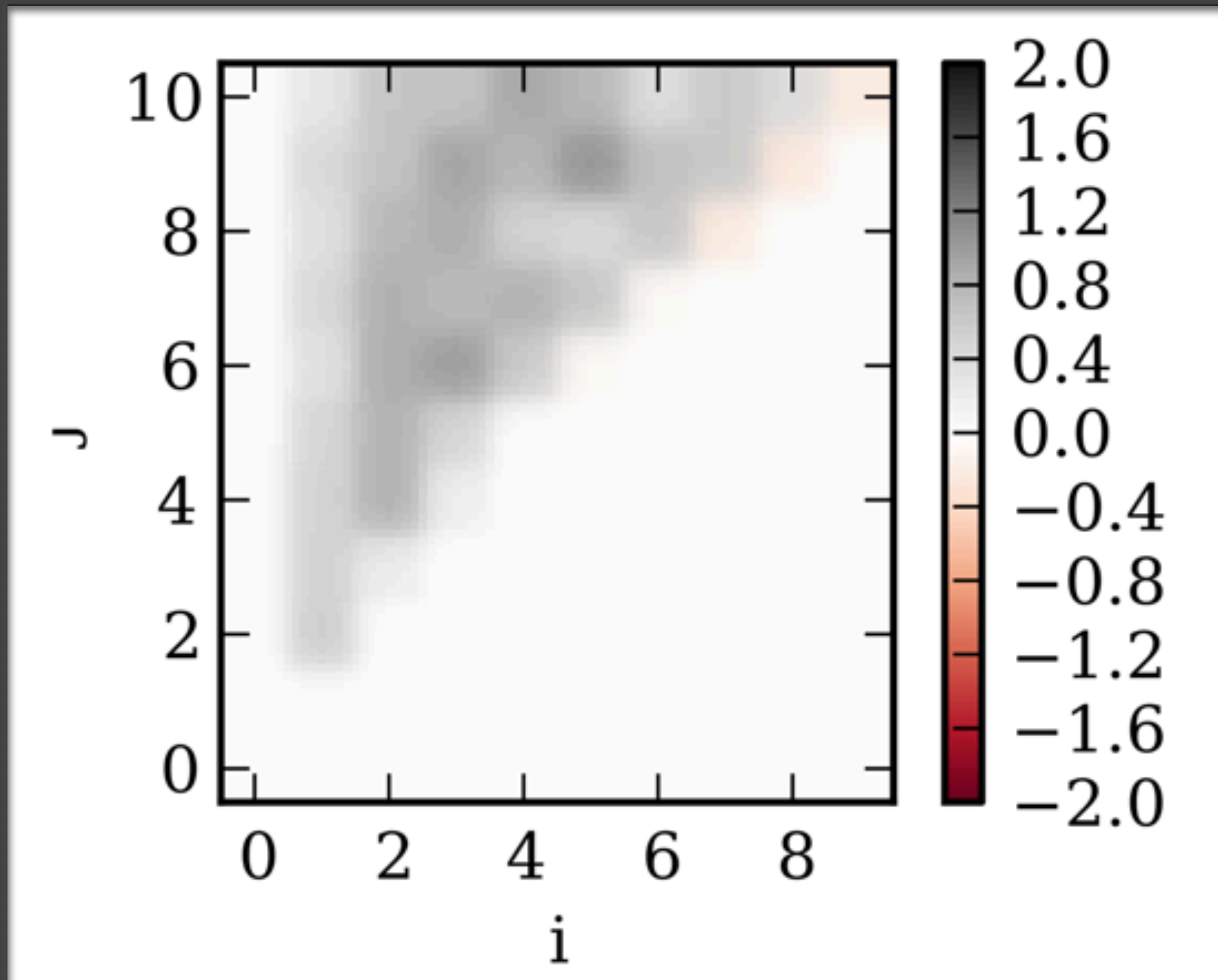
$$P(x_i = e) = \frac{\sum_{j<i} I(x_i = e) w(i-j)}{\sum_{j<i} w(i-j)}$$

recency model

we assume additivity in weights

thought experiment: learn weights, and compare additivity prediction to actual likelihoods from copying

