Bias on Search & Recommender Systems

Ricardo Baeza-Yates
Institute for Experiential AI
Northeastern University

Part of this talk appeared in
Comm. of ACM of June 2018 and is an
Updated version of my
ACM RecSys 2020 keynote

ESSIR 2022, Lisbon, Portugal, July 2022

Institute for Experiential AI

Institute for *Experiential* AI

What do we mean by *Experiential* AI?

- AI with human in the loop
- AI applied to real-world problems yielding pragmatic working solutions

Why we believe is EAI the right direction?

Much evidence that pragmatic working AI solutions have two characteristics:

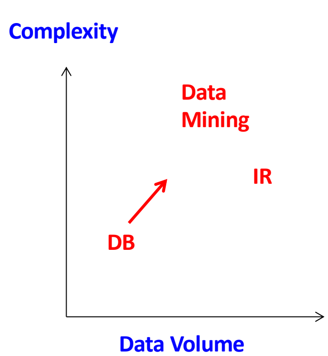
1. ***Human-in-the-loop:*** ability to bring human decision-making, common sense reasoning into the solution operation
2. ***Strong dependence on Data:*** ML and DS to leverage more quality (big) data: "We don't have better algorithms… we just have more data"

Institute for Experiential AI

A High Level View

Data / Query	Unstructured	Structured
Explicit	Information Retrieval	(Relational) Databases
Implicit	Recommender Systems	
Unknown	Data Mining	

A Bit of History



Two different points of view for data

Information Retrieval P. BAXENDALE, Editor

A Relational Model of Data for Large Shared Data Banks

E. F. Codd
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on *n*-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity
CR CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

1. Relational Model and Normal Form

1.1. Introduction

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a paper by Childs [1], the principal application of relations to data systems has been to deductive question-answering systems. Levein and Maron [2] provide numerous references to work in this area.

In contrast, the problems treated here are those of *data independence*—the independence of application programs and terminal activities from growth in data types and changes in data representation—and certain kinds of *data inconsistency* which are expected to become troublesome even in nondeductive systems.

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for noninferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the "connection trap").

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

1.2. Data Dependencies in Present Systems

The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed *without logically impairing some application programs* is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need to be removed are: ordering dependence, indexing dependence, and access path dependence. In some systems these dependencies are not clearly separable from one another.

1.2.1. *Ordering Dependence.* Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each element to participate in several orderings. Let us consider those existing systems which either require or permit data elements to be stored in at least one total ordering which is closely associated with the hardware-determined ordering of addresses. For example, the records of a file concerning parts might be stored in ascending order by part serial number. Such systems normally permit application programs to assume that the order of presentation of records from such a file is identical to (or is a subordering of) the

Volume 13 / Number 6 / June, 1970 Communications of the ACM 377

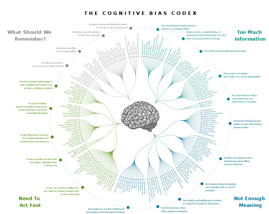
IR: 70 Years

DB: 50 Years

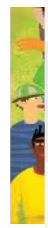
DM: 30 Years

What is Bias?

- Statistical: significant systematic deviation from a prior (unknown) distribution;
- Cultural: interpretations and judgments phenomena acquired through our life;
- Cognitive: systematic pattern of deviation from norm or rationality in judgment;



More than 100 cognitive biases!



Motivation: Impact of Bias in Web Systems

- Most web systems are optimized by using implicit user feedback
- However, user data is partly biased to the choices that these systems make
  - Clicks can only be done on things that are shown to us
- As those systems are usually based in ML, they learn to reinforce their own biases, yielding self-fulfilled prophecies and/or sub-optimal solutions
  - For example, personalization and filter bubbles for users
  - but also **echo chambers for (recommender) systems**
- Moreover, sometimes these systems compete among themselves, learning also biases of other systems rather than real user behavior
- Even more, an improvement in one system might be just a degradation in another system that uses a different (even inversely correlated) optimization function
  - For example, user experience vs. monetization

So (Observational) Human Data has Bias

Cultural Biases

- Gender
- Racial
- Sexual
- Age
- Religious
- Social
- Linguistic
- Geographic
- Political
- Educational
- Economic
- Technological

Statistical Biases

- Gathering process
- Sampling process
- Validity (e.g. temporal)
- Completeness
- Noise, spam

Cognitive Biases

Self-selection

Attempt of an unbiased (personal) view on bias in systems

Many people extrapolate results of a sample to the whole population (*e.g.,* social media analysis)

In addition there is bias when measuring bias as well as bias towards measuring it!

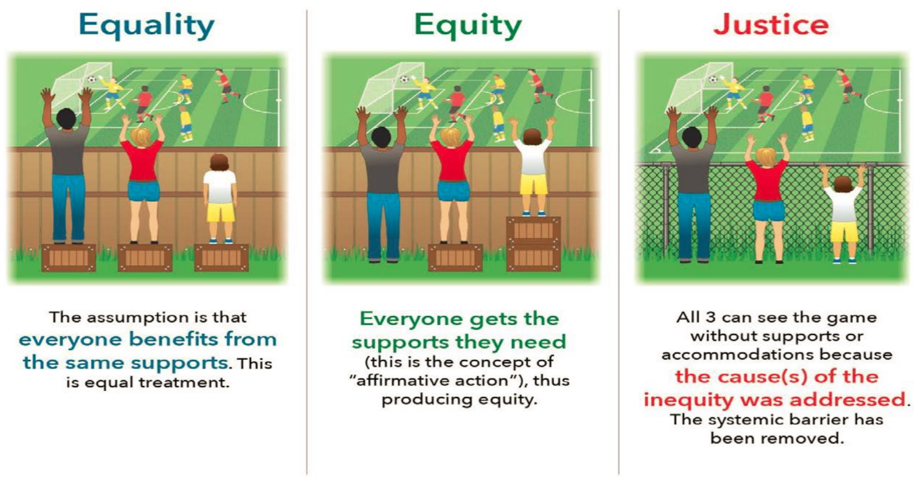
A Non-Technical Question

Biased Data → Algorithm → **Same Bias**

Neutral?
Fair?

Amplified Bias

What is being fair?



A Non-Technical Question

Biased Data → Algorithm → **Same or more Bias**

Neutral?
Fair?

Not Always!
Yes, if you harm people

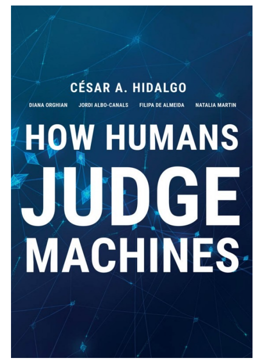
Debias the input
Tune the algorithm
Debias the output

Bias awareness!

ACM US Statement on Algorithm Transparency and Accountability (Jan 2017)

1. Awareness
2. Access and redress
3. Accountability
4. Explanation
5. Data Provenance
6. Auditability
7. Validation and Testing

Systems do not need to be perfect, they just need to be (much?) better than us

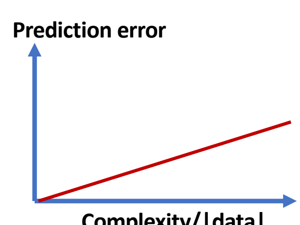


[Hidalgo at al., 2021]
Judgingmachines.com

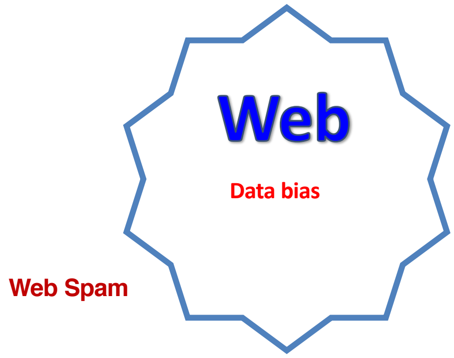
ACM Update to appear soon:
1. Legitimacy & competence

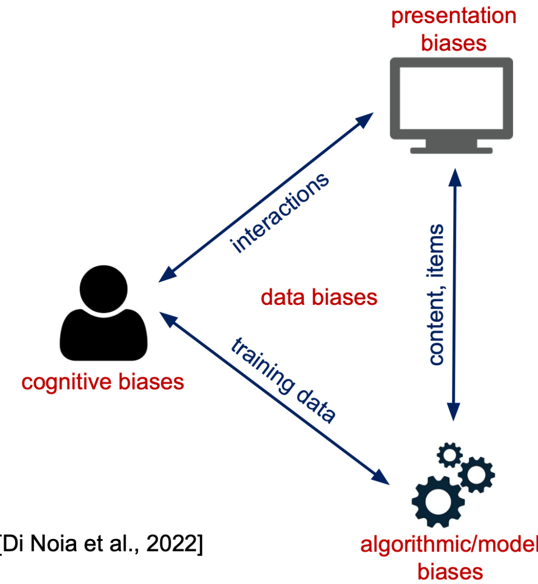
Bias in Computing Systems

- The **quality of any algorithm** is bounded by the **quality of the data that uses** (and hence of its users)
- Data bias awareness
  [Gordon & Desjardins; Provost & Buchanan, MLJ 1995]
- Bias in computing systems [Friedman & Nissenbaum 1996]
  - *"The system systematically and unfairly discriminates against certain individuals or groups of individuals in favor of others"*
- Key issues for Machine Learning
  - Uniformity of data properties
    - In the Web, distributions resemble a power law
  - Uniformity of error
  - Data sample methodology
    - *E.g.*, sample size to see infrequent events or sampling bias



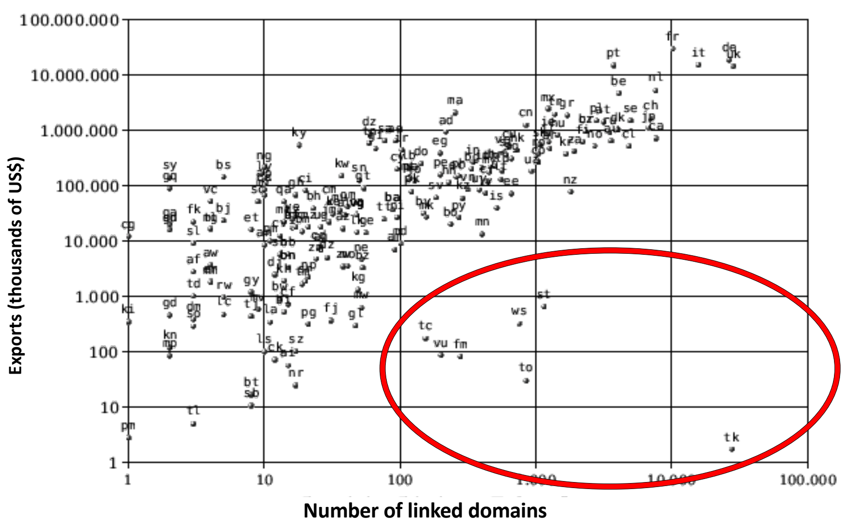
Biases on Search & RS: Web as a Case Study





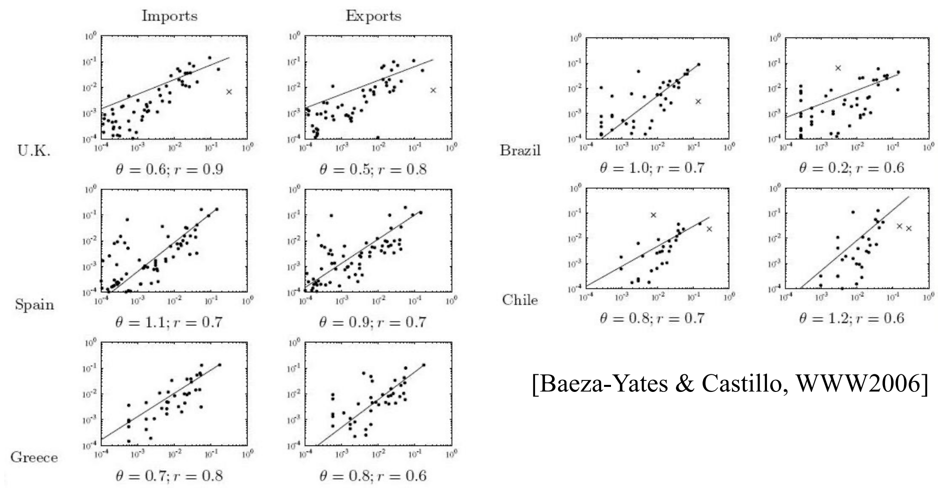
[Di Noia et al., 2022]

Economic Bias in Links



[Baeza-Yates, Castillo & López, 2005]

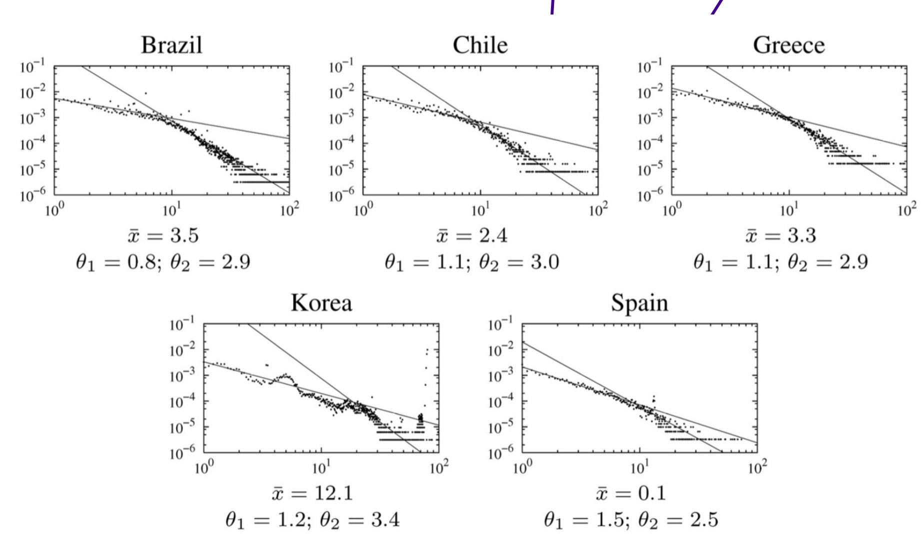
Economic Bias in Links



Cultural Bias in Content

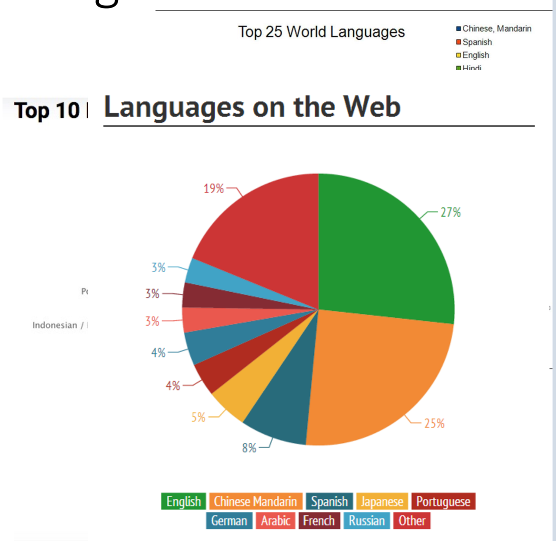
Shame

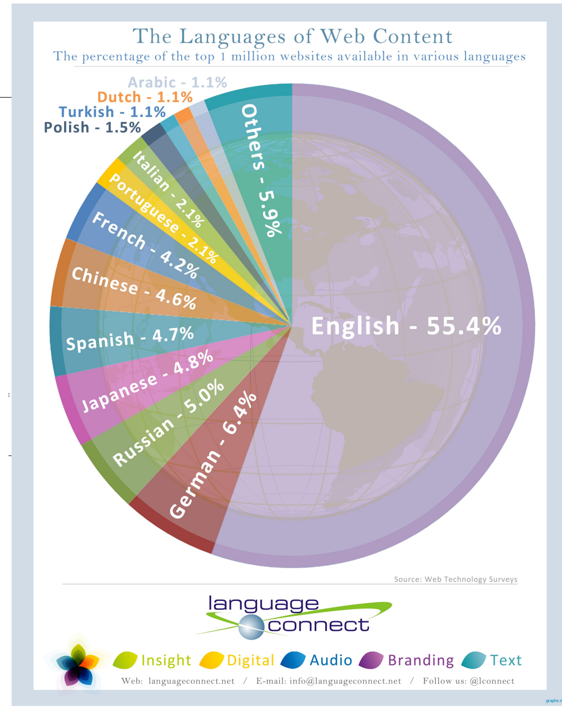
Minimal effort



[Baeza-Yates, Castillo, Efthimiadis, TOIT 2007]

Linguistic Bias





Gender Bias in Content

Tutorial on gender bias on word representations, SIGIR 2022, Madrid

- Word embedding's in w2vNEWS

Gender stereotype *she-he* analogies.

sewing-carpentry	register-nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	hairdresser-barber

Gender appropriate *she-he* analogies.

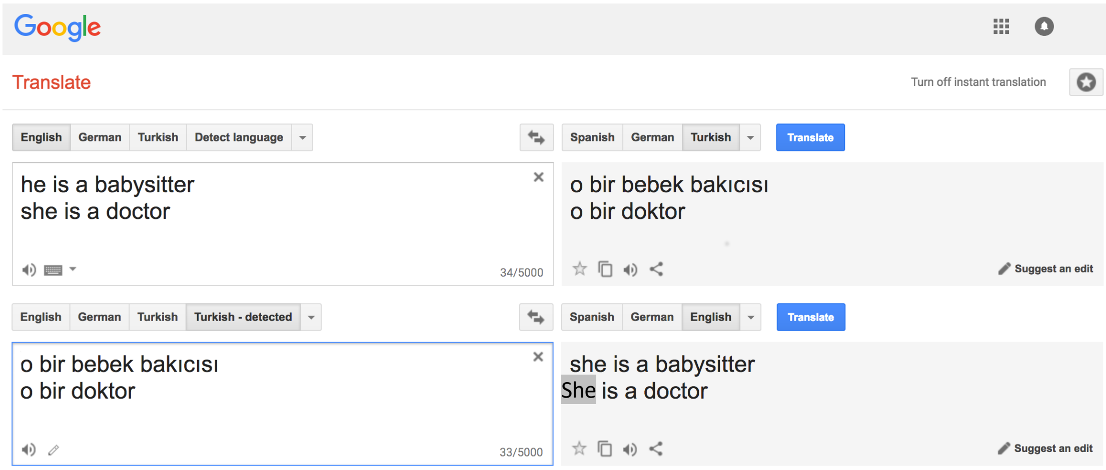
queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Most journalists are men?