recency model

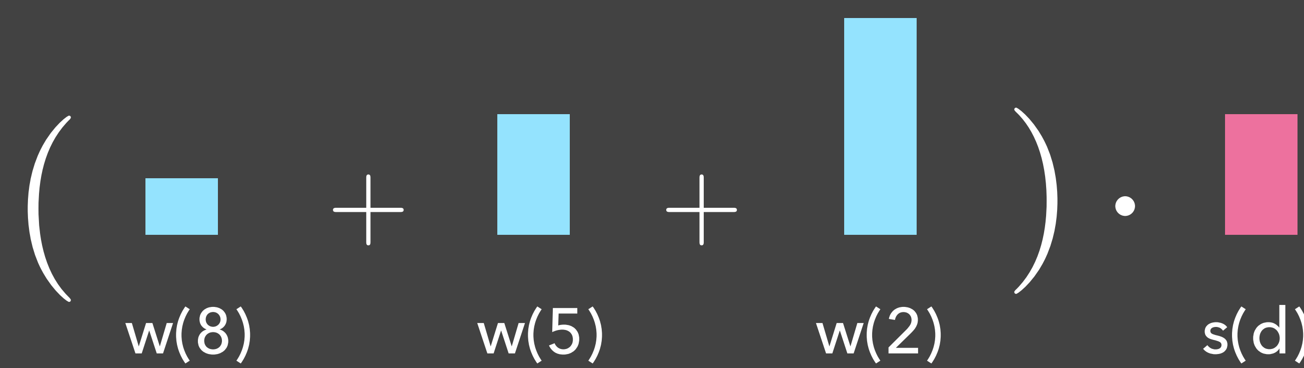


very small deviations from additivity

hybrid model

combination of recency and quality

e.g.: $P(x_i = d) \sim$



$$P(x_i = e) = \frac{\sum_{j<i} I(x_j = e) w(i-j) s(x_j)}{\sum_{j<i} w(i-j) s(x_{i-j})}$$

learning model parameters

quality model: simply the empirical fraction of occurrences

$$s(e) = \frac{1}{k} \sum_{i=1}^{k} I(x_i = e)$$

learning model parameters

recency and hybrid models:
maximize likelihood with stochastic gradient ascent

$$LL = \log \left(\prod_{i \in R} \frac{\sum_{j<i} I(x_i = x_j) w(i-j) s(x_j)}{\sum_{j<i} w(i-j) s(x_j)} \right)$$

learning model parameters

weight update:

$$\frac{\partial LL}{\partial w(\delta)} = \sum_{i \in R} \begin{cases} \frac{s(x_i)}{A_i(x_i = x_j)} - \frac{s(x_i)}{A_i(1)} & \text{if } x_i = x_{i-\delta}, \\ -\frac{s(x_i)}{A_i(1)} & \text{otherwise} \end{cases}$$

score update:

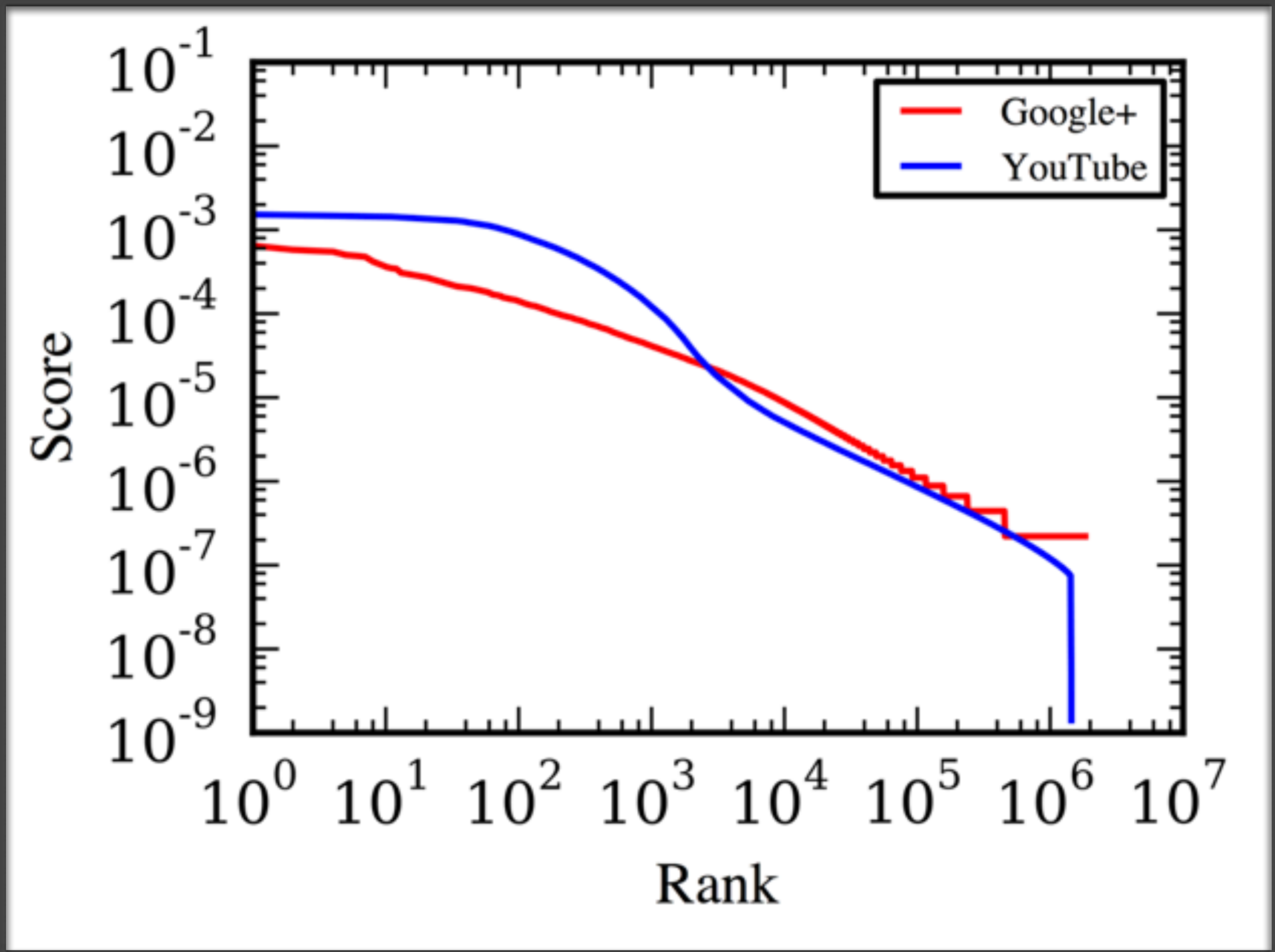
$$\frac{\partial LL}{\partial s(e)} = \sum_{i \in R} \begin{cases} 1 - \frac{A_i(x_j = e)}{A_i(1)} & \text{if } x_i = e, \\ -\frac{A_i(x_j = e)}{A_i(1)} & \text{otherwise.} \end{cases}$$

alternating updates to local maximum (not jointly convex)