[Bolukbasi at al, NIPS 2016]

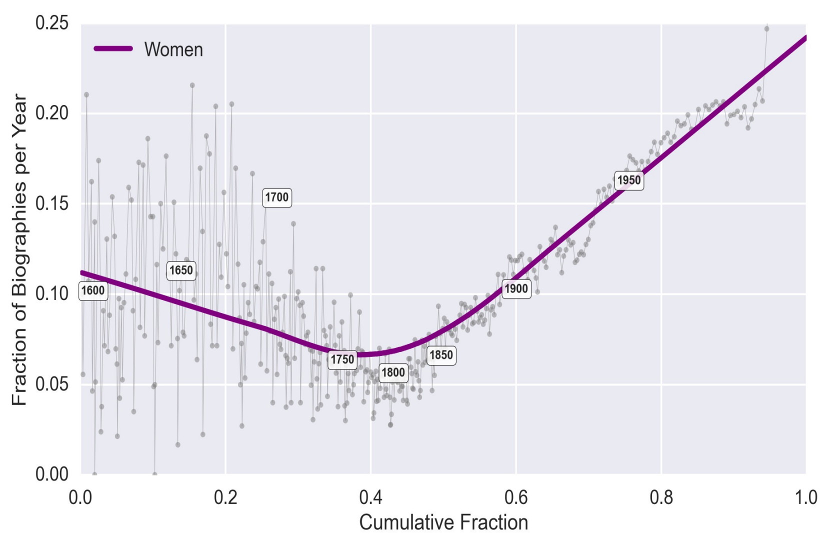
Yes, about 60 to 70% at work although at college is the inverse

Gender Bias in Translation



Gender Bias in Content

Systemic bias?

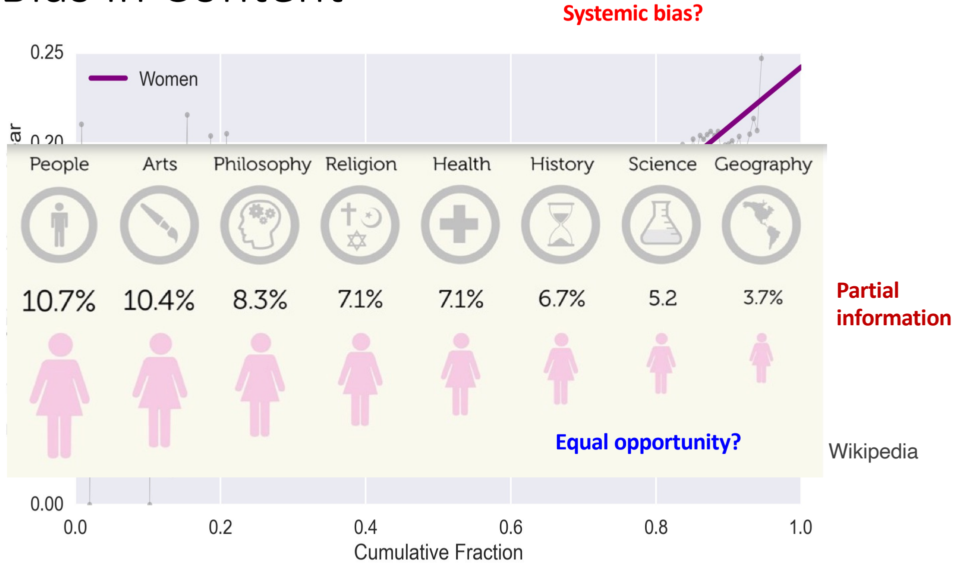


Wikipedia

[E. Graells-Garrido et al,. "First Women, Second Sex: Gender Bias in Wikipedia", ACM Hypertext'15]

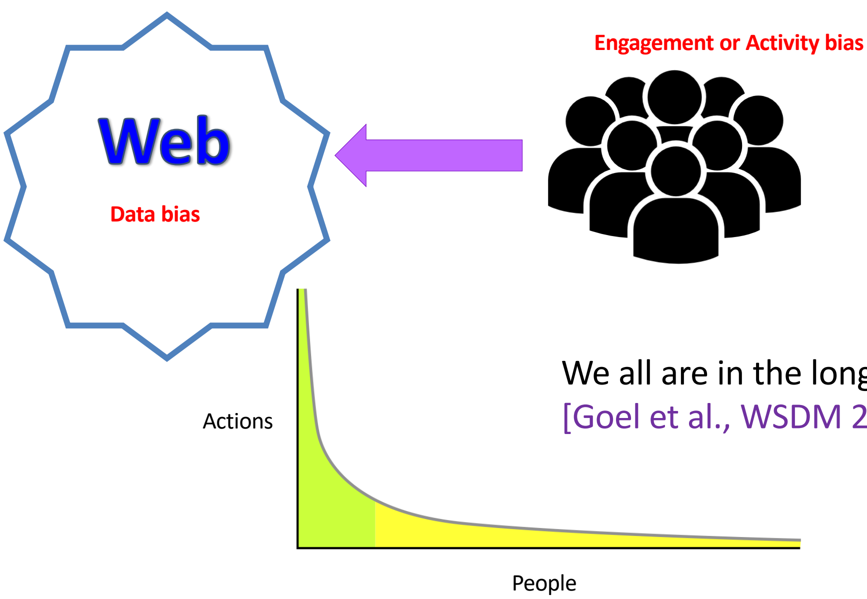
Gender Bias in Content

Systemic bias?



[E. Graells-Garrido et al,. "First Women, Second Sex: Gender Bias in Wikipedia", ACM Hypertext'15]

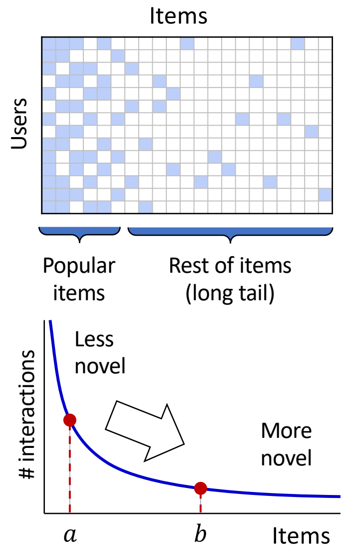
Engagement or Activity Bias



We all are in the long tail!
[Goel et al., WSDM 2010]

Courtesy of Pablo Castells

Popularity Bias in Recommender Systems



- Take care to recommended items that are not too popular
- Metrics

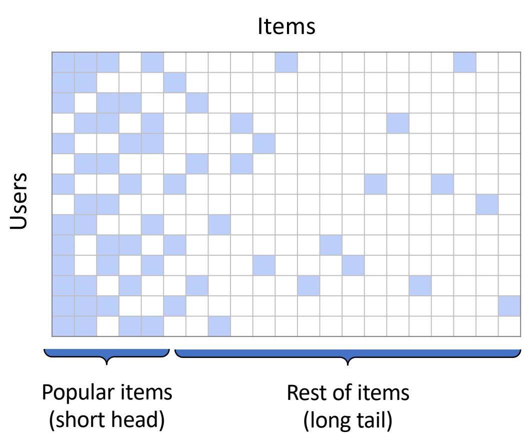
$$nov(i) = 1 - \frac{\#\text{ ratings of } i}{\#\text{ users}}$$

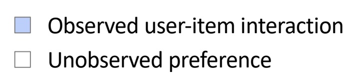
- Novelty enhancement
- Problem solved! …really?

[Vargas & Castells, RecSys 2011]

Courtesy of Pablo Castells

A self-fulfilling prophecy?

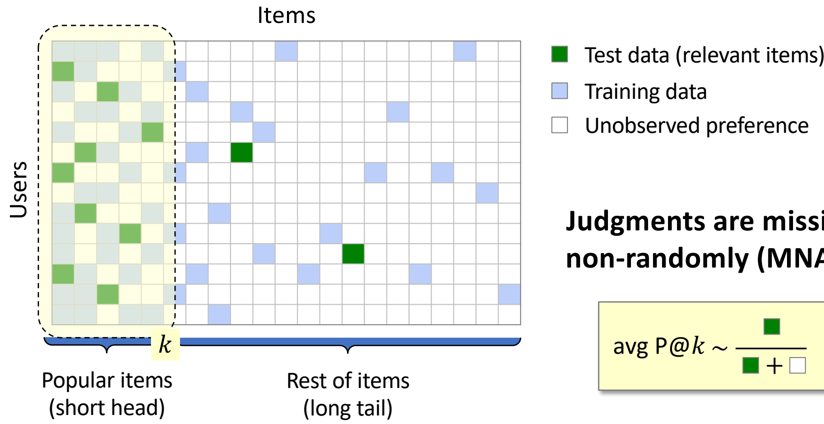




Ratings are missing not at random (MNAR)

Courtesy of Pablo Castells

A self-fulfilling prophecy?