the dynamics of repeat consumption

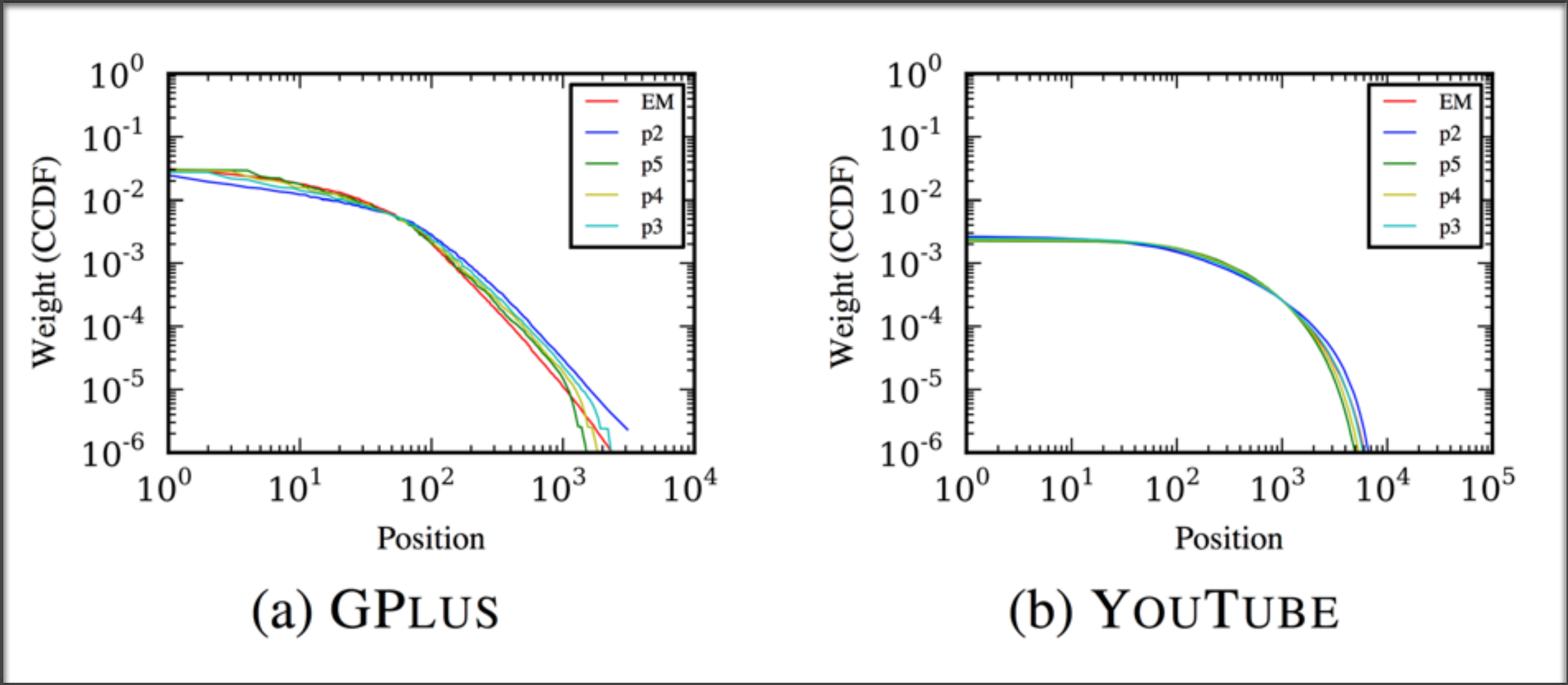
1. empirical analysis
2. models
3. experiments

experiments



scores for quality model

experiments



learned recency weights

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

log-likelihood per item of models, normalized by log-likelihood of hybrid model (which is 1.0)

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

hybrid always wins, but recency model is close

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

recency beats quality

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

learning per-item quality scores always beats setting scores to be equal to popularity

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

recency without scores >
recency using popularity as quality scores

experiments

$s(\cdot) =$ $w(\cdot) =$	popularity -	popularity learned	learned uniform	uniform learned
BRIGHTKITE	0.375	0.617	0.637	0.936
GPLUS	0.587	0.801	0.794	0.877
MAPCLICKS	0.383	0.931	0.414	0.989
WIKICLICKS	0.503	0.724	0.687	0.945
YOUTUBE	0.636	0.677	0.924	0.962

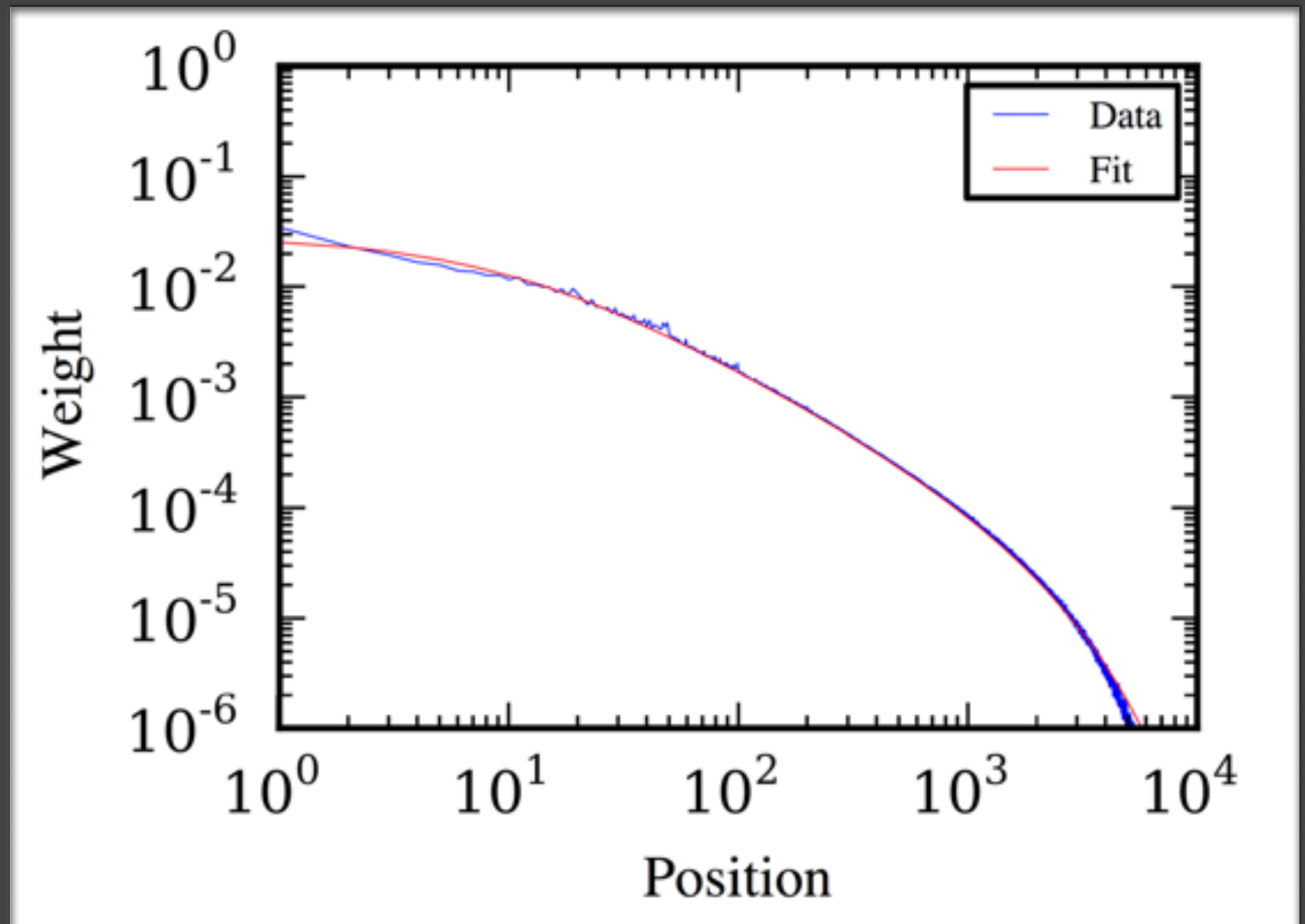
learned quality scores are quite different from popularity (Kendall-Tau coefficient of 0.44)

experiments

currently, we learn a weight for each possible previous position

can our weights be compressed?

experiments



weights follow power law with exponential cutoff

$$\Pr[x] \propto (x+\gamma)^{-\alpha} e^{-\beta x}$$

experiments

Dataset	Recency@50	PLECO
BRIGHTKITE	0.654	0.926
GPLUS	0.710	0.987
MAPCLICKS	0.668	0.921
WIKICLICKS	0.971	0.999
YOUTUBE	0.917	0.997

log-likelihood of variants of recency model
(full recency model set to 1.0)

similar results for hybrid model

conclusion

studied repeat consumption across many domains

found recency and quality to be strong empirical effects in characterizing reconsumption

developed quality, recency, and hybrid models

validated these models on lots of real data

thanks!

recency

two problems:

1. hit ratio depends on number of unique items in the sequence

2. some number of hits is expected

recency

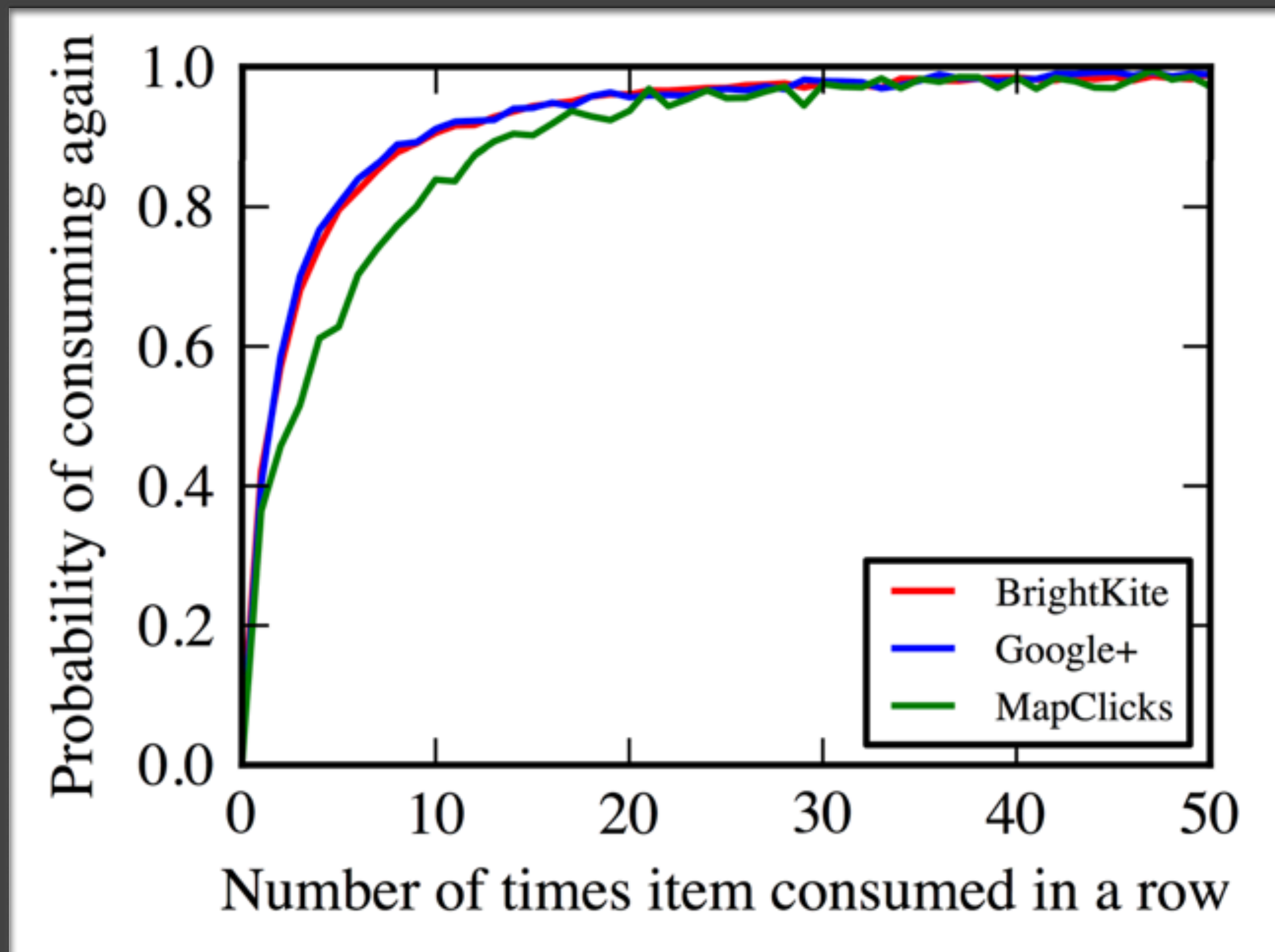
solutions:

1. use normalized hit ratio: divide hit ratio by $1 - u/c$, the upper bound on hit ratio

2. compare to normalized hit ratios on randomly shuffled version of sequences

another baseline: compare to optimal stable cache (fraction of consumptions accounted for by top k items)

satiation



no evidence of satiation in our data