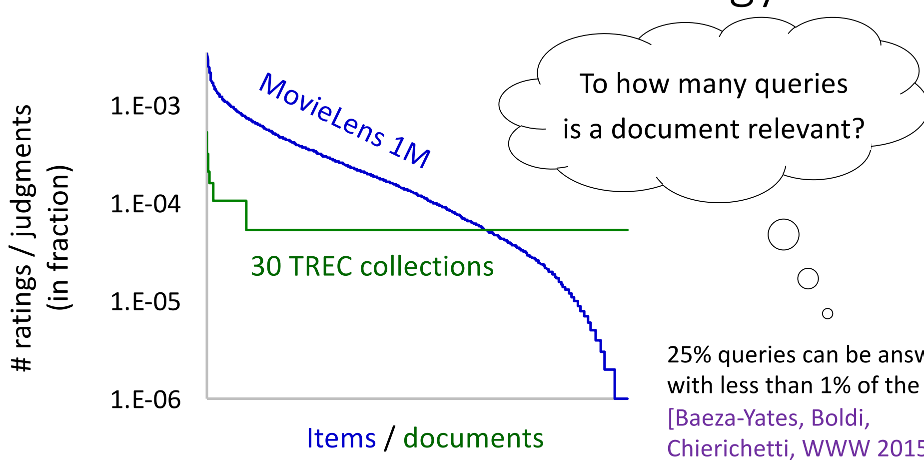
Judgments are missing non-randomly (MNAR)

Worse yet: user-system reinforcement loop (more later)

[Marlin et al., RecSys 2010]
[Steck, RecSys 2010, 2011]
[Fleder & Hossanagar, Management Sciences 2009]

Courtesy of Pablo Castells

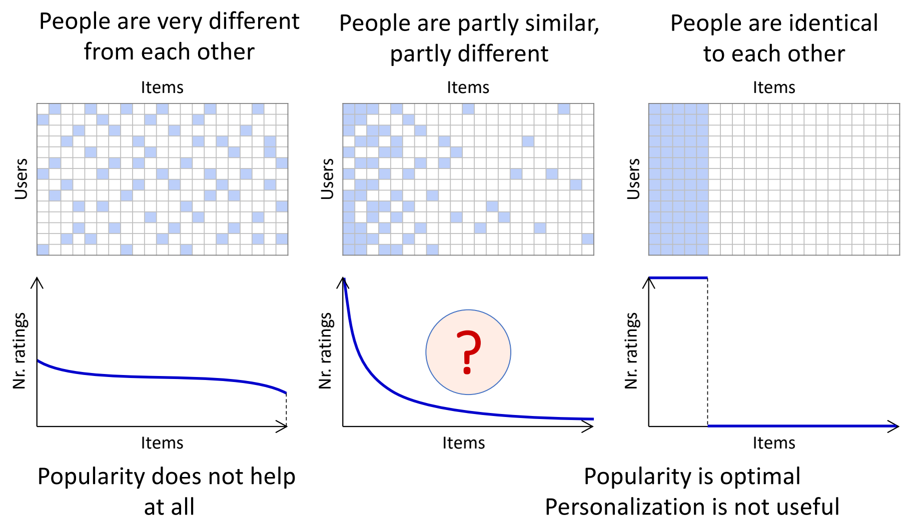
A problem for IR evaluation methodology!



25% queries can be answered with less than 1% of the URLs!
[Baeza-Yates, Boldi, Chierichetti, WWW 2015]

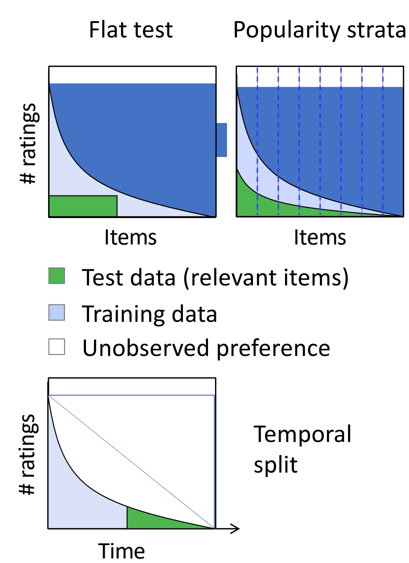
[Bellogín, Castells & Cantador IRJ 2017]

How different or similar are we to each other?



Get rid of the popularity bias!

- In the rating split
  [Bellogín, Castells & Cantador, IRJ 2017]
- In the metrics
  - Stratified recall
    [Steck, RecSys 2011]
  - Importance propensity scoring
    [Yang et al., RecSys 2018]
- In the algorithms
  [Steck, RecSys 2011]
  [Lobato et al., ICML 2014]
  [Jannach et al., UMUAI 2015]
  [Cañamares & Castells, SIGIR 2018, **best paper award**]

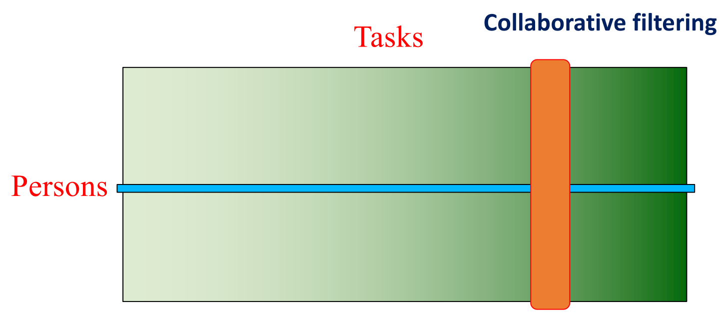


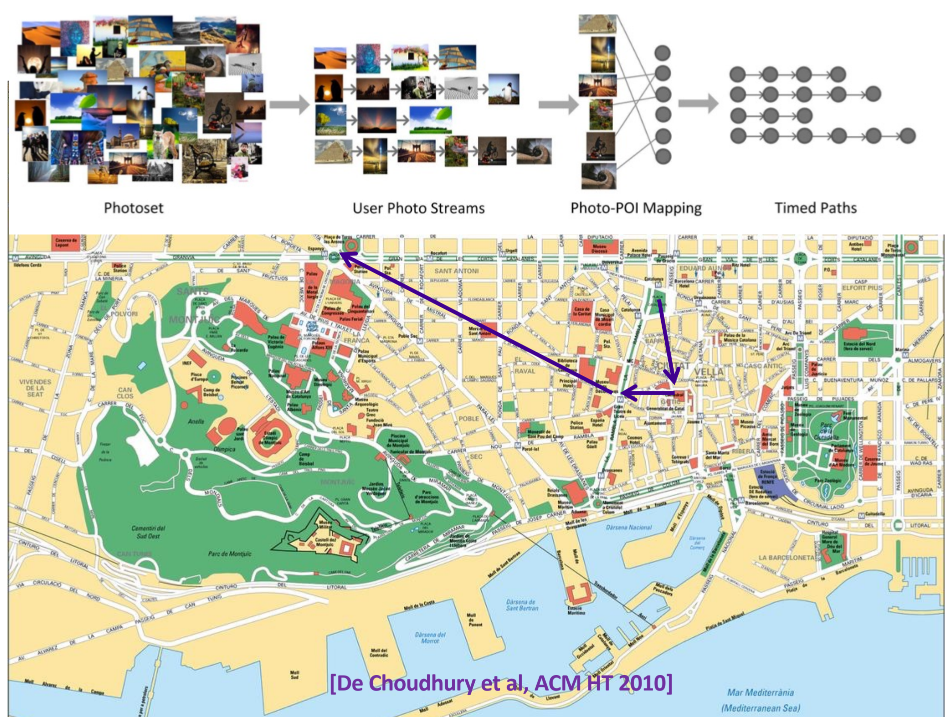
Recommending within the Long Tail

- Exploit the context (and deep learning!)

  91% accuracy to predict the next app you will use
  [Baeza-Yates et al, WSDM 2015]

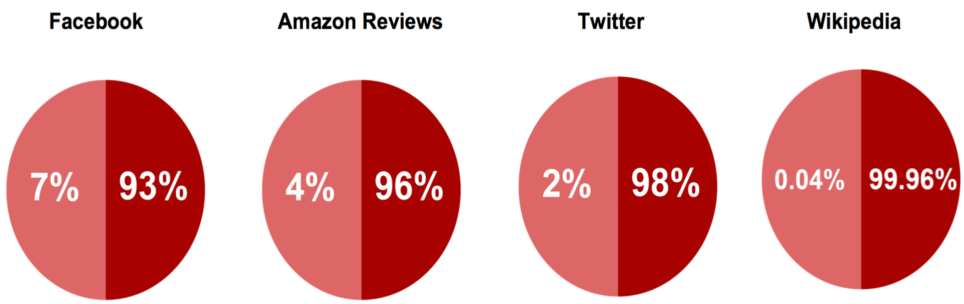
- Personalization vs. Contextualization
  Break the filter bubble! (more later)





Engagement/Activity Bias also Affects Content

Most users are passive (*i.e.,* more than 90%) – wisdom of crowds is a partial illusion
Hence, which percentage of **active** users produce 50% of the content?



[Baeza-Yates & Saez-Trumper, ACM Hypertext 2015]

Social Bias

October 2015

theguardian

ort football opinion culture business lifestyle fashion environment tech travel all sections

Amazon sues 1,000 'fake reviewers'

Online retailer files lawsuit in US against people whose names it says it does not know, claiming they offer reviews for sale

TECH

Amazon is filled with fake reviews and it's getting harder to spot them

PUBLISHED SUN, SEP 6 2020·9:00 AM EDT

Katie Schoolov

LAWSUIT

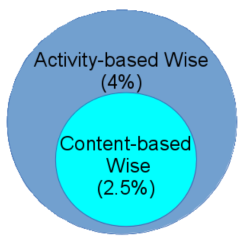
Amazon sues administrators of more than 11K Facebook groups that allegedly brokered fake reviews

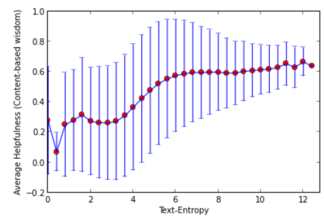
CNNWire By Brian Fung, CNN

Wednesday, July 20, 2022 12:06AM

Quality of Content?

- Adding content ⇛ Adding Wisdom ?
- We use Amazon's Reviews helpfulness
- Content-based-wise users

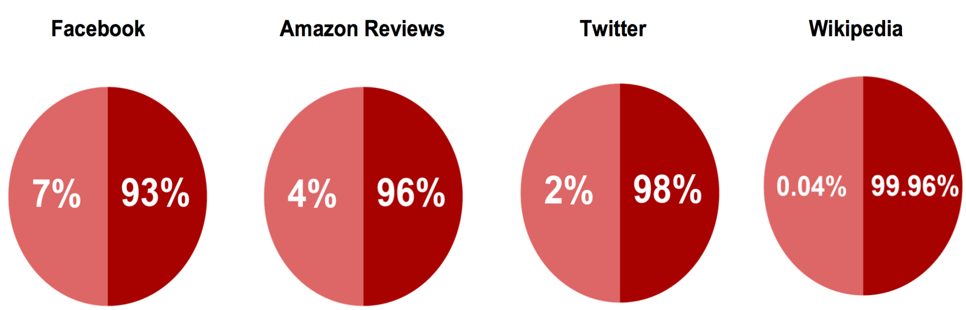




[Baeza-Yates & Saez-Trumper, ACM Hypertext 2015]

Wisdom of a Few?

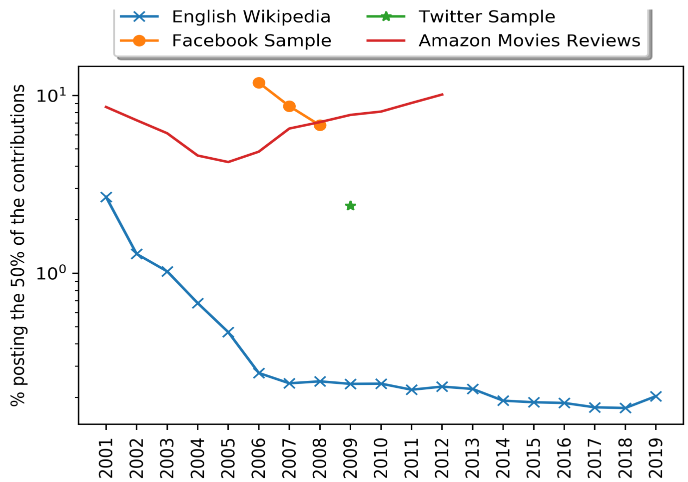
Which percentage of **active** users produce 50% of the content?
Similar to the 90-9-1 rule of Internet participation [Nielsen 2006]



[Baeza-Yates & Saez-Trumper, ACM Hypertext 2015]

Temporal Dynamics of Engagement Bias

Which percentage of **active** users produce 50% of the content?

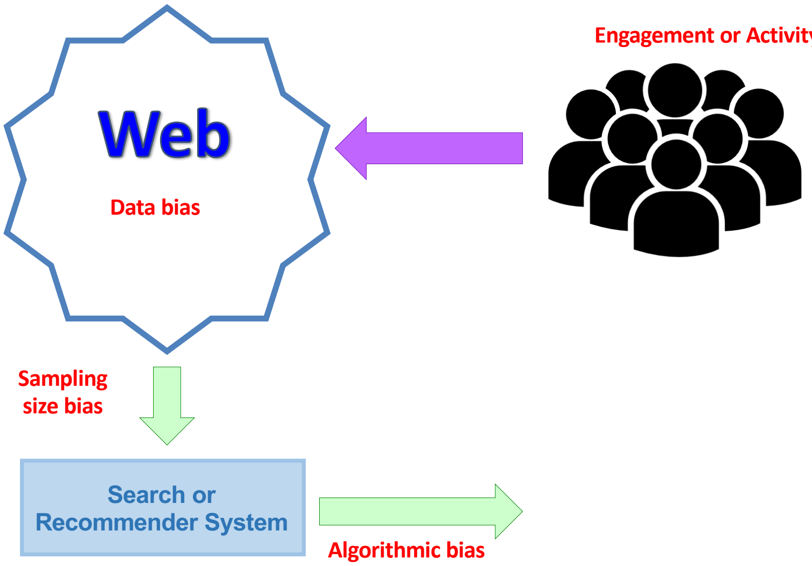


[Baeza-Yates & Saez-Trumper, unpublished]

Attention Bias: The Digital Desert

[Baeza-Yates & Saez-Trumper, ACM Hypertext 2015]

Bias in the Web



Sample Size?

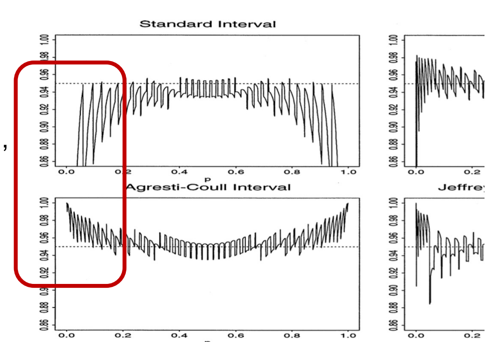
- If we want to estimate the frequencv of queries that appear with probability at least p with a certain relative error ϵ we can use the standard binomial error formula √(1-p)/*np* which works well for p near ½ **but not for p near 0**

- Better is the Agresti-Coull technique (also called *take 2*) which gives:

$$n \geq Z^2_{1-\alpha/2}\left(\frac{p'(1-p')}{\epsilon^2} - 1\right)$$

where Z is the inverse of the standard normal distribution, $1-\alpha$ is the confidence interval and $p' = p + Z^2/2$

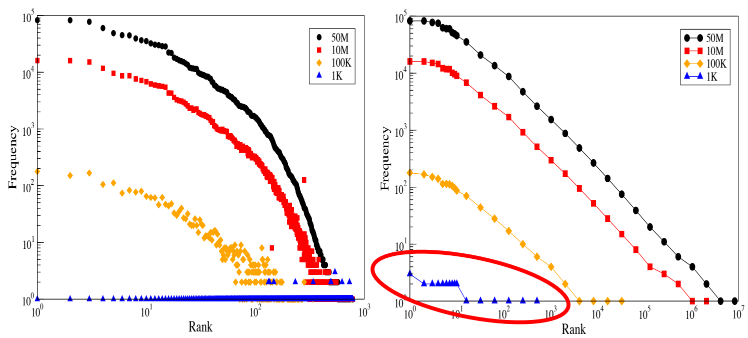
- If p = 0.1, $1-\alpha$ is 80% and ϵ is 10%, we get n = 2342. The standard formula gives n = 900!



[Brown, Cai & DasGupta, Statistical Science, 2001]
[Baeza-Yates, SIGIR 2015, Industry track]

Sampling Techniques

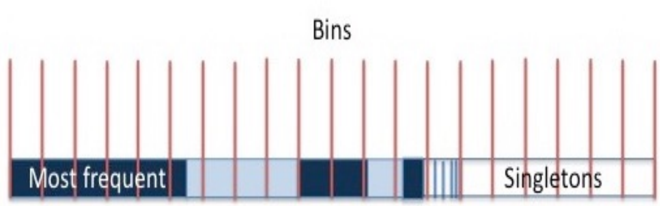
- Standard technique: $$p_q \approx \hat{p}_q(\mathcal{S}) = \frac{f_q(\mathcal{S})}{\sum_{q' \in \mathcal{S}} f_{q'}(\mathcal{S})}$$
- A good sample should cover well all the items distribution but this does not work with very skewed distributions.



[Zaragoza et al, CIKM 2010]

Incremental Stratified Sampling

- Main goal: make good samples consistent across time
- Simple idea based in stratified sampling: bins + random start point



- Bin size can be found by binary search starting with a good approximation if a query frequency model is used ($b < V/n$)
- This perfectly mimics the head of the distribution, but not the tail
- Change the bins in the tail to get the right distribution

[Baeza-Yates, SIGIR 2015, Industry track]

Fixing the Tail

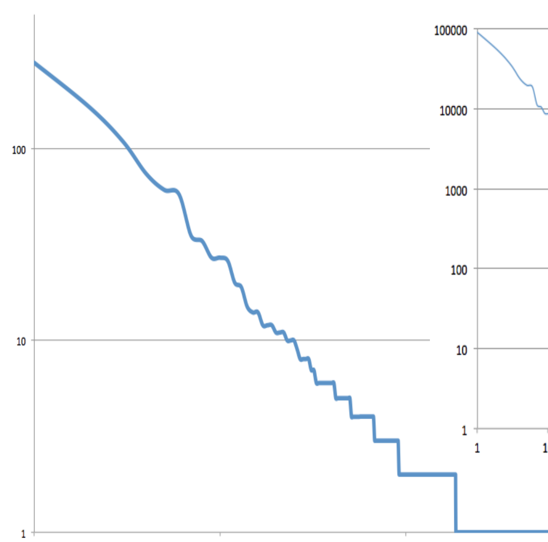
- To mimic the tail we change the binning size when we reach a query frequency of *b/*2
- If we want a singleton ratio of $\beta = S/V$ we recalculate the binning size as

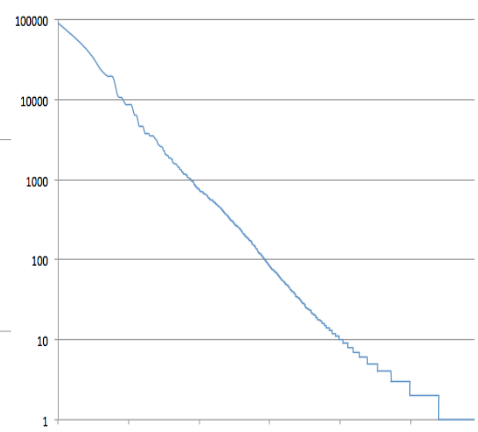
$$b' = (1-\beta)(Q-Q')/(\beta V')$$

- where Q' and V' are the partial vocabulary size and volume before changing the bin size.

[Baeza-Yates, SIGIR 2015, Industry track]

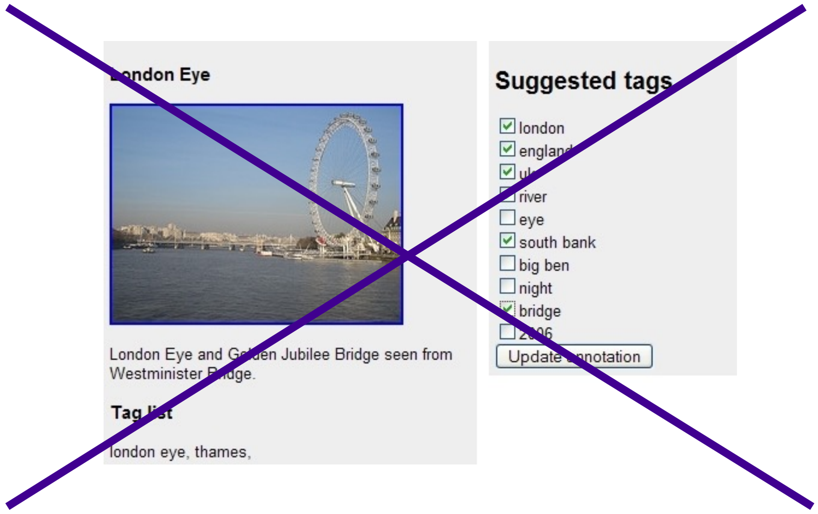
Stratified Sampling Example



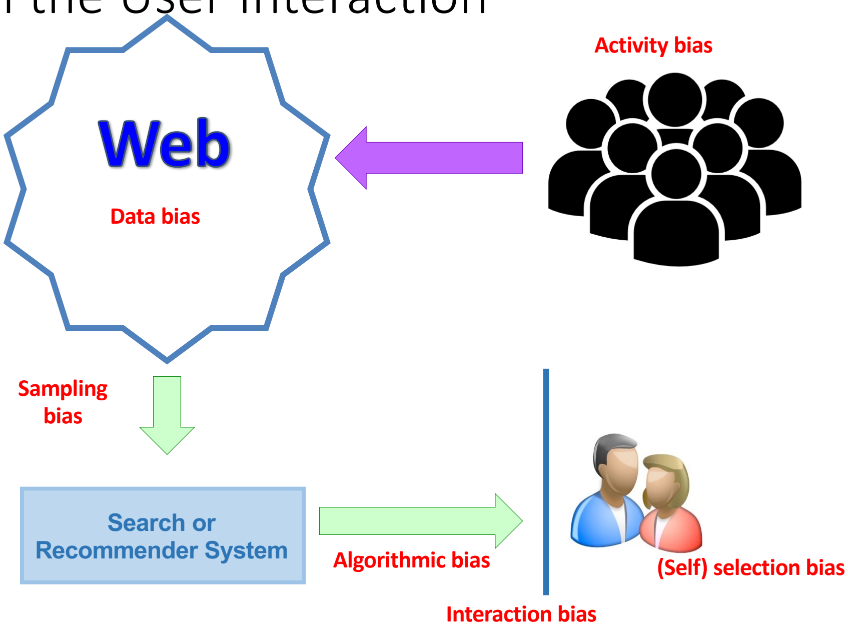


[Baeza-Yates, SIGIR 2015, Industry track]

Extreme Algorithmic Bias



Bias in the User Interaction



Most surely you will read this at the end

You will read this first

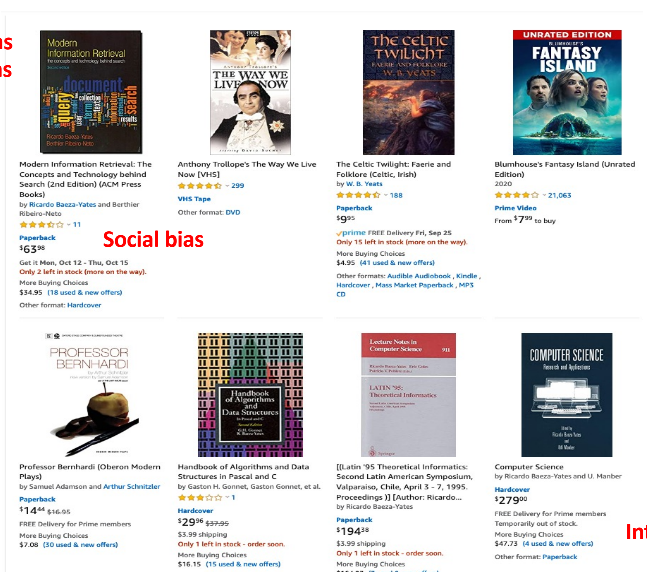
Then you will read this

Probably you will not read this: Ricardo loves geography

Bias in the Interaction

Exposure or Presentation bias

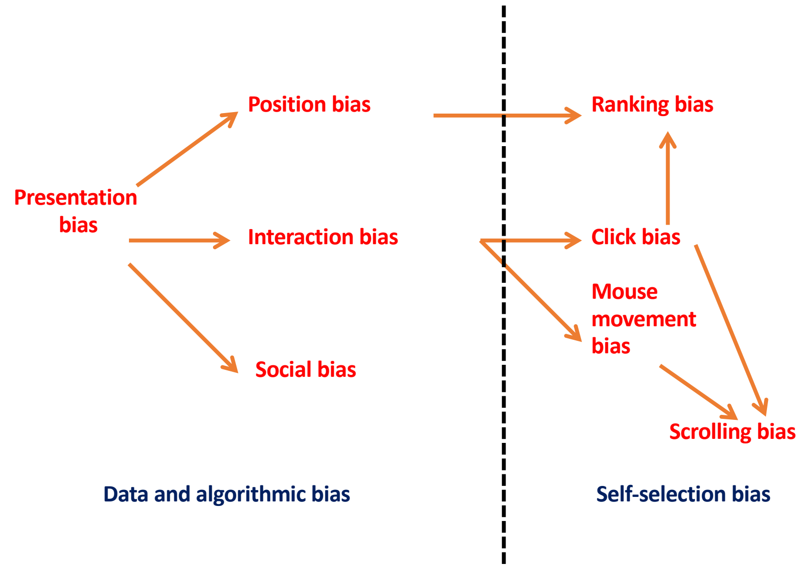
Position bias
Ranking bias



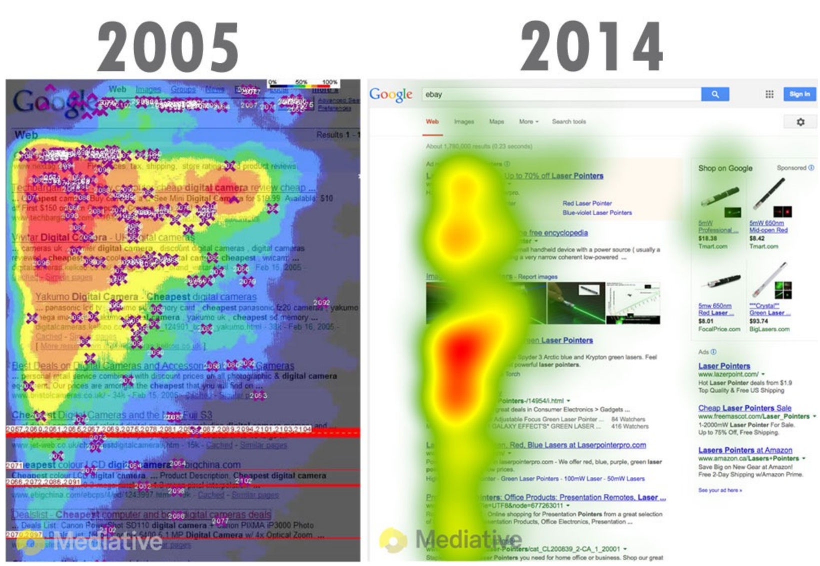
Social bias

Interaction bias

Dependencies: A Cascade of Biases!



Ranking Bias in Web Search



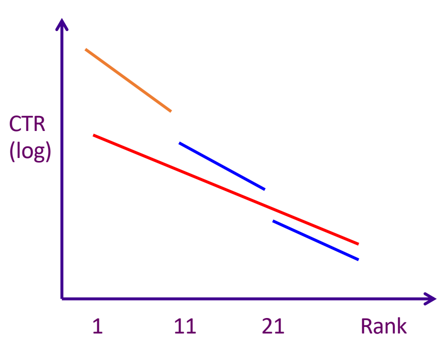
[Mediative Study, 2014]

Ranking Bias: Click Bias in Web Search

- Ranking & next page bias

Debiasing Search Clicks and Other Biases

Clicks as implicit positive user feedback



Learning to Rank with bias
[Joachims et al., WSDM 2017, **best paper**]
+ many other papers

Tune the algorithm

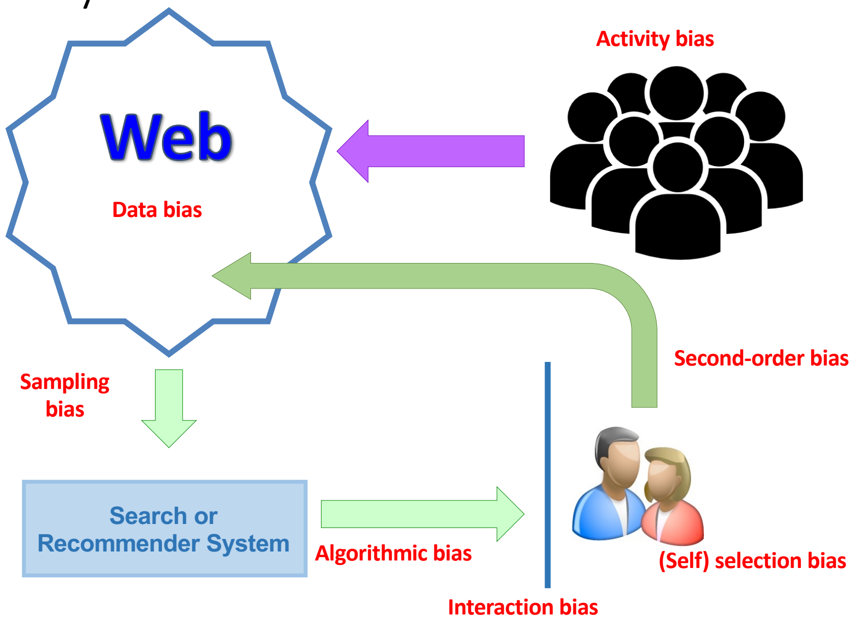
Fair rankings
[Zehlike et al., CIKM 2017]
+ many other papers

Debias the output

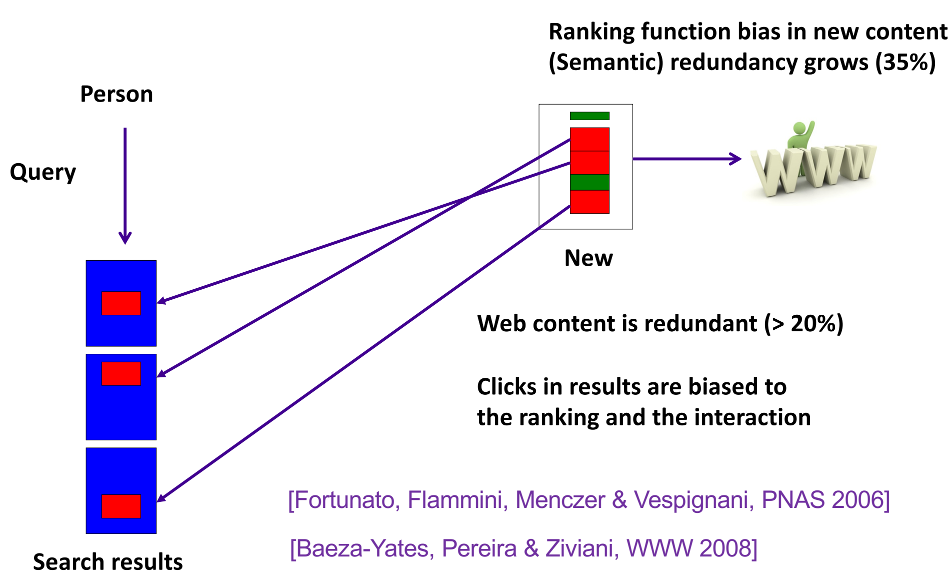
[Dupret & Piwowarski, SIGIR 2008]
[Chapelle & Zhang, WWW 2009]
[Dupret & Liao, WSDM 2010, **best paper**]

Debias the input

Vicious Cycle of Bias



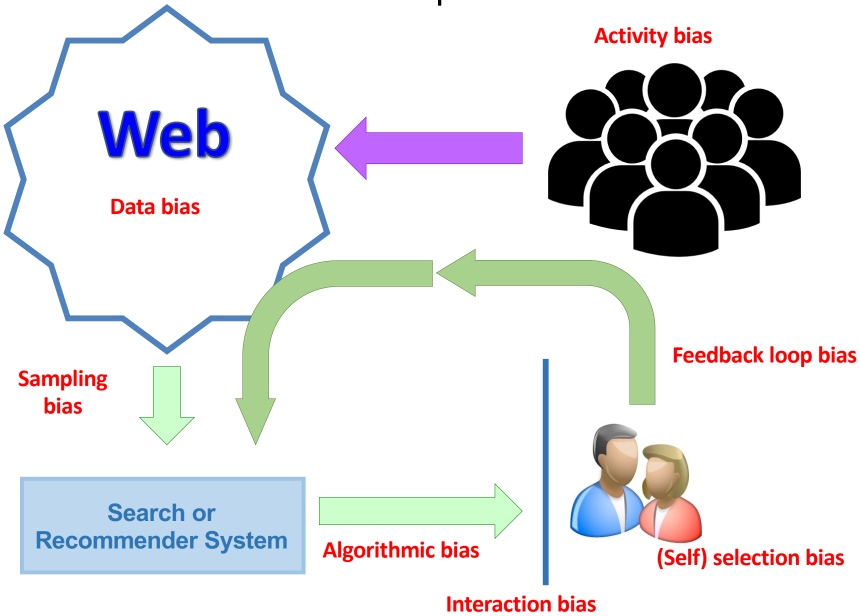
Second Order Bias in Web Content



[Fortunato, Flammini, Menczer & Vespignani, PNAS 2006]

[Baeza-Yates, Pereira & Ziviani, WWW 2008]

Bias in the Feedback Loop



Bias due to Personalization

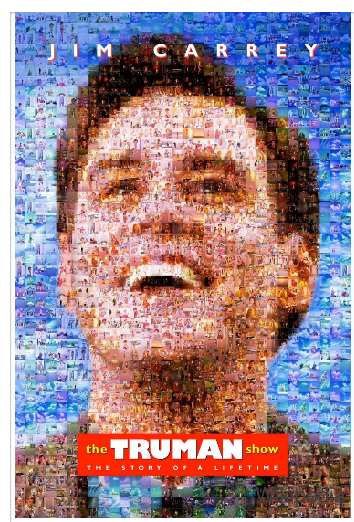
- Partially the effect of self-selection bias
- Avoid the rich get richer and poor get poorer effect
- Avoid the echo chamber by empowering the tail

Partial solutions:

- Diversity
- Novelty
- Serendipity
- My dark side

Cold start problem solution: Explore & Exploit

How much exploration is needed to counteract exposure bias?

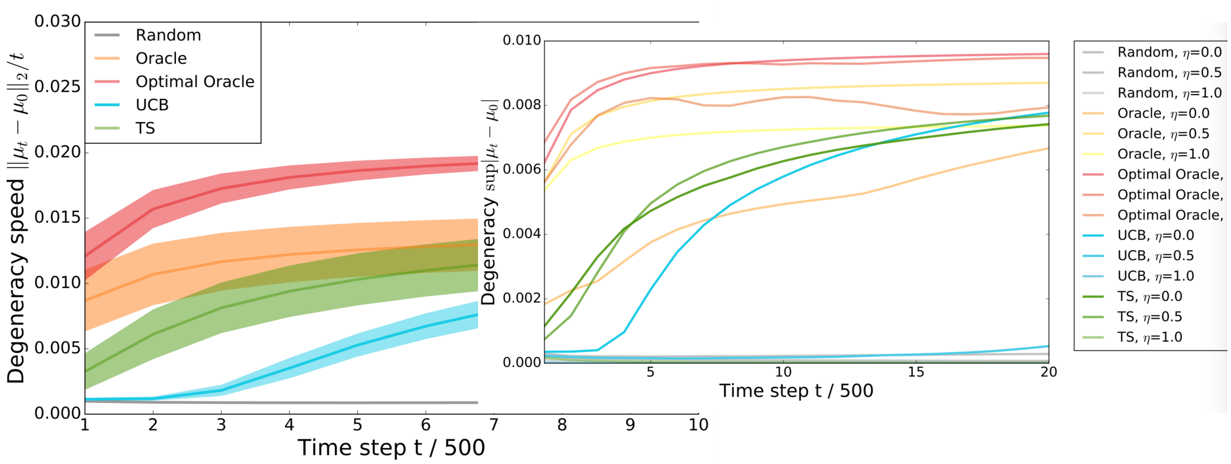


[Eli Pariser, The Filter "Bubble", 2011]

Echo Chambers in Feedback Loops

- For users
  - Filter bubbles
  - Degenerate feedback loops (e.g., YouTube autoplay)
- For systems
  - Short-term greedy optimization
  - The system is partly writing its own future (exposure bias)
  - Partial knowledge of the world if not enough exploration/traffic
  - The **system itself is also in a big bubble**!

Users' Echo Chambers in Feedback Loops



[Jiang et al. Degenerate Feedback Loops in Recommendation Systems, AAAI 2019]

Echo Chamber of the Recommender System

- Short-term greedy optimization, partial knowledge of the world
- Long-term revenue optimization is not achieved
- **Disparate impact**: unfair ecommerce/information markets

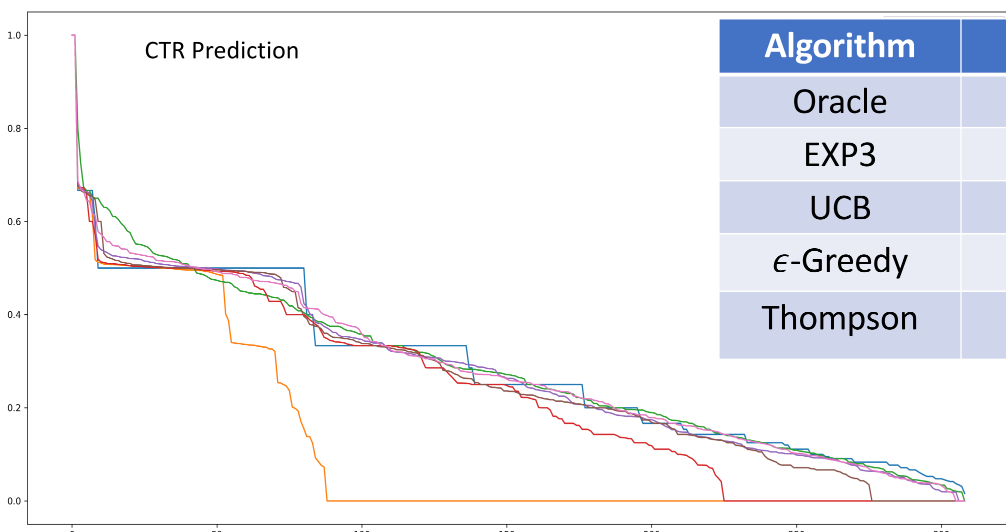
- Can we do better?
- Yes, if the amount of new traffic allows **enough** exploration for new items or any other changes in your world

$$\Delta Traffic \geq \text{Approximate}(\Delta World)$$

- Otherwise we will live in a sup-optimal solution

Counterfactual Approach

[Baeza-Yates & Delnevo, unpublished]



Algorithm	Revenue
Oracle	100%
EXP3	72%
UCB	67%
ϵ-Greedy	62%
Thompson	6%

Fairness and Ethics

- Consumers & long tail items/players are discriminated
- Matthew effect again: rich get richer, poor get poorer
- Unfair markets are unhealthy and hence less stable in the long term

- Internet Companies Antitrust - Advertising Transparency
  - Amazon's Antitrust Paradox [Khan, 2017]
- Should marketplaces sell in their own marketplace?
  - Yes, but with regulations [Hagiu, Teh & Smith, 2020]
  - Is data asymmetry ethical? (not new, but gets amplified in e-commerce)
- Fair markets could be better revenue wise
  - Fairness trade-offs [Mehrotra et al., 2018]

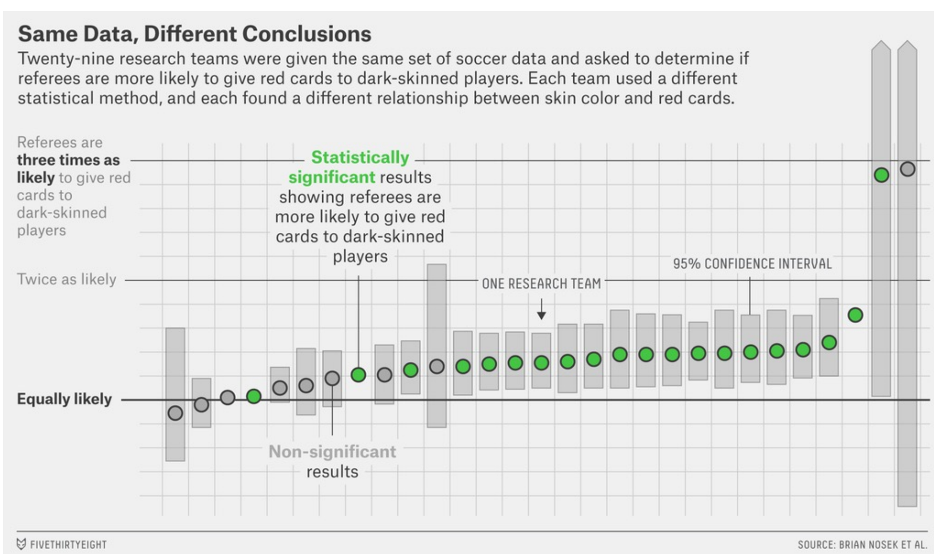
Recap

Bias \ Type	Statistical	Cultural	Cognitive
Algorithmic	♦	?	?
Presentation	♦		
Position	♦	♦	♦
Data	♦	♦	
Sampling	♦	♦	♦
Activity		♦	
Self-selection		♦	♦
Interaction		♦	♦
Social		♦	♦
Second order	♦	♦	♦

Our Professional Biases

- Problems
  - Our **big data and deep learning bias**: small data is more frequent & harder [Baeza-Yates, KD Nuggets, 2018] [Ng, IEEE Spectrum, 2022]
- Design and Implementation
  - Do systems reflect the characteristics of the designers?
  - Do systems reflect the characteristics of the coders?
- Evaluation [Silberzahn et al., COS, Univ. of Virginia, 2015] [Johansen et al., Norway, 2020]
  - Choose the right experiment
  - Choose the right test data
    - Pool bias in search test collections [Lipani et al., SIGIR 2015, CIKM 2016]
  - Choose the right metric(s)
  - Choose the right **baseline(s)**
  - Julio Gonzalo's talk: http://tiny.cc/ESSIR2019-juliogonzalo

It's Hard to Get the Truth from Data



➔ 61 analysts, 29 teams: 20 yes and 9 no
➔ [Silberzahn et al., COS, Univ. of Virginia, 2015]

What we can do?

- Data
  - Analyze for known and unknown biases, debias/mitigate when possible/needed
  - Recollect more data for sparse regions of the problem
  - Do not use attributes associated directly/indirectly with harmful bias
- Design and Implementation
  - Make sure that the model is **aware** of the biases all the time
  - Let experts/colleagues/users contest every step of the process
- Interaction
  - Make sure that the user is **aware** of the biases all the time
  - Give more control to the user
- Evaluation
  - Do not fool yourself!

The Web Works Thanks to Bias!

- Web traffic
  - Local caching
  - Proxy/network caching
- Search engines
  - Answer caching
  - Essential web pages
    - 25% queries can be answered with less than 1% of the URLs! [Baeza-Yates, Boldi, Chierichetti, WWW 2015]
- E-Commerce
  - Large fraction of revenue comes from few popular items
  - But a large fraction of revenue goes to the marketplace owner

Engagement/Activity bias

(Self) selection bias

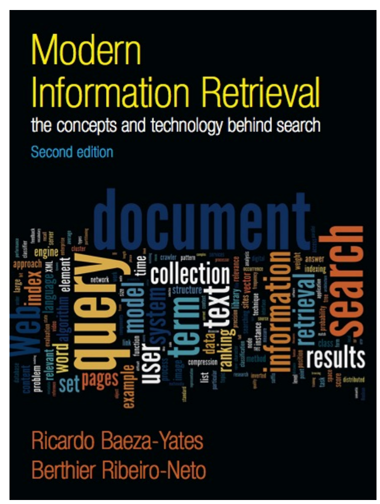
Final Take-Home Messages

- Systems are a mirror of us, the good, the bad and the ugly
- The Web amplifies everything, but always leaves traces

- We need to be aware of our **own biases**!
- We have to be aware of the biases and contrarrest them to stop the **vicious bias cycle**
- We should be **fair**
- **Plenty** of open research problems! (in **small data** even more!)

Questions?

ASIST 2012 Book of the Year Award (Biased Ad)



New Conferences that started in 2018:

AAAI/ACM Conference on AI, Ethics, and Society
http://www.aies-conference.com

ACM FAccT: Fairness, Accountability, and Transparency
http://facctconference.org

Contact: rbaeza@acm.org
www.baeza.cl
@polarbearby

Biased Questions?